Designing and Developing a Spaced Repetition Tool to Improve Learning with Recorded Lecture Videos

Eric Janto

MInf Project (Part 1) Report
Master of Informatics
School of Informatics
University of Edinburgh

2023

Abstract

In this project, we investigated how we can use an active learning method, spaced repetition, to improve learning with recorded lecture videos. Specifically, we proposed embedding interactive flashcards into recorded lecture videos, and then using a spacing algorithm to schedule review of the flashcards. This implementation allows students to recall the lecture contents at a later point in time, thereby encouraging more active learning behaviour.

To achieve this, we designed and developed a prototype of a spaced repetition tool. The development was structured into two stages: firstly, we used participatory design to inform the design process of a low-fidelity prototype. This was supported by a formative evaluation with a Human-Computer Interaction expert and peers. Secondly, we developed a high-fidelity prototype, based on the low-fidelity prototype. We evaluated it with students from the target audience. The usability of the prototype was evaluated using the System Usability Scale. A final task-based evaluation was conducted to examine students' learning experience with the prototype. The results were interpreted using thematic analysis.

Our summative evaluation found that the tool had high usability and the potential to enhance the learning experience when studying with recorded lecture videos. This project contributes to the field of educational technology by providing a unique approach to improving the learning experience with videos through the use of spaced repetition. Our findings suggest that this tool has the potential to guide future research in this area and offers a practical solution to the challenge of improving the effectiveness of educational videos.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 7235
Date when approval was obtained: 2023-01-26
The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Eric Janto*)

Acknowledgements

I would like to thank my supervisor Prof John Lee for providing outstanding guidance and invaluable feedback during the completion of this project.

I would also like to express my gratitude to all the people who helped with the evaluation of this project.

Finally, I would like to thank my family and friends who have always supported me.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

In recent years, we have observed an increase in the use of videos to deliver educational content in an online learning setting. The motivations behind using videos include the ability to reuse material, the potential to scale the number of students, and the suitability for distance learning. In particular, the Higher Education sector is a prominent application domain given the current trend of universities shifting towards a hybrid course delivery model as part of post-COVID-19 adaptations [23].

Unfortunately, several issues exist when using video for educational purposes which lead to a decreased learning outcome and diminished learning experience. With the medium inherently being a passive learning resource [36], students struggle to focus on the video content for a longer period [10]. Additionally, students may be tempted to engage in unproductive watching patterns such as binge-watching [9].

We believe that incorporating an active learning method [36], spaced repetition [15], with educational videos is a novel approach that can further improve how videos are used for online learning. This approach aims to increase learning outcomes and improve the learning experience of students.

1.2 Project Goals

The goal of this project is to investigate the suitability of using spaced repetition for learning with recorded lecture videos (RLVs). To help explore this, we strove to build a tool that integrates interactive flashcards with RLVs, stores those flashcards in a database, and presents them to students for later review.

With this goal in mind, we posed the following research questions:

1. RQ_1: How does our tool affect students' learning experience with RLVs?
2. RQ_2: How should our tool be designed to integrate spaced repetition and flashcards with RLVs to meet students' needs?

3. RQ_3: What technologies should we use to implement our tool so that it can be integrated with existing software?
   a) What existing technological infrastructure does the tool need to be integrated with?
   b) What technologies are compatible with the required infrastructure?
   c) Based on the previous questions, what technologies should be used for implementing our tool?

1.3 Overview of Research Contributions

An overview of the main contributions of this project is provided in Table 1.1. These contributions are further summarised in Section 2.8.2.

Contribution	RQ_1	RQ_2	RQ_3
Identification and development of a unique approach to improving the learning experience with RLVs using spaced repetition		x	x
Design and formative evaluation of a low-fidelity prototype		x	
Implementation of a high-fidelity prototype which instruments our unique approach to improving learning with RLVs		x	x
Usability and learning experience evaluation of the high-fidelity prototype	x	x	

Table 1.1: Contributions of this project and their addressed research questions

Chapter 2

Background and Literature Review

2.1 Learning Experience vs Learning Outcome

For clarification purposes for the rest of this report, we define learning experience and learning outcome as follows.

Learning experience refers to the entire process of acquiring knowledge and skills. This process happens through a variety of activities and interactions with the learning environment, including the teacher, peers, resources, and technologies. It encompasses the different ways in which learners engage with the material, such as reading, listening, and practicing. A learning experience is therefore not limited to a specific moment in time or a single activity, but rather a holistic and dynamic journey of discovery and learning.

By contrast, learning outcome consists of the specific and measurable knowledge and skills that learners are expected to achieve through learning. The term comprises the intended results or goals of the learning process.

2.2 Problems with Recorded Lecture Videos

RLVs are widely used to deliver educational content to students. Unfortunately, fully replacing in-person lectures with RLVs can have several negative implications for students. For example, a diminished learning experience has been reported by students, caused by reduced motivation and engagement [24]. Furthermore, the full replacement with RLVs tends to lead to a decreased learning outcome, that is, a situation where a student's ability to acquire and apply knowledge is reduced [5]. These implications are attributed to two problems with RLVs, detailed below.

2.2.1 Problem 1: Passive Learning Resource

Passive learning is the process of acquiring knowledge by being exposed to information and internalising it by some form of memorisation, whilst leaving little opportunity for a student to actively engage with the material [36]. This can lead to decreased attention, motivation, and engagement with the content [10]. Conversely, active learning assumes that by being more engaged during the learning process, the student learns more effectively and more permanently, whilst having an improved learning experience [36]. In the context of lectures, an increased learning outcome can indeed be observed when active learning elements are included for the students [4][25].

However, from a technological perspective, RLVs are inherently more compatible with passive learning. Native video players such as the HTML5 video player [50] for the web, or the QuickTime player [41] for macOS, do not offer an interface to integrate active learning into a previously recorded lecture video. Platforms such as Microsoft Teams [48] enable interactive online live teaching by providing polling and chat functionalities. These features may work well to engage students during online live lectures but are not capable of incorporating active learning elements into RLVs.

Adding active learning elements to RLVs is a desirable feature that is not supported by traditional video playback systems.

2.2.2 Problem 2: Single-Use Learning Resource

Let a learning resource that is used only once whilst taking a course be defined as a single-use learning resource. After using that resource once, a student might struggle to get more value out of it for their learning process. Thus, a multi-use resource consists of learning material which a student consults repeatedly. Such a learning resource is therefore considered more valuable for the learning process.

Students primarily use RLVs to clarify points after having attended in-person lectures, or to catch up on missed lectures, according to research [46]. This is especially problematic with students commonly engaging in two types of watching behaviour: binge-watching and segment-skipping [9]. Binge-watching is passive learning and therefore undesired. Segment-skipping can be a positive form of active engagement, especially when the student has already reviewed the material, and they can use the video to extract specific information. However, if the student skipped segments during a first watch and never went back to the ignored sections, this would lead to large amounts of an RLV remaining unwatched and a diminished learning outcome.

Here, RLVs as a single-use learning resource can potentially lead to inefficient study methods among students. To promote more efficient study methods when using RLVs, they should be presented as a multi-use resource that discourages binge-watching and the kind of segment-skipping described above.

2.3 Problem with Spaced Repetition and Recorded Lecture Videos

Spaced repetition [15] is a learning technique that involves reviewing material at increasing intervals of time. It is based on the observation that people are more likely to remember information if they encounter it multiple times spaced apart by increasingly larger intervals [35]. This technique is often used in conjunction with flashcards and algorithmic spacing, a technique which involves an algorithm to determine when to review material to increase the learning outcome.

The repeated reviewing of material, as well as the active engagement with flashcards when reviewing, implies that spaced repetition is an active learning technique which presents the contents covered by flashcards as a multi-use learning resource. There is motivation to develop an approach which allows using spaced repetition with RLVs. We review a conceptual problem in the section below.

2.3.1 Problem 3: Incompatibility of Spaced Repetition and RLVs

A common approach to integrating spaced repetition with RLVs is to divide a lecture video into sections and put each one on a digital flashcard. However, flashcards and spaced repetition are known to be only effective when each flashcard only covers a small fragment of knowledge, e.g., a single fact or the meaning of a single variable in a complex formula [34][38].

It is possible to argue that by reducing the duration of lecture video sections, the amount of information would be small enough. However, there are several reasons why this approach is not acceptable: firstly, it requires a large amount of time and manual work. This acts contrary to the earlier discussed benefit of RLVs being a reusable teaching material which can be employed without too much effort. Secondly, the lecturer structured the contents to be continuously consumed, making use of the fact that information is presented in the context of the rest of the lecture. Breaking down a lecture into infinitely smaller chunks of knowledge would remove any context and structure, potentially leading to a decreased learning outcome. And finally, one of the main benefits of spaced repetition is how time-effective the learning technique is. Having to watch sections of an RLV repeatedly when reviewing flashcards counteracts that effectiveness by slowing down the learning process and leading to a decreased learning experience.

There seems to be a fundamental incompatibility when trying to integrate spaced repetition and flashcards with RLVs. Despite that, the benefits of each method are significant enough to warrant exploration of new ways to integrate them while preserving their respective benefits. As found during the critical evaluation of existing tools later in this chapter (Section 2.5), such exploration is not supported by the current state of spaced repetition tools and video players. There is motivation to build a new technological tool to help investigate how spaced repetition can be used for RLVs.

2.4 Problem with the Integrability of Technology

When building a new piece of software, it is essential to consider the integrability of the software with its intended application domain, especially for this kind of project with the optional goal of deploying our tool in a real-world setting. In the case of our tool, its application domain is the existing technology infrastructure of the education sector.

2.4.1 Problem 4: Integrability with the Technological Infrastructure of the Higher-Education Sector

We conducted a systematic review of the frameworks and tools used to build over 70 different OpenCourseWare (OCW) web pages, as part of a research team for a presentation at an upcoming learning and teaching conference [51]. OpenCourseWare pages are online platforms that provide free and open access to educational materials from universities and other educational institutions. Preliminary results find that there is no common framework that was used for a range of them; their technological stack varies widely. Unfortunately, even the Learning Tools Interoperability (LTI) [28] standard, which aims to standardise how different educational technological tools can be combined, is not prevalent. Aiming for integrability with this standard only would not be sufficient. Regarding technology used, the only common denominator was that OCW websites either directly used native web technologies (HTML, JavaScript, and CSS), or used frameworks built on top of those technologies.

Arguably, not all teaching technology in the education sector is built the same way that OCW web pages are built. However, for the purpose of this project, we focussed solely on the reviewed OCW projects as a representative sample. Hence, the integrability problem challenges us to make our tool integrable with the majority of OCW web pages.

2.5 Critical Evaluation of Previous Work

Before forming an approach to solving the problems we previously discussed in this chapter, we conducted a comprehensive literature review of existing tools and prototypes which aim to improve students' learning experience and learning outcome when learning with educational video. As we conducted the review, we noticed that some of the tools could be grouped and reviewed together due to a similarity of functionality. For those cases, we chose a representative example and only included that one in the final evaluation. While some of these examples are targeted at different video types than RLV (e.g., recorded tutorials), they could all be repurposed accordingly. We critically evaluated each of the examples by using the previously established problems as assessment parameters.

The results, shown in Table 2.1, indicate that, to our knowledge, none of the previous work provides a complete solution for the entire problem set. The integrability problem especially remains unaddressed. For Anki, two fields were classified as 'partially' as it technically solves these problems but is not expected to lead to a sufficient learning outcome or experience. This was discussed when we presented the inherent incompatibility of spaced repetition and RLVs (Section 2.3.1).

Tool	Functionality	Active Learning	Multi-Use Resource	Spaced Repetition	Integrable
RIMES [31]	Allows for authoring, recording, and reviewing interactive multimedia exercises during an RLV	Yes	No	No	No
Elicast [39]	Embeds programming exercises within the RLV and provides just-in-time feedback as students try the exercises when they watch the video	Yes	No	No	No
Coursera [13]	Provides the functionality of prompting students with multiple-choice in-video questions	Yes	No	No	No
YouTute [42]	Allows for interacting with recorded material by providing features to tag, segment, and select certain parts of it; encourages purposeful and productive segment-skipping	Yes	Yes	No	No
Anki [1]	Allows for embedding video content on flashcards; provides spaced repetition scheduling for reviewing flashcards	Partially	Partially	Yes	No

Table 2.1: Evaluation of existing tools based on our established problem set

2.6 Proposed Solution

2.6.1 Approach and Projected Success

Our approach to solving the set of problems at hand is to design and implement a tool which adds interactivity to RLVs by inserting flashcards. This turns RLVs into an active learning resource. The flashcards will cover previously-discussed content of the lecture. Additionally, the tool will include a customised spaced repetition system which collects all flashcards encountered by the student while watching an interactive RLV and schedules them using a spacing algorithm. When reviewed later, the flashcards will link back to the RLV where they were encountered in case a student forgot how a flashcard's answer was derived. This aims to provide a solution for the single-use and incompatibility problems. The tool will be web-based and have a web component [11] to make its functionality integrable with other websites. This addresses the integrability problem.

Our approach offers a simple yet effective solution to the incompatibility problem between spaced repetition and RLVs. Instead of putting entire video sections on flashcards, we present a small chunk of knowledge that refers to parts of the lecture. By contextualizing the flashcard contents within the lecture material, our approach maintains the benefits of spaced repetition while potentially improving retention rates [22]. Overall, our approach is expected to improve both the learning experience and outcome of students using RLVs, as it provides reasonable solutions to the entire problem set.

2.6.2 Envisioned Use-Case Scenario

Before making design decisions for the tool, it was important to establish our envisioned use-case scenario for it. This scenario was also used as input for the implementation and evaluation of the tool.

Anna is a college student who is taking an online course. She has a busy schedule and finds it challenging to keep up with the course material. To help her study efficiently, she decides to use a tool that provides interactive flashcards to review lecture content.

Anna starts by watching an interactive recorded lecture video, which is part of the course curriculum. As she watches the video, she reaches a point where a flashcard is triggered. The flashcard presents a question related to the topic she just learned in the lecture.

Anna reads the question carefully, tries to recall the answer from memory, and then presses the 'Show answer' button to reveal it. Depending on whether she remembered the answer or not, she provides feedback by pressing either the 'Remembered' or 'Forgotten' button.

The tool then uses its scheduling algorithm to calculate when the flashcard should be reviewed again, based on Anna's response. Anna gets notified when the next review is due to happen.

On a particular day, Anna visits the review page of the tool to review the flashcards that *are due on that day. The tool presents the flashcards one by one. As Anna reviews each flashcard, she again provides feedback by pressing the 'Remembered' or 'Forgotten' button.*

If Anna remembers the answer correctly, the tool will schedule the flashcard for review again later. If she forgets the answer, the tool will schedule the flashcard for review sooner to help her reinforce her memory.

After reviewing all the flashcards due on that day, Anna receives a message indicating that she has reviewed all flashcards. She can then log out of the tool, knowing that she has made progress in her studies and will continue to do so with the help of the tool.

2.7 Initial Set of Design and Functional Guidelines

Based on the previous parts of this chapter and our research questions, we established a set of design and functional guidelines to support the design and implementation of the tool. Further guidelines and requirements are established later during the process of designing, implementing, and testing the tool.

The initial guidelines focus on the interface of the flashcard as it is expected to be the main interaction point with the tool.

ID	**Design Guideline**	**Justification**
DG_1	The design of the flashcard interface should attract the attention of the user.	Counter of decrease of attention with RLV; RQ_2
DG_2	The design of the flashcard interface should provide a simple and clean presentation of its contents.	Inspiration by existing software tools (Table 2.1); RQ_2
DG_3	The design of the flashcard interface should motivate the user to interact with it.	Encouragement for active learning; RQ_2
DG_4	The usability of the tool should be intuitive and easy to learn so that it does not distract from the learning process.	Goal to build a tool which suits students' needs for learning (RQ_2)

Table 2.2: Design guidelines for our tool and their justification

2.8 Methodology

2.8.1 Design Approach

Participatory design [47] is an approach to design that involves relevant parties such as end-users and experts in the design process. It typically comprises collaborative

ID	Functional Guideline	Justification
FG_1	Flashcards should be scheduled to be reviewed later using a spacing algorithm.	Maintain spaced repetition benefits; multi-use learning resource
FG_2	The flashcard should link back to the part of the RLV where it was encountered.	Present RLV as a multi-use learning resource; preserve context
FG_3	The interaction with the flashcard should be enforced for the student when watching the video.	Present RLV as active learning resource

Table 2.3: Functional guidelines for our tool and their justification

activities, such as prototyping and interviews, where users can share insights on the design. These activities help designers to better understand their needs, and the user feedback is incorporated into the design process.

We use this design approach for designing our tool since meeting students' needs is highly prioritised; this is reflected in the research questions we posed.

2.8.2 Development Stages

With our design approach in mind, we have the following development stages for our tool:

1. Definition of the relevant set of problems and their domain
   The goal of the background chapter (Chapter 2) is to research existing problems with the learning process of students when replacing in-person lectures with RLVs, and how these problems can be solved. We define an approach to solving these issues but further investigate problems with that approach (spaced repetition and video; integrability). Based on the findings, we form an initial set of design and functional guidelines. In Chapter 3, a low-fidelity prototype is constructed and evaluated to incorporate user feedback into the design process. This further defines the set of problems and their domain. This stage addresses RQ_2 and RQ_3.

2. Design of a low-fidelity prototype and formative evaluation
   Based on our findings from the literature review and the resulting guidelines, we design a low-fidelity prototype in Chapter 3. This is followed by a formative evaluation with a Human-Computer Interaction (HCI) expert and a group of students. For the evaluation, we used a think-aloud protocol [20] approach as it is particularly suitable for usability testing. The feedback from this evaluation is considered for the final design of the tool. This stage addresses RQ_2.

3. Implementation of a high-fidelity prototype
   In Chapter 4, we implement a high-fidelity prototype based on the low-fidelity prototype and the updated set of design requirements from the formative evaluation. This addresses RQ_2. During the implementation process, we review which technologies we should use based on the integrability problem and the approach

we outline in Chapter 2. This addresses RQ_3.

4. High-fidelity prototype evaluation
   During the final stage of the development process, in Chapter 5, we evaluate the high-fidelity prototype with a convenience sample [17] from the tool's target audience using a task-based evaluation method. Feedback is gathered by conducting semi-structured interviews. The results are interpreted using bottom-up reflexive thematic analysis [7]. These steps aim to answer RQ_1. We also validate whether we met the relevant design requirements by running a final usability study (RQ_2).

While commonly observed for projects of a similar nature, the separation of the development process into low-fidelity and high-fidelity prototyping stages is specifically inspired by the methodology employed by Redbond [44]. From [44], we also adopted the terminology of a 'high-fidelity' and a 'low-fidelity' prototype, as well as the overall structure of this report. A discussion of considered alternatives for the methods we used is included in the report where appropriate.

Chapter 3

Low-Fidelity Prototype Design

At this stage of the project, we chose to design a low-fidelity prototype (LFP) based on the design guidelines established in Chapter 2. The main purpose of the LFP is to identify any missing requirements that could significantly compromise the evaluation of the high-fidelity prototype (HFP). The LFP also serves the purpose of guiding the implementation of the HFP.

This chapter addresses the second research question: '*How should our tool be designed to integrate spaced repetition and flashcards with RLVs to meet students' needs?*'

3.1 Overview of the Design

Figure 3.1 shows an overview of the design of the LFP that will serve as a guide to the reader throughout the rest of this chapter. The diagram depicts the two different aspects of the LFP: first, the static user interface design mock-up, and second, the added interactivity necessary for a more meaningful usability evaluation and requirements analysis.

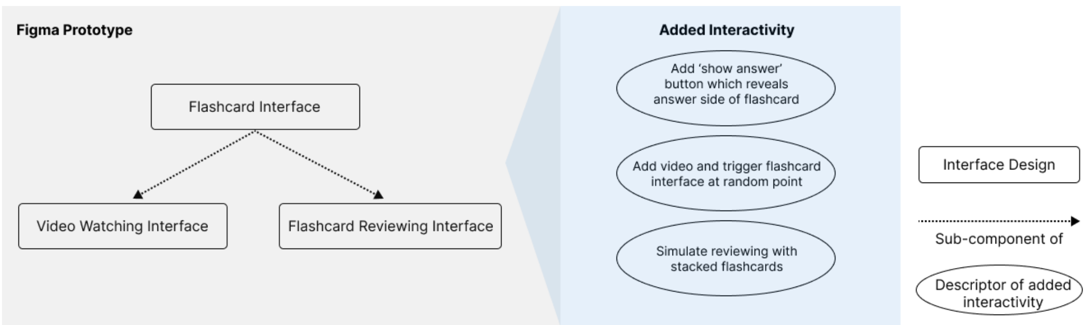


Figure 3.1: Overview of the low-fidelity prototype

3.2 Description of the Design

The design process of the LFP was split into two parts. For the first part, we developed a prototype of the user interface using Figma [18]. This is outlined in the left part of Figure 3.1. Figma was particularly suitable for our purposes due to its low learning curve and the possibility to create user interfaces without being required to write any code. This first part was based on the design guidelines and the envisioned use-case scenario established during the literature review in Chapter 2.

However, for a meaningful evaluation, it was critical to include the following user interactions, which could not be prototyped using Figma only:

1. *Interacting with a flashcard:* The user interacts with the flashcard and reveals the answer by pressing the corresponding button.
2. *Interacting with a flashcard while watching the RLV:* The user watches the RLV and encounters a flashcard. They interact with the flashcard as described above.
3. *Reviewing flashcards:* The user uses the review interface to interact with multiple flashcards, one after another.

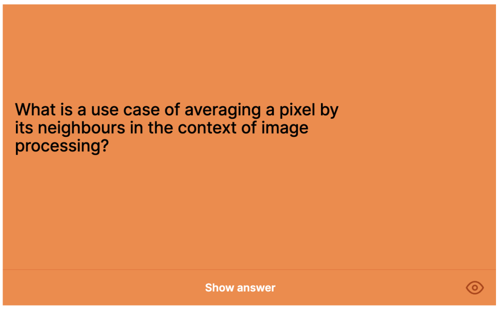
For the second part of the design process, we augmented the prototype with custom code to allow for those interactions. This is outlined in the right part of Figure 3.1. We achieved this by using images of the design mock-up in a simple web application. Using images rather than custom-coded interfaces simplified the process considerably, allowing for flexible adaptation of the LFP to changes in the interface requirements.

Overall, the LFP can be broken down into three parts: an interactive flashcard interface, an interactive video watching interface, and an interactive flashcard reviewing interface. What follows is a detailed report of each part of the LFP.

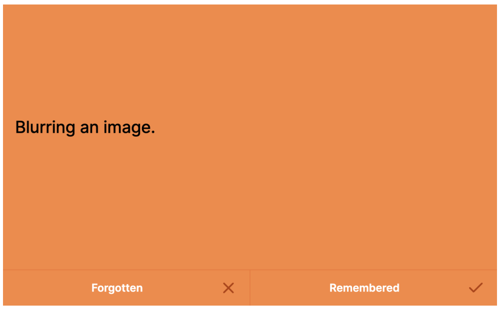
3.2.0.1 Flashcard Interface

The flashcard interface prompts the user with a question and presents the answer when the ‘Show answer’ button (Figure 3.2 a)) is pressed. Before revealing the answer, the user is tasked to try and think of it in their mind. After revealing the answer, they are asked to provide feedback about whether they managed to successfully recall it (‘Remembered’ button, see Figure 3.2 b)). If they did not recall the answer, the user should press the ‘Forgotten’ button.

The design of the flashcard interface was inspired by the flashcard application we reviewed in Chapter 2, Anki [1]. We strove for a clean and simple presentation of the flashcard contents (DG_2). This was achieved by surrounding both the question and the answer text with enough space, and by giving the text on the interactive buttons a different colour. Icons were added to the buttons to encourage user interaction with the flashcard (DG_3), and to make the usability of the flashcard more intuitive (DG_4). Finally, we chose a background colour which attracts the attention of the user (DG_1) without distracting from the contents (DG_2), based on the study by Camgöz et al [8].



(a) Front page showing the question



(b) Back page showing the answer and the feedback buttons

Figure 3.2: Design drafts for the flashcard interface of the low-fidelity prototype

3.2.0.2 Video Watching Interface

The video watching interface serves the purpose of presenting the RLV and prompting the user with flashcards as they watch the video. For the LFP, the functionality is limited to showing a single flashcard 30 seconds into watching the video.

The video watching interface of the LFP serves as a guiding example only as our tool can be embedded and used on other websites with video. When embedded, only the flashcard interface will be visible on the external website. This is further explained in Chapter 4.

The video watching interface is presented on a single page. It includes a video title, a video player, and a flashcard area. The video player is a native HTML media player [50]. Below the media player, the flashcard area displays a placeholder where flashcards will be presented as the user watches the video. It instructs the user to watch the video to reveal flashcards to interact with (see Figure 3.3 a)).

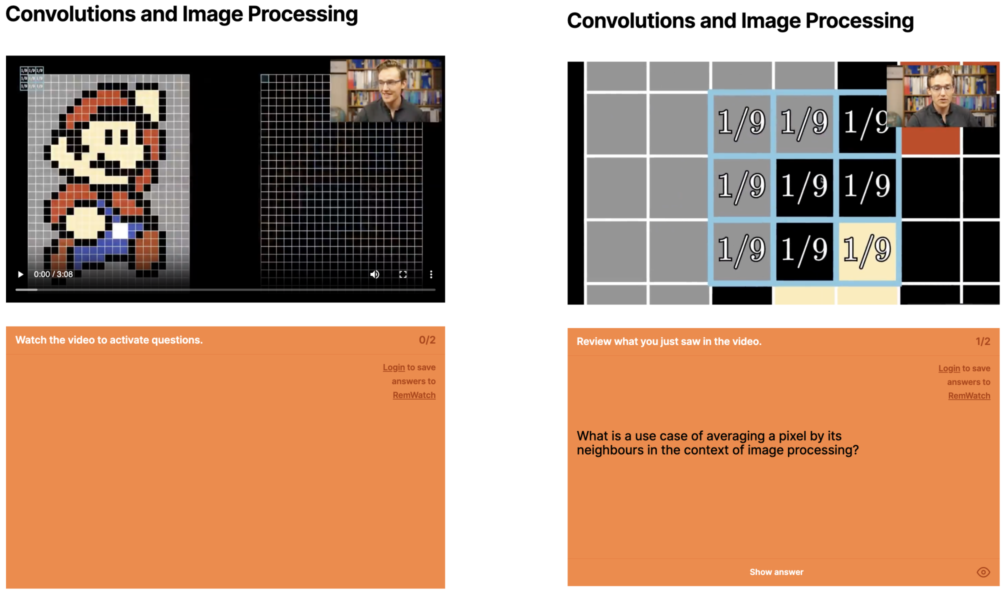
When the user encounters a flashcard trigger point in the video, the corresponding flashcard is presented in the flashcard area, and the user is tasked to further interact with it as described in the previous section (see Figure 3.3 b)).

Including explicit instructions in the flashcard area corresponds to DG_1 and DG_4: the aim is to motivate the user to start watching the video to reveal flashcards, and then to let the user interact with them. Furthermore, it explains how to use the tool.

We also integrated a counter of how many flashcards have been encountered so far, and how many have yet to be "discovered" when watching the video. This corresponds to DG_3: the intention is to motivate the user to keep discovering new flashcards later in the video. We are aware that this might lead to the opposite effect where the user sees that they already encountered all flashcards and their motivation to finish the video decreases. However, we think that this concern is negligible as it can easily be circumvented by embedding flashcards towards the end of a video.

3.2.0.3 Reviewing Interface

The reviewing interface is encountered when the user progresses to reviewing the flashcards they previously encountered via the video watching interface. This will be a



(a) The interface when showing the placeholder, with video title and user instructions

(b) The interface when showing a flashcard, with updated instructions

Figure 3.3: Video watching interface of the low-fidelity prototype

few days after watching the video, depending on the spacing algorithm we decide to implement for the HFP.

For the LFP, the interface consists of a title and the flashcard interface as a subcomponent of the design (see Figure 3.4). We added a counter of how many flashcards are left to be reviewed. This should motivate the user to review all the flashcards (DG_3).

The interface also features a link which brings the user back to the part of the lecture where the current flashcard was encountered ('Jump back to explanation'). While we included this in the design, its functionality was not implemented until we developed the HFP and could not be evaluated for the LFP.

3.3 Evaluation with Experts and Peers

At this point, we carried out a formative evaluation of the LFP with experts and peers whose expertise was expected to be valuable for identifying additional requirements for the design.

3.3.1 Participants

The details of each participant for the evaluation can be found in Table 3.1. This specific convenience sample [17] was found to be the right choice for this evaluation due to the combined expertise of the participants in the field of HCI, education, and interactive videos.

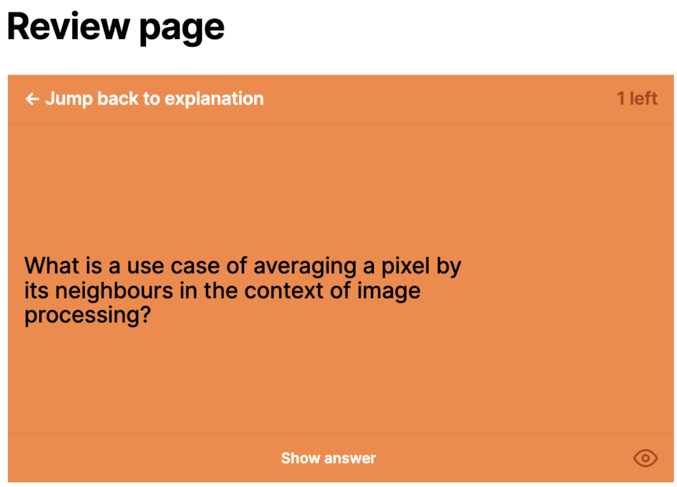


Figure 3.4: Flashcard reviewing interface of the low-fidelity prototype

Participant	Status	Area of Expertise
E1	Expert (academic researcher)	HCI, interactive streaming videos for education
S1	Undergraduate student	Informatics studies, currently also conducting research in the domain of interactive videos for education
S2	Undergraduate student	Informatics studies, currently also conducting research in the domain of interactive videos for education
S3	Undergraduate student	Informatics studies, currently also conducting research in the domain of interactive videos for education

Table 3.1: Details of the low-fidelity prototype evaluation participants

3.3.1.1 Materials

We used the following materials during the evaluation: low-fidelity prototype with added interactivity; a study instruction sheet; laptop (password-protected); video conferencing application; note-taking application.

3.3.1.2 Procedure

We ran two iterations of the study, first with the academic researcher (E1), and then with the group of student participants. The study was structured into three stages:

1. The participant or group of participants was given a short presentation about the general idea of this project and its goals. The presentation was supported by the interface design drafts created for the LFP.
2. We presented our envisioned use-case scenario from Section 2.6.2. For this step too, we used images of the user interface developed for the LFP to facilitate the participants' understanding of our tool.
    a) For the evaluation with the academic researcher, we gave them the instruction sheet and let them interact with the LFP themselves.
    b) For the evaluation with the student group, we demonstrated the instruction sheet steps whilst sharing our screen. This was due to the meeting being remote and the prototype only being available locally on our machine.
3. As the participant(s) interacted with the prototype, they were encouraged to share their thoughts and opinions at the same time. We discussed suggestions and additional requirements as they were following the instructions.

The evaluation with E1 was split into multiple sessions as they interacted both with an early-stage version of the LFP as well as with the version described in this chapter. The group evaluation was conducted within a single session. For both cases, the evaluation strictly followed the procedure we just described.

The evaluation method followed the approach of a think-aloud protocol [20]. This method was suitable as we were interested in participants' thoughts on usability as they interacted with the prototype or witnessed those interactions. Alternatives, such as a debrief-protocol [21] where feedback is gathered in retrospect, were considered as well but found to be less suitable due to the research goal of this study.

3.3.1.3 Data Collection and Analysis

Both the data collection and analysis aspects of this formative evaluation were kept informal. During the requirements discussion, we took notes of what participants said to collect feedback data. While this method of collection carries the risk of missing data items (compared to recording a session, for instance), we found it suitable for our purposes because we had the opportunity to clarify unresolved questions and revisit missed points.

The analysis process involved manually inspecting the notes and identifying additional requirements by considering each proposed requirement and evaluating whether it was crucial to add to the HFP. This was done by acknowledging the problems and initial set of guidelines presented in Chapter 2.

More formal analysis methods, such as full thematic analysis, were considered but found unsuitable for this kind of formative feedback data, as we were only interested in one theme at this point in the project: missing requirements.

3.3.2 Results

This section presents the results obtained during the formative evaluation of the LFP.

For the evaluation with the group, we noticed that we did not gather any feedback from S3. While we were still able to extract useful feedback data during that session, this might indicate that the think-aloud protocol was not the best choice for a group setup.

Reviewing Procedure
The idea of reviewing flashcards was overall well-received when it was presented using the LFP. S1 remarked: '*I think this is a good idea, I could imagine wanting to use this*.' The other participants were also convinced by the general concept; S2 commented that they could generally see the procedure of reviewing flashcards as beneficial for the learning outcome.

When asked to review flashcards, E1 wondered, '*How do I know when to review flashcards*?' Along these lines, S2 suggested: '*I would probably need a notification system to remind me when I need to review flashcards again*.' S1 said that they '*would definitely need a notification system*'.

Usability Issues
For this category, we received limited feedback from the student group because we were unable to let the student participants interact with the LFP itself during the study.

During the evaluation with E1, the overall impression of the usability was mixed. While the participant was able to follow through with the instruction sheet, they had to ask for clarifications for some of the steps. When asked whether the LFP was intuitive to use and the navigation self-explanatory, E1 responded that they were '*a bit confused on how to use the system*'. With '*how can I answer the flashcard?*' and '*I do not understand where I can write my answer*', E1 gave feedback on the usability of the flashcard, suggesting that the intended use of flashcards in the LFP was not intuitive enough. Further usability issues occurred when E1 was asked to review encountered flashcards: '*It is confusing to me what the review platform is and what the video watching page is, those two things look very similar.*'

E1 concluded their comment on the usability with the following suggestion: '*Maybe some documentation explaining your application and how to use it would be useful, that would clear it up for me*'.

Identified Flaws
E1 identified a flaw when they interacted with the LFP: '*I don't think the count of the*

remaining flashcards in the video is properly working, which leads to confusion.' This refers to the count element of the video watching interface, shown in Figure 3.3.

Further Suggestions
E1 made further suggestions to improve the LFP. They commented on using the LFP from the perspective of an instructor: '*As an instructor, how do I add flashcards to my video? I think this should be uncomplicated and even possible for people without a technical background.*'

When asked how they felt about the idea of showing the question on the answer page, all participants responded positively.

3.3.3 Impact on Design

Table 3.2 presents the additional requirements we identified during the study. They were extracted from the feedback notes we took during the evaluation, in combination with the write-up of the results presented in the previous section.

For each suggested requirement, we evaluated whether it should be directly implemented ('implemented'), implemented in parts ('partially implemented') or planned to be implemented in the future ('future work'). This was based on several factors, including the estimated priority of the suggestion, by whom it was suggested and in what context, how much it is estimated to affect the evaluation of the final prototype, and the given time limit for this project in combination with the estimated effort to implement the requirement.

No	Suggestion	Priority	Decision	Suggested by
1	Provide a user guide or documentation on how to use this application.	High	Implemented	E1
2	Fix the count of how many flashcards are remaining the video.	High	Implemented	E1
3	Display the question on the answer page of the flashcard.	Medium	Implemented	E1, S1, S2
4	Add introductory text elements to the video watching interface.	Medium	Implemented	E1
5	Clearly distinguish between the review and the video watching parts of the prototype.	High	Partially implemented	E1
6	Add a way of notifying the user when a flashcard is due to be reviewed.	Medium	Future work	E1, S1, S2
7	Make it easy for instructors without a technical background to add flashcards to their videos.	Medium	Future work	E1

Table 3.2: Suggestions and additional requirements identified during the formative evaluation of the low-fidelity prototype

Chapter 4

High-Fidelity Prototype Implementation

At this stage of the project, the goal is to implement an HFP based on the design prototype from the previous chapter, and on the results of its evaluation. The main purpose of the HFP is to be a tool which can later be used for evaluating the learning experience of students. After summarising the technologies used for its implementation, the chapter progresses to a description of the implemented tool. We also report the testing process we employed to facilitate and validate the prototype implementation, followed by a brief comment on a key implementation challenge.

This chapter addresses two of our research questions: *'What technologies should we use to implement our tool so that it can be integrated with existing software?'* (RQ_3), and *'How should our tool be designed to integrate spaced repetition and flashcards with RLVs to meet students' needs?'* (RQ_2).

4.1 Platform and Resources

For the tool development, we chose the resources and platform listed below based on their ability to implement the LFP design, the integrability problem, and our previous experience with using them for software development.

Platform
The tool is a web application which can be used in any modern browser. We chose to develop a web application rather than OS-native software (e.g., an Android application) to reach a large target audience whilst avoiding cross-platform development challenges.

Next.js Front-End
Next.js [37] is a React-based [43] framework for building web applications. As we had identified three different main components of the tool during the LFP design (flashcard, review, and video watching interface), this framework was particularly suitable as it allowed for a component-based development approach. Furthermore, we had previous experience with this framework, which proved to be helpful considering the given time limit. We used the framework for file-based routing of the web application, and to expose API endpoints for the web component.

Web Components
Web components [11] are a standard component model for the web that enables encapsulation and reusability of customized HTML elements. We used web components for our tool because they are compatible with native web technologies like HTML and JavaScript, and offer a solution to the integrability problem (Section 2.4.1). Alternatives like React and other JavaScript frameworks also offer component development for encapsulating and reusing functionalities, but they are not backwards-compatible with native HTML-based websites. In contrast, web components are forward-compatible and can be used with any platform built with these frameworks.

Vercel Hosting
For the evaluation of the HFP, we had to make the prototype accessible via a public domain, as opposed to running the prototype locally on our machine. To this end, we used Vercel [49] as a hosting provider. Vercel is suitable for this project since it offers hosting and domain names at a free tier whilst providing production quality. This was important as there was no monetary budget for this project.

Cloud Firestore Database Hosting
Cloud Firestore [19] provides hosting for NO-SQL databases. We selected this database hosting provider as there is a well-developed SDK which facilitates connecting to the database via HTTP endpoints.

OAuth Authentication
OAuth [2] is used as a third-party authentication provider to enable and manage user authentication processes. OAuth was chosen due to its compatibility with Next.js, and the ability to protect specific API routes within the web application with an authentication layer. This allowed us to securely manage reading and writing access to the database.

4.2 Description of the Implemented High-Fidelity Prototype

For the implementation of the HFP, we considered the design and functional guidelines presented in Chapter 2. We also recognised the additional requirements we identified during the evaluation of the LFP. The design and functionality of the LFP were maintained and only altered where necessary. Additionally, the HFP provides the following:

- An integrable web component which can be used by developers to make their videos interactive.
- An interactive flashcard component which dynamically retrieves and renders flashcard content from our database.
- A template which serves as an example and guidance for how to add interactivity to videos with the web component. This corresponds to the video watching interface of the LFP.
- A spacing algorithm to calculate user-specific review dates for flashcards, optionally based on previous review data.

- A review page where users can create a user account, log in, and review flashcards that they previously encountered on websites with videos made interactive with the web component.

4.2.1 Integrable Web Component

The HFP provides a web component which can be used to add interactive video functionality to external websites. Once added to a website, the web component communicates with our tool via exposed HTTP endpoints. This allows for retrieving necessary flashcard data, displaying that data in the flashcard component on the external website, and returning the feedback a user gave upon reviewing the flashcard ('remembered' or 'forgotten') to our tool. The returned feedback will be used as input for the spacing algorithm.

The web component provides further functionality to the external website's video player: it automatically pauses and resumes the video when a flashcard is encountered; it scrolls the flashcard into the view when it needs to be reviewed; and it removes the user's ability to continue the video manually until the flashcard is answered. These implementation details are based on DG_4 and RQ_2, aiming to make the tool easier to use and to enforce increased engagement. This is further discussed in Section 4.5.

4.2.2 Spacing Algorithm

The HFP uses the following spacing algorithm for calculating the review dates for flashcards:

$$N(C,L,A) = \begin{cases} C+1 & \text{if } L = \varnothing \text{ and } A = 1 \\ 2 \times (C-L) & \text{if } A = 1 \\ C & \text{otherwise} \end{cases}$$

where N is the next due date for the flashcard review, C is the current due date, L is the last due date of the specific flashcard, and A is the user feedback for a flashcard (1 for 'remembered' and 0 for 'forgotten').

The algorithm doubles the interval between review dates if the student reviews a flashcard successfully in a row. If it is a first review, the flashcard is scheduled for the next day. If they answered incorrectly, the algorithm resets the review history and schedules the flashcard to be reviewed on the same day again.

This is a variation of the Leitner-schedule [45] and conforms to the basic concept of spaced repetition [15]. While there are more sophisticated algorithms (e.g. Anki's scheduler [16]), our algorithm suffices for prototyping and proof of concept purposes.

4.2.3 Flashcard Component

The flashcard component of the HFP functions similarly to the design mock-up developed for the LFP: it presents the user with a question, reveals the answer upon the user's

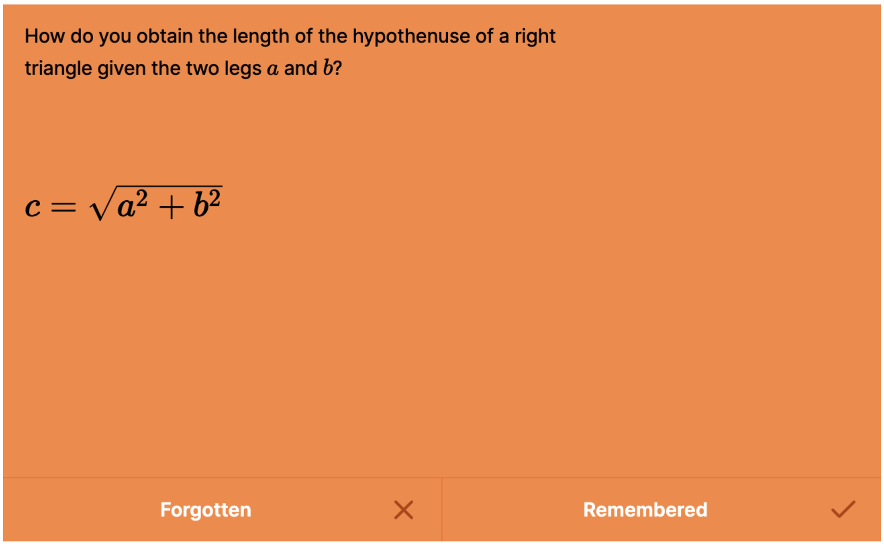


Figure 4.1: Updated flashcard answer page for the high-fidelity prototype

pressing the corresponding button, and finally asks for feedback on how the reviewing went.

Based on suggestion 3 from the evaluation results of the LFP (Table 3.2), the flashcard component displays the question above the answer of the flashcard, as reference for the user when they reveal the answer. This is shown in Figure 4.1.

The component can display text, optionally formatted with Markdown [33]. It is also capable of rendering mathematical expressions. The support of displaying further types of media (e.g., images, audio) is discussed in the Future Work section of Chapter 6.

The user feedback to a flashcard is stored in the database, along with the review date.

4.2.4 Interactive Video Website Template

The HFP template closely follows the video watching interface that was developed for the LFP. Presented on a single page, it includes a video title, a video, and a flashcard area. As per suggestion 4 from the LFP evaluation results in Table 3.2, we added text areas which can be used to augment the video with complementary text about its contents (Figure 4.2).

4.2.5 Reviewing Page

Users encounter the reviewing page of the HFP when they review the flashcards that they previously encountered via the video watching interface. The reviewing page uses the spacing algorithm to determine which flashcards are due to be reviewed based on the data in the database and presents them to the user for review.

For the HFP, the design of the reviewing page is very similar to the LFP. Following suggestion 5 from the LFP evaluation results, we added a brief paragraph which clearly states the purpose of the reviewing page, and differentiates it from any video watching

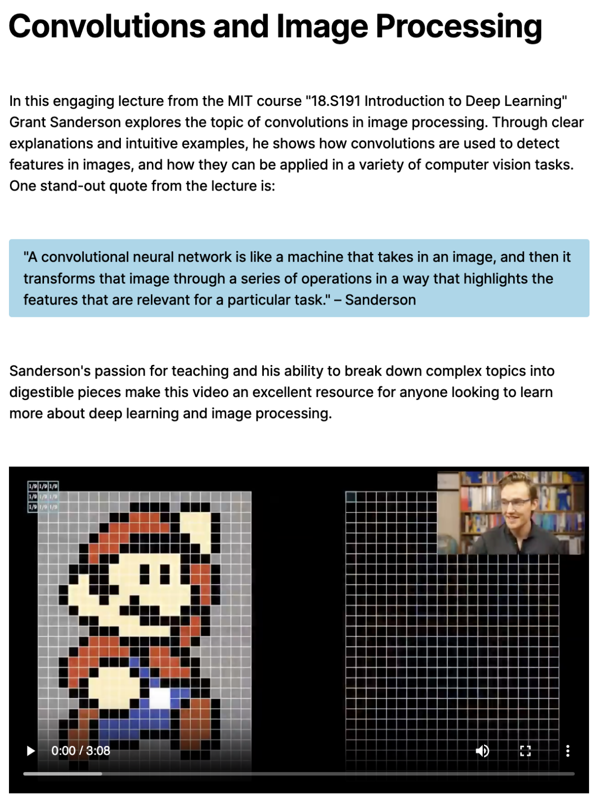


Figure 4.2: Updated interactive video template for the high-fidelity prototype, with additional context notes about the video. Note that the flashcard component is cut off in this image; it is still located below the media player.

websites. This also adheres to design guideline DG_4. This solution is somewhat crude (hence the ‘partially implemented‘ status in Table 3.2), but is expected to have the desired effect and can be further justified by the given time limit for this project.

4.3 Testing

The implementation of the HFP was supported by a thorough software testing process. Testing plays a crucial role in enhancing the reliability and validating the functionality of a tool before it is used in the real world. This is particularly significant when considering the consequences of a faulty system. In this context, we carried out testing in four consecutive steps, following the recommended process and utilising techniques from the literature [40]. The objective was to ensure that the tool had basic functionality working for prototype evaluations, as well as meeting the requirements for real-world deployment. By utilising a structured testing process, we were able to thoroughly validate the functionality of the tool.

4.3.1 Step 1: Initial Analysis

For this step, we first conducted a requirements analysis to identify aspects of our prototype which need testing. This requirements analysis was different to the analysis in Chapter 2 which yielded the design and functional guidelines as it focused on the goal of reliability and correctness of software. We addressed different levels of requirements (system, unit, integration) and established initial approaches for testing them. This involved translating qualitative requirements into measurable quantitative requirements. Finally, we assessed the appropriateness of the testing approaches.

The results of this step are a statement of requirements and an overview of testing approaches, both of which can be found in Appendix C.

4.3.2 Step 2: Design and Implementation of a Test Plan with Instrumented Code

For this step, we constructed a test plan with instrumented code:

- *Construction of a test plan:* we assigned priorities to the requirements that were identified during step 1. We then analysed any additional needs for testing the requirements thoroughly.
- *Assessment of the test plan:* we assessed the quality of the test plan in terms of three relevant types of risk which were described by Pezze et al [40]: personnel risk, technology risk, and scheduling risk. Where necessary, we proposed mitigation approaches.
- *Code instrumentation:* we identified tasks of the test plan which required additional code to carry out the testing, and developed the instrumentation.
- *Evaluation of the instrumentation:* we assessed the instrumentation code on whether it would be adequate to facilitate the testing process, and whether there were any potential improvements for it.

The results of this step are a comprehensive test plan consisting of a statement of testing needs for the requirements, a task and test specification, and a description of the implemented instrumentation code, along with its evaluation. These documents can be found in Appendix D and in the project materials.

4.3.3 Step 3: Application of Testing Techniques

For this step, we carried out the test plan from step 2 and applied a range of testing techniques, including End-To-End testing, penetration testing, and accessibility testing. We documented the results of the testing and assessed them on the following criteria, where applicable: accuracy, precision, repeatability, cost, execution time, and ease of use.

The results of this step are a report documenting the testing process and an evaluation statement of the testing results, to be found in the project materials provided with this report.

4.3.4 Step 4: Evaluation of the Limitations of the Testing Process and Code Inspection

For the final step of the testing process, we evaluated the limitations of the testing process by identifying gaps and omissions for each test task. We also inspected the code as a final means to identifying potential weaknesses in our prototype. For the inspection, we followed a classic checklist inspection approach [40].

Appendix E comprises the results of this step: a statement of deficiencies and omissions during the testing process, and an inspection documentation sheet.

4.3.5 Summary and Impact of the Testing Process

We conducted a thorough testing process to ensure increased reliability of our tool and to detect issues at an early stage. The testing results are thoroughly documented in the appendix of this report. Where necessary, we adapted the HFP to remove weaknesses and bugs that were exposed during the testing process.

4.4 Implementation Challenges

The implementation phase of a project is a critical step towards achieving its objectives, and it often comes with challenges. In this section, we examine one key implementation challenge encountered in the context of our research project. By focusing on this challenge, we aim to provide a detailed understanding of the complexity involved in implementing our proposed solution. This will help future researchers and practitioners to better anticipate and navigate similar implementation challenges in their own work. Including this implementation challenge in this report is essential because it allows us to reflect on the practical aspects of our research and its potential impact in the real world.

4.4.1 Cross-Origin-Resource-Sharing Restrictions

In addressing the integrability challenge discussed in the literature review, we made the decision to use a web component to encapsulate the interactive video functionality, allowing it to be integrated onto other websites. However, this decision led to some challenges. Specifically, when the component was integrated on external websites, it still needed to communicate with our prototype to retrieve flashcard contents, among other things. CORS (Cross-Origin Resource Sharing) [12] restrictions posed a significant obstacle to this communication. Allowing any cross-origin communication to access the HFP and its API to read and write from the database (by using the 'origin-all' option) would have been a bad idea, as it could have resulted in a DoS (Denial of Service) or other security attack. To overcome this challenge, we decided to embed parts of our front-end as an iframe [27] that handles the communication between the component and the prototype. This solution also had the added benefit of keeping users logged in across multiple tabs and different websites which use the web component of our tool.

4.5 Considerations for Evaluation

During the development of the LFP, we had several additional ideas about implementation details based on the initial guidelines presented in Table 2.2 and 2.3. However, we could not implement these ideas for the LFP without adding unnecessary complexity to the code, which would have hindered the evaluation of the LFP. As a result, these details could not be confirmed during the evaluation of the LFP. We will refer to these details as 'unconfirmed implementation details'. These details primarily focus on the web component, for two reasons. First, the web component was not a part of the LFP, and thus these details could not be confirmed during the LFP evaluation. Second, the web component is what ultimately provides the interactive functionality to the video. As a result, even small changes to the web component can have a significant impact on how users perceive and interact with the video.

No	Relevant HFP Component	Unconfirmed Implementation Detail	Motivated by
1	Web component	The web component stops the video when a flashcard is triggered.	FG_3, DG_4
2	Web component	The web component prevents the video from being continued when a flashcard is triggered.	FG_3, DG_4
3	Web component	The web component automatically resumes the video when a flashcard has been answered.	FG_3
4	Web component	The web component automatically scrolls to the flashcard interface when it needs to be reviewed, and back to the video after the flashcard has been answered.	FG_3, DG_4

Table 4.1: Unconfirmed implementation details of the high-fidelity prototype

Chapter 5

High-Fidelity Prototype Evaluation

At this stage of the project, we conducted two evaluation studies using the HFP. Firstly, a usability study which addresses the second research question: '*How should our tool be designed to integrate spaced repetition and flashcards with RLVs to meet students' needs?*' Secondly, a study about the learning experience with the HFP which addresses the first research question: '*How does our tool affect students' learning experience with RLVs?*'

The chapter first provides an overview of the participants selected for the studies, and then presents the usability study in Section 5.2 and the learning experience study in Section 5.3. We conclude this chapter with a discussion of our findings.

5.1 Participants

For both studies, we used the same group of participants. This simplified the administrative process of running the studies without being expected to have a considerable effect on the results. Table 5.1 provides an overview of the participants and their details.

Participant	**Background**	**Age**	**Previous Experience with Flashcards**	**Previous Experience with RLVs**
1	Undergraduate student, Biomedical sciences	22	Yes	Yes
2	Undergraduate student, Music	23	Yes	Yes
3	Graduate student, Biochemistry	24	Yes	Yes

Table 5.1: Details of the high-fidelity prototype evaluation participants

This specific convenience sample of participants was found to be suitable for the studies as they were all students from different academic backgrounds, and therefore a reasonable representation of the target demographic of our tool [17].

5.2 Usability Evaluation

The aim of the usability study was to validate that the HFP meets the design and functional guidelines that we established in Chapter 2. Although a similar evaluation was conducted for the LFP in Chapter 3, this was important to re-assess at this stage due to the alterations we made to both the functionality and the design of the prototype during the implementation of the HFP.

5.2.1 Materials

We used the following materials to conduct the study: high-fidelity prototype; participant's information sheet and consent form (Appendix H); user documentation for our tool (Appendix A); detailed study protocols; System-Usability-Scale (SUS) [6] feedback form (Appendix G); laptop (password-protected); audio recording device (smartphone, password-protected). The SUS form was slightly adapted to make the questions easier to understand in the context of our tool.

5.2.2 Procedure

The study was conducted in-person over three sessions, one with each participant. We audio-recorded each session. After each session, we copied the recording from the smartphone to the laptop to create a backup of the feedback data.

At the start of each session, we let the participant read the participant information sheet and fill out the consent form. Then, we asked about the background of each participant, to obtain Table 5.1. We then presented them with the detailed study information sheet and the user documentation to deepen their understanding of the concept of the tool, and how to use it.

Presenting a potential user with documentation beforehand follows suggestion 1 from the LFP evaluation (see Table 3.2).

At this point, we proceeded with a task-based evaluation, following each of the interaction protocols. During the evaluation, the participant was asked to rate the tool's usability using the SUS form. The SUS is an industry-standard survey that measures the usability of a system through ten questions assessed on a 5-point Likert scale [29]. Given our small sample size of participants, the SUS was deemed appropriate for our purposes. However, it should be noted that the SUS may not be suitable for all contexts and research questions, and its limitations should be carefully considered. The SUS evaluation focuses on the overall usability of a system and does not provide detailed information on specific usability issues [14]. However, this limitation is not relevant for us as our main objective was to obtain a general assessment of the tool's usability.

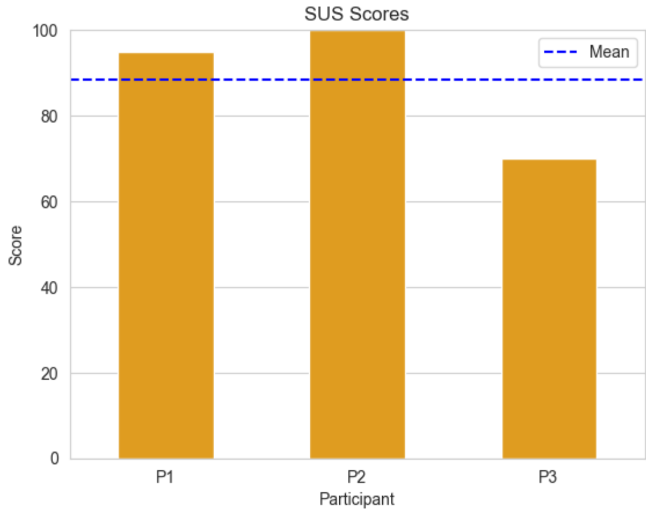


Figure 5.1: Calculated SUS scores from the participant feedback

5.2.3 Data Collection and Analysis

After the sessions had been completed, we transferred the answers submitted via the feedback form to a Jupyter Notebook [30] and evaluated them by calculating the SUS score for each participant.

5.2.4 Results

Overall, the participants were confident with using the tool and did not encounter major usability issues. All tasks stated in the protocols were successfully completed by each participant without the need of additional help.

System-Usability-Score
As shown in Figure 5.1, the HFP achieved a mean SUS score of 88.33. The lowest score was 70 (P3), and the highest score was 100 (P2). These results are discussed in the discussion section of this chapter.

5.3 Learning Experience Evaluation

The goal of this study was to determine how students respond to the tool in terms of their learning experience. We were interested in two aspects: first, the learning experience with the interactive RLV and integrated flashcards. Second, the learning experience with the flashcard reviewing process and spaced repetition scheduling.

5.3.1 Materials

The following materials were used for the study: high-fidelity prototype; example RLV (5min long, about image processing); detailed study guide, including an interview guide; laptop (password-protected); audio recording device (smartphone, password-protected). The choice of the topic of the RLV was found to cause issues due to the varied academic backgrounds of the participants; this was not expected as we estimated the content to be suitable for any high-school graduate. This is further discussed in the Section 5.4.

5.3.2 Procedure

This study was conducted with the same group of participants as the usability study. Since the previously-presented participant information sheet and its corresponding consent form also covered this study, we could skip this step. The study was carried out in person, with a separate session for each participant. We audio-recorded the sessions.

Each session was structured into four phases:

1. *Explanatory phase:* explanation of the idea of our tool and how to use it.
2. *Video watching interaction phase:* interaction with the interactive video.
3. *Flashcard reviewing interaction phase:* interaction with the reviewing page.
4. *Final feedback phase:* gathering of final feedback on the learning experience.

The explanatory phase could be omitted for two of the three participants as they still remembered the explanations from when they participated in the usability study.

During the phases, we used a debriefing rather than think-aloud method to gather feedback. This was to decrease the risk of the participants being distracted from the actual learning experience as they were interacting with the prototype. Except for the explanatory phase, each phase also entailed a semi-structured interview to allow for an open-ended discussion about the participant's learning experience.

During the evaluation, we purposefully left the definition of learning experience open-ended, only providing some pointers to give the participants an idea of what was meant. This was to account for aspects of a participant's learning experience that we could not anticipate and might have been omitted in a participant's answer given a too-rigid definition of the term.

5.3.3 Data Collection and Analysis

Like the usability study for the HFP, we transcribed the audio recordings using an automatic transcription tool. We manually corrected the parts of the transcripts that were relevant but wrongly transcribed by the transcription tool.

We carried out reflexive thematic analysis [7] to interpret the data. We chose a recursive bottom-up approach because we did not have a clear idea of what themes to expect, due to the open-question design of the semi-structured interviews. Similarly, for the coding process of the data we chose a deductive approach, as outlined in [7]. Initially, there was an unusually large number of codes, but a few of them could be merged or eliminated at a later stage of the analysis.

As guidance throughout the analysis, we created a mind map with an overview of established themes, sub-themes, and their relations (shown in Appendix B). The established top-level themes were: subject learning outcome, subjective learning productivity, impact of the interface, concerns, and external factors.

5.3.4 Results

Overall, all participants attributed a positive learning experience to using the tool, and could imagine using it in a real-world setting to increase their learning outcome when learning with RLVs.

Subjective Learning Outcome
This theme explores how much participants believed they were learning from the RLV when watching the video interactively with the prototype. This was subjective and not evaluated against a formal test of the participants' knowledge of the lecture. This theme also covers the perceived learning outcome when using the reviewing page of the tool.

Two out of three participants (P1, P2) experienced an increased subjective learning outcome. Both commented that the contents of the RLV addressed by the flashcards stood out more and were better remembered than the rest of the lecture. When asked whether this meant that things not covered by flashcards were harder to remember than they usually would have been, P2 responded that they '*maybe would have remembered other things but the overall memory of contents is definitely better* [when using the tool]'.

P1 stated that lecture contents were easier to remember when answering the inserted flashcards incorrectly ('*learning from mistakes*'), and they also felt that reviewing flashcards repeatedly increased the learning outcome. P1 concluded their feedback on their subjective learning outcome by comparing it to attending in-person lectures: '*It reminds me of live lectures where lecturers put in little quizzes or little summaries in between topics within a long 1-hour lecture. And I find that when they do that, I remember things better.*'

P2 perceived their learning outcome to be increased because of the review feature of the tool: '*Recalling what I've done in terms of flashcards or looking back to parts of the lecture to explore what I've forgotten, to then fill in smaller gaps . . . I feel I definitely learn more with lecture videos this way.*'

P3 reported an insufficient subjective learning outcome and commented: '*I felt like I still did not understand the lecture content.*' They also added that they were not confident with the mathematical complexity of the lecture content.

Subjective Learning Productivity
This theme explores how productive the learning process felt to the participants as they were interacting with the HFP.

All participants stated that they experienced increased motivation, attention, or engagement when watching the interactive video. Especially interacting with flashcards during the RLV was found to increase the focus on the lecture content, as reported by all participants. P2 summarised this experience by saying: *'To be presented with the flashcards made me focus because if I was* [. . .] *drifting off for a few seconds, I might miss something important and then I wouldn't be able to answer that flashcard. I would* *have to go back. I was compelled to focus.'* P3 also reported: *'If you get the flashcard wrong, I would definitely want to go back and rewatch rather than just keep on going.'*

All participants felt more motivated to try and understand the content of the RLV. P2 said: *'It definitely felt like it kept me engaged and* [...] [the tool] *made sure of that because if I had not been able to answer the flashcard, I would have gone back.'* Indeed, P1 reported a similar experience: '[There is] *less risk of just watching the lecture and* [...] *not understanding the content.'*

P3 commented on the general feeling of productive learning by saying that *'If I get it* [i.e., answer correctly], *then I feel more confident and then I feel like that boosts my learning experience.'*

P1 noticed a more productive learning process as they felt that they could omit taking notes whilst watching the RLV without a negative impact on their subjective learning outcome. In the end, they found: *'I find myself not having to take as many notes to explain a topic to myself in my notes.'*

The reviewing process of the flashcards after watching a lecture was generally met with positive feelings. Both P1 and P3 mentioned they can see reviewing flashcards with spaced repetition being a good way of learning with RLVs, *'without having to watch a video over and over again'* (P3). P2 mentioned they *'would need to see how learning with such a tool and reviewing is like over a longer period of time.'*

Impact of the Interface
This theme is about features of the user interface of the HFP, and how participants thought they affected the learning process. It consists of two sub-themes: positive effects on the learning experience, and negative effects on the learning experience.

P2 commented directly on how the interface positively affected the learning experience for them: *'The learning experience was quite visual and I think it's good to see the question, see the answer and making the connection between them.'* They also commented about the flashcard design: *'I like the colour orange. I think it's warm and it draws my attention.'*

P1 had mixed feelings about the interface. They said: *'I like the fact that* [the tool] *stops the video and then automatically displays the flashcard that corresponds to the information that I've just covered, in front of me.'* However, they also displayed confusion when interacting with the review page interface and then jumping back to the RLV: *'I think it was* [...] *confusing to have the option to answer the flashcard both on the review page as well as here* [on the video watching page].'

P3 found the interface overall sufficient for their learning experience but they said that they would prefer if clicking the 'Jump back' button on the review page would not link to the part of the RLV where the flashcard is triggered but rather a few moments earlier when the content of the flashcard was explained.

Concerns and Improvements
This theme consisted of three sub-themes: overall conceptual concerns, potential improvements for the tool, and additional requirements or needs from the student perspective.

One conceptual concern of using the tool for learning was brought up by P3 when they mentioned that the only things that they learned were the parts covered by the flashcards. They concluded: *'I can't really remember anything else he* [the lecturer] *said.'*

In terms of potential improvements, P2 mentioned they could imagine that it would be helpful to have additional supporting content located on the video watching page: *'On* [Microsoft] *Teams, you have access to files that have been shared and comments made throughout the lecture, which could be useful to have here as well.'* P3 provided a suggestion for the interface when jumping back to an RLV from the reviewing page: *'It would be nice if it had come up with both the flashcard and the video in full view.'* They also suggested improving the spacing algorithm by adding an additional input parameter, namely, whether parts of an RLV had to be rewatched to answer a flashcard.

Additional requirements included a request for notifications when flashcards are due to be reviewed (P1, P3) and an optional format for providing an answer to the flashcard question (P2). P2 provided the example of being able to write or draw an answer as they try to think of it, enabled by the interface, and then revealing the answer page of the flashcard.

External Factors
This theme is about mentioned external factors and how they affected the perceived learning experience of the participants.

All participants mentioned that their learning experience was most likely positively influenced by the teaching skills of the lecturer from the example RLV. Additionally, P2 commented that their learning experience might have been negatively impacted if the content covered by the flashcards had not been as straightforward.

Implementation Considerations
At the end of Chapter 4, we mentioned certain implementation details of the tool which should be considered during the evaluation of the HFP (Table 4.1).

Detail 1 (automatic stopping) did not receive any negative or positive feedback from any of the participants. Detail 2 (forced answering) was positively received (3/3 participants), with the same argument across the board that they would otherwise feel tempted to skip a question, and that it increased the attention. Detail 3 (automatic resuming) did not get any negative feedback but was also not otherwise mentioned in the feedback. Detail 4 (automatic scrolling) received positive feedback from all participants, but P2 suggested to leave some space at the top of the video player when scrolling back to it after having answered the flashcard, so that they would not need to fear about parts of the video being cut off at the top.

5.4 Discussion

We ran two different types of studies to evaluate the HFP: one to assess usability, and one to assess the learning experience. The participants did not encounter any major usability issues and were able to complete all tasks from the usability evaluation without any additional help. For the learning experience evaluation, the consensus was that the tool improved learning with RLVs.

The SUS scores indicate good usability of the HFP: two scores were above 85 which corresponds to excellent usability levels [3]. The third score was 70, which indicates good usability with few issues [3]. The mean score of 88.33 is above the average for educational technology [52]. This might have to do with the fact that the HFP has a low number of elements to interact with, and is, hence, of decreased complexity in comparison to other educational tools.

We identified that the perceived learning experience of participants consisted of two crucial elements: the subjective learning outcome, i.e., how much of the RLV content participants thought that they learned, and the subjective learning productivity, i.e., how effective the process of learning the lecture content felt. The fact that all participants commented on their perceived learning outcome without directly being asked any questions on this topic suggests that the feeling of how much students learn affects their learning experience, a result which correlates with previous research [26].

Two out of three participants (P1, P2) experienced a feeling of increased learning outcome during the evaluation. Only P3 felt that their learning outcome was insufficient; their unease with the mathematical complexity of the example RLV suggests that this might have to do more with the choice of RLV than with the HFP itself. This would also address the conceptual concern presented in the results.

All participants reported an increased feeling of learning productivity and increased attention compared to their previous experience when learning with RLVs. This result might be skewed since we were only able to let participants watch and interact with a 5min-long RLV, rather than an RLV of typical length. This is a potential compounding factor which we were unable to circumvent during the design of the study (e.g., by using a longer RLV) due to administrative limitations and the availability of the participants. We identified that the presenting skills of the lecturer and their teaching style could also be a compounding factor.

The participants made several suggestions for improving the tool, including a more sophisticated spacing algorithm, and further interface improvements. We think that all the suggestions are valid and should be addressed in future versions of the prototype.

None of the implementation details from Table 4.1 received negative feedback, and a few were explicitly mentioned in a positive context. This suggests that these details can be maintained for future versions of the tool.

To conclude, the HFP was found to be sufficiently usable and to improve students' learning experience when learning with RLVs. This corresponds to the projected results from Chapter 2, except for the improved objective learning outcome for which we did not obtain any results. We also identified that the results were subject to several compounding factors (choice of RLV content and its complexity, presenting skills of the lecturer and their teaching style, the length of the RLV, etc.), some of which were classified as external, i.e., not possible to circumvent using our evaluation methods.

Chapter 6

Discussion, Conclusion, and Future Work

This project investigated how we can build a spaced repetition tool for recorded lecture videos that suits the needs of students. For this purpose, we posed three research questions which were addressed at various stages of the development of the tool. In this chapter, we reflect on the research questions, discuss the limitations we encountered, and provide an outlook on possible future work. The chapter ends with a conclusion.

6.1 Reflecting on the Research Questions

1. **How does our tool affect students' learning experience with RLVs?**
   This question was addressed during the evaluation of the HFP with three participants who were sampled from the target audience of the tool (Chapter 5).

   Thematic analysis of the evaluation feedback data yielded results on the following aspects of the learning experience of the participants: subjective learning outcome, subjective learning productivity, impact of the interface, concerns, and external factors.

   The subjective learning outcome was reported to be improved, in comparison to passively learning from RLVs (two out of three participants). There were reports of an increased ability to remember the content covered by flashcards that the participants interacted with. One participant reported an insufficient subjective learning outcome which we attributed to the specific choice of RLV and the mathematical complexity of its content, rather than being a consequence of using the tool. All participants reported an increased feeling of learning productivity compared to their previous experience when learning with RLVs.

   The user interface of the tool was found to be beneficial for the learning experience of the participants; they made a few suggestions how it could be improved, e.g. by changing where the 'Jump back' button on the review page leads to. Further improvements where participants thought that they would improve their learning experience included supporting content for the RLV on the video watching page, a more sophisticated spacing algorithm, and a notification system for flashcards due to be reviewed. Unconfirmed implementation details from Table 4.1 did not affect the learning experience negatively; some of them were attributed to a positive learning experience.

Overall, we found that the learning experience of students when using the tool was perceived improved compared to learning with RLVs without the tool. This correlates with the projected results from Chapter 2. However, the learning experience evaluation was subject to a number of compounding factors, as mentioned in the discussion from Chapter 5. Thus, any final conclusions drawn from our results should be applied with caution.

2. **How should our tool be designed to integrate spaced repetition and flashcards with RLVs to meet students' needs?**
This question was addressed by reviewing the relevant literature during the background chapter (Chapter 2), during the design of the LFP and its formative evaluation (Chapter 3), and during the final usability evaluation of the HFP (Chapter 5).

   In Chapter 2, we identified four problems relevant to the design of the tool: the problem of RLVs inherently being a passive learning resource (Section 2.2.1), of RLVs being a single-use learning resource (Section 2.2.2), of the seemingly incompatible nature of spaced repetition and RLVs (Section 2.3.1), and the integrability problem (Section 2.4.1).

   Based on these problems, we established an initial set of design guidelines (Table 2.2). We followed these guidelines closely during the design of the LFP (Chapter 3). Then, we conducted a formative evaluation of the LFP design with an HCI expert and three students who work on projects related to this one. The design of the LFP was generally well received during the evaluation and suitable for the goal of this project. We documented any further suggestions that came up during the evaluation (Table 3.2) and ranked them by priority. Finally, we adapted the LFP accordingly to meet the updated design requirements.

   We then developed an HFP (Chapter 4) based on the design of the LFP. We evaluated the design and usability of the HFP using the System-Usability-Scale with a convenience sample from the target audience (students or recent graduates with previous RLV learning experience). The HFP received two excellent scores and one good score, with a mean of 83.33. This indicates that the final prototype is a good representation of how technology should be designed around our project goal, with a mean usability score that is above the average of educational tools [52].

3. **What technologies should we use to implement our tool so that it can be integrated with existing software?**
This question was addressed during the background chapter (Chapter 2) and the implementation of the HFP (Chapter 4).

   a) **What existing technological infrastructure does the tool need to be integrated with?**

This sub-question was answered in Section 2.4.1 by reviewing preliminary results of an ongoing research project which investigates the differences in technological stack used for various OpenCourseWare platforms [51]. We found that there was no common demoninator, and the most plausible solution would be to integrate with native web technologies (HTML, CSS, JavaScript).

b) **What technologies are compatible with the required infrastructure?**
Based on our findings for 3a), we deduced in Section 4.1 that web components [11] would be a suitable type of technology to involve in the implementation of our tool due to the possibility of integrating it with any natively built website, and the forward-compatibility with various other frameworks.

c) **Based on the previous questions, what technologies should be used for implementing our tool?**
This sub-question was addressed in Section 4.1 of the HFP implementation chapter by considering the previous parts of this research question. Additionally to web components, we found that the following resources were suitable for implementing the tool: Next.js for front-end development, Vercel for hosting the tool, Cloud Firestore for hosting a database, and OAuth for authentication and user management. Web-applications as target platform were found to be suitable for our project due to their projected ability to reach a large part of the target audience, and due to avoiding cross-platform development issues this way.

6.2 Limitations

Due to the limited time available to carry out the project, we only developed an HFP of the tool. While we progressed with the development of the prototype as far as possible, there were requirements and suggestions that were not implemented in the end (flagged as 'partially implemented' and 'future work' in Table 3.2). The limited time also affected the choice of research questions and evaluation strategy. Insights into students' objective learning outcome achieved with the tool, rather than only into their learning experience, would have been a useful result to obtain. However, due to the nature of the spaced repetition learning technique, this would have likely needed to be a quantitative long-term study, such as seen for [32]. Such an evaluation would have been unfeasible given the time limit.

The limited time also prevented us from testing the integrability of our tool with the existing technological infrastructure in praxis. We had to rely on theory-based findings from the literature review for this aspect.

Lastly, the learning experience evaluation study was limited to showing a single RLV of a short length only. Given more time and participants with schedules that were less busy, we could have used RLVs of various length and about different topics to obviate the risk of compounding factors impacting the significance of our results. Furthermore, the study was limited to participants who had previous experience with both flashcards and learning with RLVs.

However, despite these limitations, we managed to show qualitatively that the tool we developed for this project has the potential to guide future work in this field.

6.3 Future Work

For the second year of this project, the plan is to approach the challenges and suggestions described below with similar methods as we employed this year. This especially refers to an iterative prototyping approach which is informed by evaluation results. A suggested timeline is presented in Appendix F.

The limitations of this project suggest several avenues for future work. First, we recommend implementing the suggestions flagged as 'partially implemented' or 'future work' in Table 3.2. Future versions of the tool could incorporate features that capture participants' perceived learning outcome, allowing for a more comprehensive understanding of the relationship between learning outcome and learning experience. As outlined in the description of the flashcard component in Chapter 4, the tool could also support other types of media to be rendered by flashcards, e.g. images and audio.

Second, we suggest analysing intructors' needs for using our tool in a real-world setting for teaching purposes. The HFP should be adapted accordingly.

Furthermore, we recommend conducting an additional evaluation study that measures learning outcome quantitatively, potentially by employing the tool for use by a university course for a semester. This would also test the integrability of the tool with existing technological infrastructure in praxis. It would also allow for a final learning experience study, with more room to obviate previously discussed compounding factors.

Finally, we suggest conceptual work in this field, exploring different ways of integrating spaced repetition and video beyond what was investigated in this project. Our tool could inspire other researchers to investigate novel ways of integrating these techniques to further enhance learning with videos.

6.4 Conclusion

This project presented a unique approach to improving the learning experience with RLVs through the use of a spaced repetition tool. We used a participatory design process to develop a design prototype that was evaluated by an HCI expert and peers. Based on the feedback received, a high-fidelity prototype was implemented and evaluated with students. In summary, the research contributions are the identification and development of a unique approach to integrating spaced repetition and RLVs, the design and formative evaluation of a low-fidelity prototype, the implementation of a high-fidelity prototype, and the usability and learning experience evaluation of the high-fidelity prototype. Our tool was found to be both usable and effective in enhancing the learning experience of students when learning with RLVs. As educational videos become increasingly prevalent in teaching, the significance of this research lies in providing a potential solution to improve learning with RLVs, and to guide future research in this field.

Bibliography

[1] Anki, https://apps.ankiweb.net/, March 2023.

[2] Auth0, https://auth0.com/, March 2023.

[3] Aaron Bangor. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3), 2009.

[4] Charles C. Bonwell. Enhancing the lecture: Revitalizing a traditional format. *New Directions for Teaching and Learning*, 1996(67):31–44, 1996.

[5] Cynthia J. Brame. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences Education*, 15(4):es6, December 2016.

[6] John Brooke. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 1996. Num Pages: 6.

[7] David Byrne. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56(3):1391–1412, June 2022.

[8] Nilgün Camgöz, Cengiz Yener, and Dilek Güvenç. Effects of hue, saturation, and brightness: Part 2: Attention. *Color Research & Application*, 29(1):20–28, February 2004.

[9] Andreas Chai. Web-enhanced procrastination? How online lecture recordings affect binge study and academic achievement. Technical report, Griffith University, 2014.

[10] Chih-Ming Chen and Chung-Hsin Wu. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80:108–121, January 2015.

[11] Web Components, https://developer.mozilla.org/en-US/docs/Web/Web_Components, March 2023.

[12] Cross-Origin Resource Sharing (CORS), https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS, March 2023.

[13] Coursera, https://www.coursera.org/, March 2023.

[14] Joseph S. Dumas and Jean E. Fox. Usability Testing: Current Pratice and Future Directions. In *The Human-Computer Interaction Handbook*. CRC Press, 2 edition, 2007. Num Pages: 22.

[15] Hermann Ebbinghaus. Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, 20(4), October 2013.

[16] Damien Elmes. What spaced repetition algorithm does Anki use? - Frequently Asked Questions, https://faqs.ankiweb.net/what-spaced-repetition-algorithm.html, March 2023.

[17] Ilker Etikan. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1):1, 2016.

[18] Figma, https://www.figma.com/, March 2023.

[19] Firebase, https://firebase.google.com/, March 2023.

[20] Marsha E. Fonteyn, Benjamin Kuipers, and Susan J. Grobe. A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, 3(4):430–441, November 1993.

[21] Roxane Gardner. Introduction to debriefing. *Seminars in Perinatology*, 37(3):166–174, June 2013.

[22] Duncan Godden and Alan Baddeley. When does context influence recognition memory? *British Journal of Psychology*, 71(1):99–104, February 1980.

[23] Neil Guppy, Dominique Verpoorten, David Boud, Lin Lin, Joanna Tai, and Silvia Bartolic. The post-COVID-19 future of digital learning in higher education: Views from educators, students, and other professionals in six countries. *British Journal of Educational Technology*, 53(6):1750–1765, November 2022.

[24] Vanessa Gysbers, Jill Johnston, Dale Hancock, and Gareth Denyer. Why do Students still Bother Coming to Lectures, When Everything is Available Online? *International Journal of Innovation in Science and Mathematics Education*, 2011.

[25] Mehmet Ali Gülpinar and Berrak Ç. Yeğen. Interactive lecturing for meaningful learning in large groups. *Medical Teacher*, 27(7):590–594, November 2005.

[26] Matti Haverila. Prior E-learning Experience and Perceived Learning Outcomes in an Undergraduate E-learning Course. *MERLOT Journal of Online Learning and Teaching*, 7(2), 2011.

[27] iframe, https://developer.mozilla.org/en-US/docs/Web/HTML/Element/iframe, April 2023.

[28] Learning Tools Interoperability, https://www.imsglobal.org/activity/learning-tools-interoperability, March 2023.

[29] Susan Jamieson. Likert scales: how to (ab)use them. *Medical Education*, 38(12):1217–1218, December 2004.

[30] Project Jupyter, https://jupyter.org, March 2023.

[31] Juho Kim, Elena L. Glassman, Andrés Monroy-Hernández, and Meredith Ringel Morris. RIMES: Embedding Interactive Multimedia Exercises in Lecture Videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1535–1544, Seoul Republic of Korea, April 2015. ACM.

[32] Douglas P Larsen, Andrew C Butler, and Henry L Roediger III. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical Education*, 43(12):1174–1181, December 2009.

[33] Markdown, https://daringfireball.net/projects/markdown/, March 2023.

[34] Andy Matuschak. How to write good prompts: using spaced repetition to create understanding, https://andymatuschak.org/prompts/, December 2020.

[35] Arthur W. Melton. The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 9(5):596–606, October 1970.

[36] Norbert Michel, John James Cater, and Otmar Varela. Active versus passive teaching styles: An empirical study of student learning outcomes. *Human Resource Development Quarterly*, 20(4):397–418, September 2009.

[37] Next.js, https://nextjs.org, March 2023.

[38] Michael Nielsen. Augmenting Long-term Memory, http://augmentingcognition.com/ltm.html, July 2018.

[39] Jungkook Park, Yeong Hoon Park, Jinhan Kim, Jeongmin Cha, Suin Kim, and Alice Oh. Elicast: embedding interactive exercises in instructional programming screencasts. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pages 1–10, London United Kingdom, June 2018. ACM.

[40] Mauro Pezzè and Michal Young. *Software testing and analysis: process, principles, and techniques*. Wiley, Hoboken, N.J., 2008. OCLC: ocn123408179.

[41] QuickTimePlayer, https://support.apple.com/en-gb/guide/quicktime-player/welcome/mac, March 2023.

[42] Susen Rabold, Stuart Anderson, John Lee, and Neil Mayo. YouTute: Online Social Networking for Vicarious Learning. In *Proceedings of ICL2008*, pages 1–14, Villach Austria, September 2008. Kassel University Press.

[43] React, https://react.dev/, March 2023.

[44] Aimee Redbond. *Designing and Developing a Technology-based Tool to Help Children with Autism Cope with Changes*. Bachelor's Thesis, University of Edinburgh, May 2021.

[45] Siddharth Reddy, Igor Labutov, Siddhartha Banerjee, and Thorsten Joachims. Unbounded Human Learning: Optimal Scheduling for Spaced Repetition. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1815–1824, San Francisco California USA, August 2016. ACM.

[46] Moira Sarsfield and John Conway. What can we learn from learning analytics? A case study based on an analysis of student use of video recordings. *Research in Learning Technology*, 26(0), December 2018.

[47] Douglas Schuler and Aki Namioka, editors. *Participatory Design: Principles and Practices*. CRC Press, 1 edition, December 2017.

[48] Microsoft Teams, https://www.microsoft.com/en-gb/microsoft-teams/group-chat-software, March 2023.

[49] Vercel, https://vercel.com/, March 2023.

[50] HTML video, https://developer.mozilla.org/en-US/docs/Web/HTML/Element/video, March 2023.

[51] Vidminas Vizgirda, Fiona McNeill, and Brian Mitchell. Practical Approaches to Freeing University Material Stockpiles. In *OER23 Conference: Advancing Open Education Practices*, Inverness Scotland, April 2023. Association for Learning Technology.

[52] Prokopia Vlachogianni and Nikolaos Tselios. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education*, 54(3):392–409, May 2022.

Appendix A

High-Fidelity Prototype – User Documentation

The purpose of this documentation is to be brief and cover most relevant aspects of our tool for participants of potential studies. As such, it aims to be informal, friendly, and easy to understand. It is not exhaustive. Some of the terminology may differ from what is used in the strictly academic context of our dissertation report. In the documentation and during the studies, we called our tool 'RemWatch'.

Concept of a flashcard
A flashcard is a card bearing information on both sides. The way it is often used is that it contains a question or prompt on its front and the answer on its back.

When you encounter a flashcard, the idea is that you try to think of the answer, mentally, before revealing the answer. This is called "answering" a flashcard. If it's not the first time you answer that flashcard, we call that "reviewing" it.

The digital version of a flashcard often has a button to reveal the content of its back side, shown in Figure A.1.

Flashcard system
A flashcard system tries to predict the best time to review a flashcard next. Ideally, this is the day you are predicted to forget the answer of a flashcard.

It uses feedback from you to determine this. Such feedback usually comes in the form of "were you able to answer the question? yes/no".

What is RemWatch?
RemWatch is a tool I built for my dissertation. It helps me to investigate certain research questions around how to improve learning with lecture recordings by making them interactive.

RemWatch does two things: 1. It integrates flashcards into lecture recordings. If you watch such a recording, you will have to answer any encountered flashcards. 2. It collects any encountered flashcard and shows it on its review page when it thinks it's the best time to review it again.

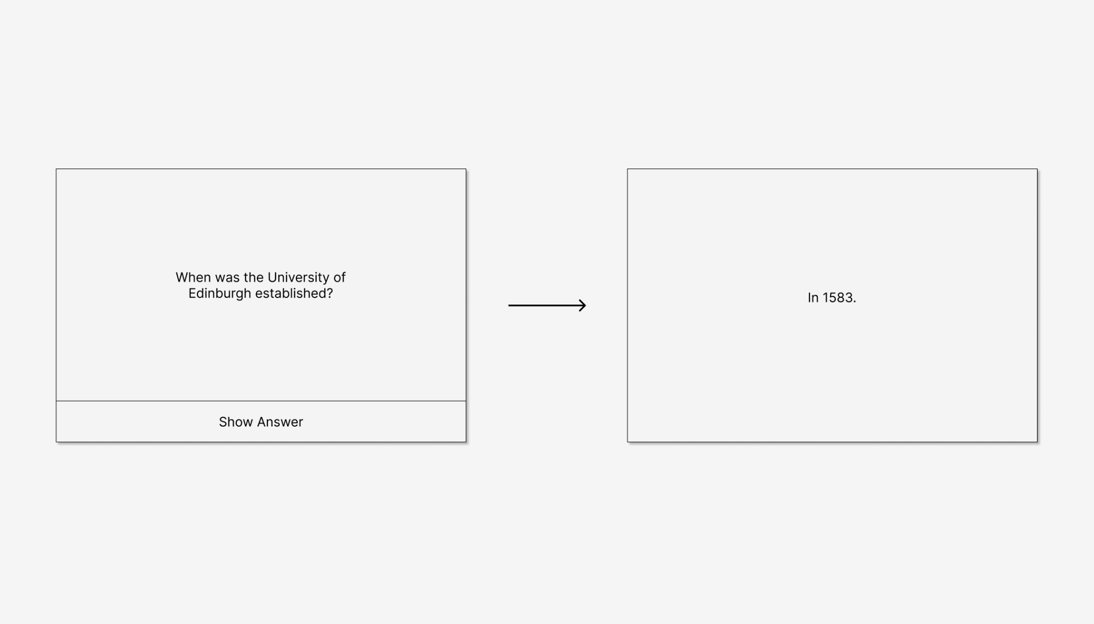


Figure A.1: Concept of a digital flashcard

How to use RemWatch?
You will first have to watch a lecture recording which has embedded flashcards. If you are a participant of a study, you will be provided with a link to such a recording.

When you encounter a flashcard, it will look like the screenshot in Figure A.2.

You will have to:

1. Read the question.
2. Try to think of an answer, in your mind.
3. Reveal the answer ("Show answer"), compare it to the one you thought of.
4. If you knew the answer, click "Remembered". Otherwise, click "Forgotten".

To review flashcards, go to the review page and log in (you will be provided with log-in details). If you can't remember an answer, you have the option of jumping back to the specific part of a lecture where the flashcard was encountered.

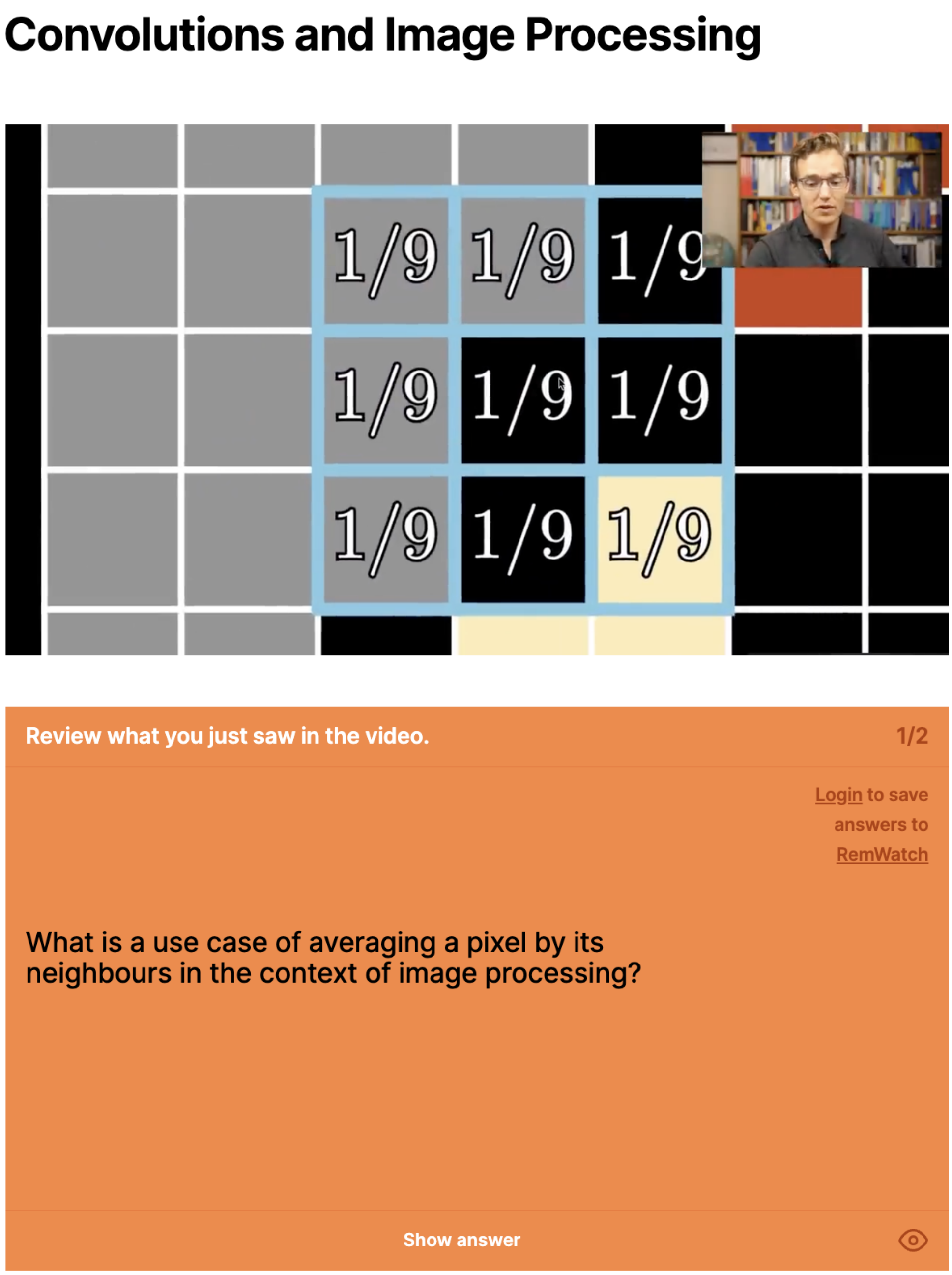


Figure A.2: Screenshot of video watch page with flashcard

Appendix B

Thematic Analysis – Mind Map

Mind map created for the thematic analysis of the data collected during the learning experience evaluation, to be found in Figure B.1.

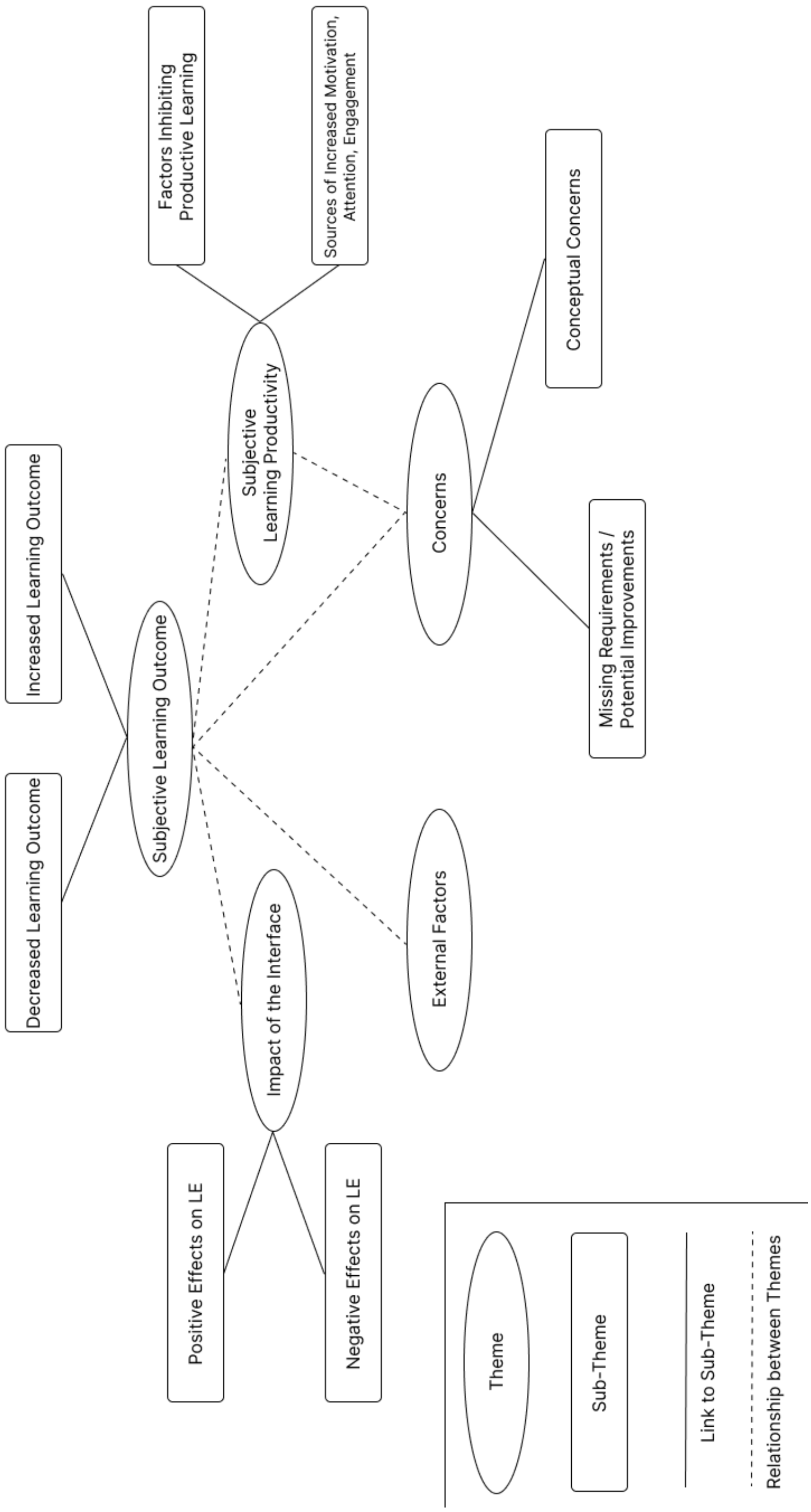


Figure B.1: Mind map of themes discovered during a bottom-up thematic analysis of learning experience evaluation data

Appendix C

Testing – Requirements Analysis and Approaches

Testing approaches for the requirements specified in Table C.1:

- R1: End-to-end testing using Functional Testing
- R2: Penetration testing using Combinatorial Testing
- R3: Manual accessibility testing using Functional Testing
- R4: Continuous Development Testing Pipeline
- R5: Top-Down Functional Combinatorial Testing
- R6: Performance Testing
- R7: Unit Testing using a combination of Structural and Functional Testing
- R8: Performance Testing

ID	Requirement	Type, Level, Property	Translation to Measurable Requirement
R1	The system should prevent unauthorised access to and tempering with the database via its UI.	functional, system, safety	
R2	The system should protect its API against Denial-of-Service attacks.	functional, system, robustness	
R3	The system UI should be accessible to visually-impaired people.	qualitative, system, accessibility	The system UI should conform to WCAG standards and hence be possible to be navigated via keyboard only.
R4	The system should not have software fragments which endanger the availability of the system to the end-user.	qualitative, system, availability	The system should have a feature to prevent software fragments which cause compilation failure from being deployed.
R5	The system's API should integrate correctly with commercial off-the-shelf (COTS) components provided by the authentication provider so that certain routes are auth protected.	functional, integration, security	
R6	The system's API should have seamless integration with the database by using minimal latency communication.	qualitative, integration, performance	Given a sufficient network connection, the system should be able to transfer data to and from the database via its API at the speed of 50kb/s, at minimum.
R7	The spaced repetition scheduling unit should predict the correct next due date based on given review history and the given user answer, conforming to the Leitner-system specification. This means that it should double the previous review interval $x\_new = 2 * x\_prev$ or resetting it to a single day.	functional, unit, correctness	
R8	The spaced repetition scheduling unit should be efficient.	qualitative, unit, performance	The unit should calculate the next due date in $<$100ms.

Table C.1: Testing-specific requirements of our tool

Appendix D

Testing – Test Plan

This appendix provides a test plan specification for identified tasks [T], including tests (Figure D.1).

To purposefully assign the tasks, we estimated how long the task will take, and what is necessary for the task to be possible: resources, code, data, results from earlier tasks. The test plan is specified in Table D.1, tasks associated with testing activities are coloured.

Task ID	Description	Needs	Pass Condition
T1	Perform an E2E-test on the review platform UI to test for R1	prior completion of T3, T16	no authorised access is gained
T2	Perform an E2E-test on the flash-card UI to test for R1	prior completion of T3, T16	no authorised access is gained
T3	Research basic concepts of E2E-testing, and tools to realise them		
T4	Find suitable tool which performs mock DoS attacks to verify R2		
T5	Prepare means to analyse results from simulated DoS attacks		
T6	Perform manual human UI accessibility testing	prior completion of T8	the user can navigate all elements via keyboard
T7	Static code analysis (not carried out as non-critical)		
T8	Research testing framework for human UI accessibility testing		
T9	Analyse T6 results	prior completion of T6	
T10	Test for compilation ability to verify R4		system compiles successfully
T11	Integrate the compilation ability test into a continuous testing environment	prior completion of T10	
T12	Exhaustively test for authenticated access of API routes	prior completion of T16	can only get authenticated API access if user is authorised, otherwise not
T13	Create a synthetic data generator		
T14	Generate of relevant input values		
T15	Create relevant input data for the unit, unit testing	prior completion of T14	unit returns expected result for relevant input
T16	Create scaffolding to programmatically simulate registering and logging in a user, so that a driver or other part of a test program achieves the needed authentication status		
T17	Instrument code to log how UI interactions are interpreted by the system		
T18	Test performance of scheduling unit to verify R8		performs <100ms
T19	Create scaffolding for the scheduling unit to accept different date formats for testing purposes		

Table D.1: Test plan consisting of a task specification, task needs, and a schedule

Appendix E

Testing – Omissions and Inspection

Omissions and deficiencies of the test plan from Appendix D are summarised in Table E.1. Results of a classic checklist inspection [40] are found in Table E.2. The test plan was updated accordingly. Overall, the checklist exposes a few weaknesses of the test plan, including unclear fail conditions for testing tasks, unaddressed risks, and unclear system level of tasks. As to the unaddressed risks, the test plan does not include them since they were ranked as negligible for this system and the given tight time schedule. However, this is a clear limitation of the test plan.

Significant omissions & deficiencies	Comment	Potential Remedies	Affected Tests
Absence of appropriate tools / development	The usage of the React framework made it extremely difficult to compute test coverage via traditional profiling tools due to the interwoven structure of programmatic logic and UI elements of the files.	More research into appropriate structural analysis tools.	T18
Test coverage	The chosen test cases showcased for this portfolio do not cover much of the system. This is due to limited time available, and having the goal of planning and implementing a range of different tests rather than tests with high coverage.	More time budgeted towards A&V processes.	
Use of inadequate tool	The use of Jest was not appropriate for calculating the test coverage as it could not deal with transpiled code for calculating the test coverage, which was a specific requirement stated in the coursework specification. There are Babel plugins for solving this, but I did not have the time to make them work for my system setup.	Use a different testing framework or different system setup which allows for coverage calculation.	
Lack of data	The API robustness tests (e.g. the DoS mock attacks) are only somewhat indicative of the system meeting the respective requirement. If we had real-world data of how likely the system is to experience such attacks, and at what scale, this would be invaluable training data and would lead to more sophisticated and useful test cases.	Long-term: add API-use monitoring functionality to the system for recording such kind of data	T4, T12

Table E.1: Significant omissions, deficiencies, and potential remedies in the testing process

Feature	How to check	Property to check	yes	no
REQUIREMENTS TO BE TESTED	For each requirement R, does the plan include	a priority rating (high / moderate / low)	✓	
		a justification for the priority rating	✓	
		a comment on A&T needs (even if ”no needs”)	✓	
A&T NEEDS	For each item in each A&T needs section, does it include	a reference to an associated task [T_N] OR direct resolving of the need	✓	
TASK SPECIFICATIONS	For each task, does it include	a description of what the task involves	✓	
		stated needs to complete the task	✓	
		an estimated completion time (ECT)	✓	
		an indication of which level it is tailored towards		✓
		a justified assignment into a software lifecycle phase	✓	
RISK ASSESSMENT & EVALUATION	Are the following risks addressed	personnel risks	✓	
		technology risks	✓	
		schedule risks	✓	
		development risks	✓	
		execution risks	✓	
		risks from critical requirements	✓	
RISK MITIGATION	Do the following risks have a mitigation plan associated with them	personnel risks	✓	
		technology risks	✓	
		schedule risks	✓	
		development risks	✓	
		execution risks	✓	
		risks from critical requirements	✓	
PASS/FAIL CRITERIA	For each task associated with testing, does it include	a clearly stated criterion stating the pass condition	✓	
		a pass condition which implies when the test fails		✓

Table E.2: Classic checklist inspection for our test plan and system. A&T refers to Analysis and Testing processes

Appendix F

Timeline for Next Year

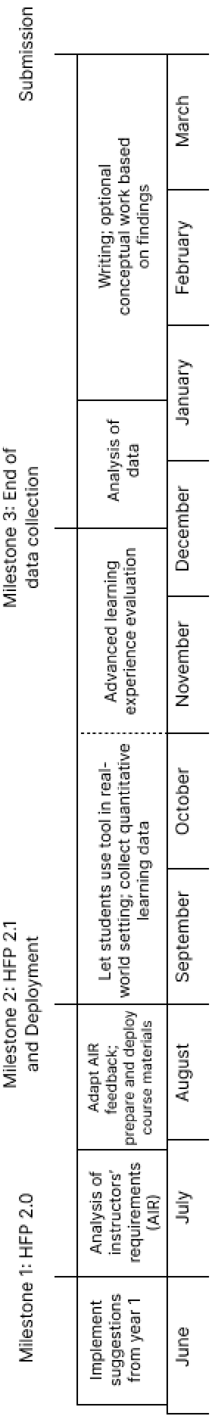


Figure F.1: Rough timeline for work to be carried out in year two of this project

Appendix G

System-Usability-Scale Feedback Form

Usability Feedback

1. I think I would like to use this system frequently

system = integrated flashcards in lecture recordings, how you just tested them

1	2	3	4	5
Strongly disagree				Strongly agree

2. I found the system unnecessarily complex

1	2	3	4	5
Strongly disagree				Strongly agree

3. I thought the system was easy to use

1	2	3	4	5
Strongly disagree				Strongly agree

4. I think that I would need the support of a technical person to be able to use this system

1 2 3 4 5

Strongly disagree Strongly agree

5. I found the various functions in this system were well integrated

"functions" are any things that make the system work, such as the answer button, or the automatic stopping of the video when a flashcard needs to be answered.

1 2 3 4 5

Strongly disagree Strongly agree

6. I thought there was too much inconsistency in this system

Inconsistency regarding the usability

1 2 3 4 5

Strongly disagree Strongly agree

7. I would imagine that most people would learn to use this system very quickly

1 2 3 4 5

Strongly disagree Strongly agree

8. I found the system very cumbersome to use

1 2 3 4 5

Strongly disagree Strongly agree

9. I felt very confident using the system

1	2	3	4	5
Strongly disagree				Strongly agree

10. I needed to learn a lot of things before I could get going with this system

1	2	3	4	5
Strongly disagree				Strongly agree

Appendix H

Participants' Information Sheet and Consent Form

Participant Information Sheet

Project title:	Interactive streaming videos for educational use
Principal investigator:	John Lee
Researcher collecting data:	Eric Janto
Funder (if applicable):	

This study was certified according to the Informatics Research Ethics Process, reference number 7235. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

Eric Janto (student)

John Lee (supervisor)

What is the purpose of the study?

There are two purposes of the study:

1. Identify the usability of my software.
2. Investigate whether the software improves students' learning experience with lecture recordings

Why have I been asked to take part?

The research target group includes anyone who is a University student. It is preferred yet not necessary that students have any of the below backgrounds:

- Students who have used flashcard software before (such as Anki, Quizlet)
- Students with experience in User Interface design
- Students who have experience with using lecture recordings to study for a subject

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until you submit one of the questionnaire forms associated with the study without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

If you decide to take part, these kinds of data are being collected:

- Your academic background (degree, year)
- Your prior experience with using flashcards
- Your answers to questions regarding specific user interface and usability of software
- Your learning experience around using my system

The means of collection will be one or multiple questionnaire forms. You will be asked to interact with the system whilst following along with it / them.

You might be asked to participate in an optional follow-up workshop. In the workshop, you would be asked to use the flashcard system in various ways to investigate the student's learning experience with the system.

The study is expected to take 1-2h in total at most.

You will not be video recorded.

The study will happen at multiple occasions, in-person (for the first meeting and potential workshop) and offline as a self-study (for follow-up questionnaire forms). For in-person meetings, the location and time will be agreed on with the participant (you) but should happen during semester 2 and at a University facility or other public space.

Compensation.

There is no monetary compensation for your participation in this study (£0). You may be provided with refreshments.

Are there any risks associated with taking part?

There are no significant risks associated with participation

Are there any benefits associated with taking part?

Potential refreshments may come in the form of a cup of tea, snack, or coffee.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher (Eric Janto), principal investigator (John Lee), and secondary dissertation marker (Luo Mai).

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Eric Janto via email: s1975761@ed.ac.uk.

If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Eric Janto via email: s1975761@ed.ac.uk.

General information.

For general information about how we use your data, go to: edin.ac/privacy-research

Participant number:____________________

Participant Consent Form

Project title:	Interactive Streaming Videos for educational use
Principal investigator (PI):	John Lee
Researcher:	Eric janto
PI contact details:	john.lee@ed.ac.uk

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1. I agree to being audio recorded.

Yes **No**

2. I allow my data to be used in future ethically approved research.

Yes **No**

3. I agree to take part in this study.

Yes **No**

Name of person giving consent	Date dd/mm/yy	Signature
Name of person taking consent	Date dd/mm/yy	Signature

THE UNIVERSITY *of* EDINBURGH informatics