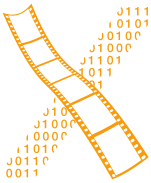

Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Automating User Skill Evaluation to Provide Matching DIY Tutorials

Master's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Ricarda Rahm

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr.-Ing. Ulrik Schroeder

Registration date: 28.01.2022
Submission date: 27.07.2022

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
Para. 156 StGB (German Criminal Code): False Statutory Declarations
Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence
(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.
(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures

List of Tables

Abstract

Finding DIY tutorials for which the required skill set matches the user's skill set can be frustrating for users. There are no clear characteristics or indicators of tutorials that help to find a matching tutorial. Cognitive biases in self-assessment about one's skills further complicate the search. In this work, we want to explore methods to automatically assess a user's skills to empower the user with more filtering options. For this, we have limited the work to the subject of sewing. The first research objective is to identify factors that professionals and instructors use to assess the quality of sewing work or the hand-crafting skills of a sewing person. In a semi-structured interview, we identified eleven factors used by professionals to assess sewing skills. Subsequently, we derived concepts that use these factors as assessment mechanisms to assess a user's skills. The second research objective of this work is to investigate to what extent our proposed concepts, to assess the user's skills, are accepted by users. To this end, we set up a mixed-methods user story, where we let users interact with the prototype. In a survey and a semi-structured interview, we gather data to evaluate the extent of user acceptance. In statistical investigations, we find out that there is no significant difference between the acceptance of the proposed prototypes. Further, we analyze the survey and the interviews and conclude that all proposed concepts are accepted, but we also find evidence that they are not suitable for all hobby sewers.

Überblick

Die Suche nach DIY-Tutorials, bei denen die geforderten Fähigkeiten mit denen des Nutzers übereinstimmen, kann für Nutzer frustrierend sein. Es gibt keine eindeutigen Merkmale oder Indikatoren, die dabei helfen, eine passende Anleitung zu finden. Kognitive Verzerrungen der Selbsteinschätzung der eigenen Fähigkeiten erschweren die Suche zusätzlich. In dieser Arbeit werden wir Methoden zur automatischen Bewertung der Fähigkeiten eines Benutzers erforschen, um Nutzer mehr Filtermöglichkeiten zur Verfügung zu stellen. Im Rahmen dieser Arbeit, haben wir uns auf das Thema Nähen beschränkt. Das erste Forschungsziel besteht darin, Faktoren zu identifizieren, die Fachverständige und Ausbilder der Nähtechniken nutzen, um die Qualität einer Näharbeit oder die handwerklichen Fähigkeiten einer nähenden Person zu bewerten. In einem halbstrukturierten Interview konnten wir elf Faktoren identifizieren, die Fachverständige zur Beurteilung von Nähfertigkeiten verwenden. Anschließend haben wir daraus Konzepte abgeleitet, die diese Faktoren als Bewertungsmechanismen zur Beurteilung der Fähigkeiten eines Anwenders nutzen. Das zweite Forschungsziel dieser Arbeit ist es, zu untersuchen, inwieweit die von uns vorgeschlagenen Konzepte zur Bewertung der Fähigkeiten von Anwendern akzeptiert werden. Zu diesem Zweck haben wir eine Mixed-Methods-Studie durchgeführt, in der wir die Nutzer mit den jeweiligen Prototypen interagieren ließen. Anschließend, sammelten wir in einer Umfrage und einem halbstrukturierten Interview Daten, um den Grad der Benutzerakzeptanz zu bewerten. In statistischen Untersuchungen fanden wir heraus, dass es keinen signifikanten Unterschied zwischen der Akzeptanz der vorgeschlagenen Prototypen gibt. Anschließend haben wir die Umfrage und die Interviews analysiert und sind zu dem Schluss gekommen, dass alle vorgeschlagenen Konzepte akzeptiert werden. Allerdings haben wir auch Hinweise darauf gefunden, dass sie nicht für alle Hobbynäher geeignet sind.

Acknowledgements

First, I would like to thank Prof. Dr. Jan Borchers and Prof. Dr.-Ing. Ulrik Schroeder for examining this thesis.

Special thanks to my supervisor Marcel Lahaye, for the insightful discussions, motivating words, and enthusiastic mentoring.

Without study participants, this work would not have been possible, so I thank you for your time and your contributions to this work. Thanks to all people who supported the search for participants.

Laura and Alex, thanks for your emotional support and all the helpful advice you gave me!

Julian, I'm more than thankful for your unfailing, loving support.

Thank you, mom and grandma. Even if you don't really understand what I'm doing "with computers," I'm grateful you're listening and supporting me anyway. Last but not least: Thanks, dad — for thinking it was a good idea to put me in front of computers as a 3-year-old. I wish you could witness how it turned out.

Conventions

Throughout this thesis we use the following conventions.

Codes and implementation symbols are written in typewriter-style text.

`Code`

The whole thesis is written in American English.

To include people of all gender identity, we use singular *they*.

The *editorial we* is used.

Quotations from interviews were mainly translated from German. For technical terms of sewing, the technical dictionary by Schubert [2005] was used.

Chapter 1

Introduction

When users are about to start a new handcraft project and search for tutorials to learn how to do it, they will come across many tutorials from which they have to choose. One factor in deciding on a tutorial is that users estimate whether they would complete the project successfully. However, this causes challenges for users because tutorials do not have clear properties that users can use to determine that it matches their skills. On the one hand, this tedious search can lead to decision fatigue (Pignatiello et al. [2020]), as it is challenging to decide on a tutorial without these clear properties. On the other hand, it can also be frustrating to start tutorials and only later realize that the required skills of the tutorial do not match their own.

Searching tutorials can be frustrating

In this context, matching means that the user's handcrafting skill set and the tutorial's required skill set match together. Matching tutorials implies that a user can rework shown techniques without seeking further help and that the user does not have to quit the project.

Tutorials should match user's skills

Platforms where users can find tutorials include social networks like YouTube, TikTok, or Instagram, blogs, or websites explicitly built for tutorials like Thingiverse. On YouTube, one of the possible ways to find a tutorial is to search for specific terms, e.g., "DIY Table for Beginners." In the search results, users could determine whether the video matches their skills based on the headline, description, or

No clear properties that indicate skill level

thumbnail. On TikTok and Instagram, tags, the description, or the content itself could reveal which skills users need. All of the mentioned websites share that there is no clear indication about what skills are needed to accomplish the tutorial. A website that offers sewing patterns, Burda Style, indicates with displayed stars the needed skill level of the tutorial. One star equals that the pattern is easy to do, and four stars indicate it is challenging.

Challenges of assessing own skills

However, this is not the only challenge the user needs to solve to find a matching tutorial. The other challenge is to evaluate their own skill set or skill so that they can understand if the required skill set equals their own skill set. Related literature by Kruger and Dunning [1999] shows that self-assessment is affected by cognitive bias. However, the problem of cognitive bias affects not only the user but also the creator of the tutorial. Moreover, it is possible that the estimation of the required skill level set, which a creator might indicate, is biased by the creator as well.

Objective to empower the user to find matching tutorials

In this work, we want to contribute a solution to empower the user with more information to decide on a tutorial. To this end, the system must understand the user's skill set. Since there is research that self-assessment is unreliable, we want the system to perform an automated skill assessment. With this estimation, the system can offer additional filtering options and recommendations that the user can use to decide on a tutorial.

Outline of the research

To get to this solution, we first need to explore possibilities to enable a system to assess and quantify the crafting skills of a user in order to filter the tutorials. For this purpose, we have decided to focus on the field of sewing technology. First, we want to identify factors to assess the quality of a home-sewn product or the skills of hobby sewers. Next, we want to derive concepts that automatically measure the factors for these identified factors. Lastly, we will prototype the concepts and evaluate their user acceptance.

Our research goal is to answer the following research questions:

RQ1 What are factors that professionals and instructors

use to assess the quality of sewing work or the handcrafting skills of a sewing person?

RQ2 To what extent are our proposed concepts for assessing user skills accepted by users?

Chapter 2

Related Work and Background

Since this work is exploratory, we first try to find indications or inspirations in other scientific work that contribute to finding a solution for how to assess maker skills automatically. One possibility for the user to find a tutorial matching their skills is to self-assess their skills and search for tutorials that appear to match. However, the literature shows that self-assessment is often error-prone, but other publications still propose it as a method to evaluate students. Further, we will examine approaches to assess practical skills to see what methods for automated skills already exist and to estimate if we could use one of them in our context. A part of teaching art courses is evaluating handcrafted products and their students' skills. Therefore, we will also briefly look at related literature on this topic. Finally, we consider related work about skill assessment in sewing.

2.1 Using Self-Assessment to Assess Skills

Some humans are "Unskilled and unaware of it" (Kruger and Dunning [1999]). This statement resulted from their

Humans facing issues to be objective about their skills

research about a phenomenon of cognitive bias, which is called the "Dunning-Kruger Effect." The effect describes that people with little skills tend to overestimate their skills, whereas people with many skills rather underestimate themselves. The reason for this lies on the metacognitive level, which figuratively describes how people think about their own thinking (Metcalfe et al. [1994]). On the metacognitive level, humans with limited skills are not yet aware of what they cannot do and therefore cannot evaluate whether they are performing well or poorly. On the other hand, people who have more skills tend to be more aware of what they already can not do and tend to underestimate their skills. For this reason, the trivial option of simply asking users what they are good at or bad at and filtering the tutorials accordingly to the self-assessment is not applicable.

Criteria-based self-assessment is a supposed method to evaluate students

However, research findings related to metacognition have also shown that self-assessment can help achieve learning progress (Andrade and Valtcheva [2009]). The authors state that students should identify their strengths or weakness in their work and that "self-assessment is a process of formative assessment during which students reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria." The research in this area implicates for our work that we can not rely on self-assessment entirely. Still, we could use criteria-based self-assessment like proposed from Andrade and Valtcheva [2009] to evaluate the user skills automatically and recommend matching videos based on self-assessment.

2.2 Analyzing Activities for Skill Assessment

Assessing visual data

Another option to assess a person's skills is observing and analyzing the activities. Observing could be done visually, such as the authors Funke et al. [2019a] did for assessing the skills of surgeons. This approach is based on deep-learning video classification.

Assessing sensory data

Another method to evaluate the skills is the classification of sensory data, similar to the work of Gong et al. [2019] where their instrument analyzes fabrication activities and their expertise with machine learning. In this work, we will exclude the idea of using sensory data to analyze the skills because we want to provide a solution that makers can use at home and thus without the need for devices that can record sensory data. Another way to assess a user's skills is to evaluate how they interact with the system.

Assessing the interaction with the tutorial system

There is research about detecting if users have a novice or skilled use with a user interface Hurst et al. [2007]. Therefore, they consider mouse events and menu data. Related to our problem, this could inspire us to observe how users interact with tutorials, for example, which tutorials they search, how often they pause or rewind, or check which tutorials they have completed.

2.3 Performance Assessment in Art Classes

Using exams, observation or assessing portfolios to assess skills

As makers produce a handcrafted item, we can also consider how art teachers assess their students when their task is to create something handmade. We found a survey contribution from Clark [2002] that evaluates methods art teachers can use to assess the students. As assessment methods, they listed, for example, the assessing portfolios, exams, or direct observation. The latter, we could realize with video or sensory data analysis. Another option is an exam, so we could let the user do a test to assess their skills.

2.4 Assessing Sewing Skills

Skill assessment with an exam and evaluation of a work sample

In a course for basic clothing construction, Blood and Owens [2015] describe that they used to have exams in their classes but added a practical component. In the exam, students have to take an ordinary written exam, where they

can use all course materials like notes and textbooks. In addition, they have to sew a bodice during the exam and submit that for evaluation, but the authors do not explain with which criteria they assess this working sample.

Criteria to assess a home-sewn product

Instead, we found a document that lists quality standards for home-sewn items by Hendrickson et al. [2004]. As a rough estimation, they record that the garment should be: attractive, flat and smooth, free from bulk, have secure stitching at a uniform distance from edge or fold, it's functional and durable. Further, they list categories to evaluate, for example, the general appearance, details (like buttonholes, zippers), fitting, facings, and seams. In another document by the same authors (Kay et al. [2000]), we found a checklist for a workshop offered for children and young adults, which had the following top categories so that they can self-assess the progress of their sewing skills:

- Sewing Tools and Machines
- Fabric Knowledge and Skills
- Pattern Knowledge and Skills
- Construction Knowledge and Skills
- Fabric Care Knowledge and Skills

2.5 Proficiency Levels of Sewing

Vocational training plans

For the vocational training as a custom tailor, there is a plan that defines the sewing skills to be learned for every apprenticeship year (Berufsbildung [2008]). So it would be an option to adapt it in the level division. However, hobby-makers like the unstructured and exploratory nature of experiencing tutorials (Desjardins and Wakkary [2013]). Therefore we expect it is not be promising to stick to vocational training plans in hobby sewing.

There is no official level classification in sewing skills

In hobby sewing, there is no set definition by which sewing skills are divided into levels. To get an idea of what would be possible level divisions, we can take a

look at a renowned fashion magazine that has been published monthly since 1950. Burda Style claims to be one of the most prominent fashion magazines in the world (Style [2022b]) and publishes patterns for hobby sewers in their magazines and on their website.

Using four circles to indicate pattern's skill level

On the German edition of their website (Style [2022a]), they use four circles to indicate the pattern's skill level; easy, medium, advanced, and challenging. It is unclear what criteria they use to categorize the instructions in the levels. Unfortunately, a request to the publisher on how the classification comes about remained unanswered.

Hint that four level skill indication is frustrating

While researching this, we came across a blog post (Patterns [2018]) where a blogger represents what they think about the level division in this magazine:

> Blogger: "There were two problems with this. Not every issue of Burda magazine had even one half circle rated pattern, let alone two or three for me to continue honing my skills like a good little seamstress I was. If I was to follow this line of thinking and only practice patterns within 'my level' I would most certainly mess up my fabric, waste everyone's time by going above my level and spend the rest of my life living in a cardboard box under the bridge. I wasn't sure whose time I would waste and which bridge would become my homestead, but I really didn't want to risk finding out. The other problem was that super easy patterns were plain and quite frankly patterns classified as two circles seemed more fun. And I ogled three circle patterns like an awkward teenage boy checking out his crush but who is too afraid to make eye contact and say hello. No, I simply had to work my way up, the proper way. Who was I to challenge the system with my radical ideas? Sewing police would surely get me."

Even though this blogger's contribution is not scientific but of a more entertaining nature, there is an indication that

the classification into levels for hobby sewers seems to be frustrating.

Chapter 3

Exploratory Study to Identify Factors of Skill Evaluation

The first step in exploring how a system can evaluate automated sewing skills is to understand what factors instructors and teachers of sewing techniques use in practice to assess sewing skills. Following, we will address this research question:

RQ1 What are factors that professionals and instructors use to assess the quality of sewing work or the handcrafting skills of a sewing person?

In the following, we will report the methodology to determine how experts assess sewing skills, the findings, and the discussion.

3.1 Methodology

We conducted semi-structured interviews to generate data about sewing skill assessment methods. In generating and

Semi-Structured interviews with Theoretical Sampling

analyzing the data, we used elements from Grounded Theory (GT), but the product of our study will be a detailed description of sewing skill assessment methods and not a theory that "explains what is happening in a domain," which is, according to Cole and Gillies [2022], a clear distinction from GT. To enrich or challenge already gathered data, we used the Theoretical Sampling (GT): After each interview session, we analyzed the audio transcript. Depending on past interviews, we intensified follow-up on specific questions.

3.1.1 Participants

Professionals or instructors in sewing techniques

We recruited n=10 participants from Germany (all female) by writing invitation emails to local sewing course providers, universities of applied sciences, universities, and vocational schools that offer courses or majors that involve learning sewing techniques. Table 3.1 lists the proficiency of the interviewed persons. We interviewed a mixture of people who teach sewing practically, e.g., in precourses, or who teaches theoretically, e.g., fashion design or clothing technology professors. The language of the interviews was German, we translated all used citations to English.

3.1.2 Procedure

Examination of students and assessment of sewed products

We conducted the study remotely via video chat. On average, a session lasted around 30 minutes. We interviewed the participants in German, and we translated all following quoted text passages into English. First, we asked for the occupational routines with the students. If they worked practically with students, we questioned how they group the students into different levels, how the instructors determine the experience of their students, and how they observe a learning process. If the participant was part of examining students, we asked how they execute exams and what forms of testing they use. As our goal is to understand how a computer system could rate someone's sewing skills,

ID	Proficiency
E1	25 years of experience in sewing and teaches pre-courses at a university
E2	Professor of pattern construction
E3.1	Professor for clothing technology
E3.2	Research Associate teaching clothing technology
E4	Head of a sewing workshop at a university of applied science
E5	Head of a tailor shop at a university of applied science
E6	Teaches manufacturing at a university of applied science
E7	Teaches costume conception at a school of art
E8	Head of a sewing workshop at a university of applied science
E9	Conducts sewing classes for interested students at a university of applied science

Table 3.1: Proficiency of interviewed professionals.

which is a rather virtual approach, we asked how they handled distance learning during the pandemic and how they solved encountered problems.

Assessing sewn product

To learn how they assess a finished sewing work, we presented a sewed product from back, front, and the inside (see figure 3.1). We asked the participants to inspect the sewed item and describe what properties they reflected upon for assessing the quality. To understand if we could match already done projects or used materials with a skill level, we questioned whether they could determine a specific level depending on used materials or techniques and how they would be considered typical as beginners or advanced skills. The last question was about the organization of sewing skills to learn if the skills are hierarchical or if students can learn them independently from each other.

The whole study design is attached in the Appendix B.1

Figure 3.1: Sewed product example.

3.1.3 Analysis

Thematic Analysis with Initial Coding and Focused Coding

After each interview, we transcribed the recorded audio. On this data, we applied a 'Bottom-Up Thematic Analysis' (TA), described by Clarke et al. [2015]. For this, we first familiarized ourselves with the data, generated codes, searched for themes, reviewed the themes and created a report about the identified factors. For the coding, we used the coding methods, which Saldana [2015] suggested for GT: In the first coding cycles, we used *Initial Coding* to divide the transcript into topic-specific segments. As a second cycle method, we applied *Focused Coding* to build up categories based on thematic or conceptual similarity.

3.2 Findings

With Thematic Analysis we identified three coded main *themes*:

```
THEME: Assessing skills of a person
THEME: Assessing quality of a sewed product
THEME: Organization of sewing skills
```

The themes with the code `Assessing skills of a person` and `Assessing quality of a sewed product` we will following identify as *factors*, that explain how experts assess sewing skills and the information about the organization of Sewing Skills. The corresponding coded subthemes can also be understood as subfactors.

For the factor *Assess skills of a person*, the participants gave us information about what sewers do in the sewing process, from which they can conclude a certain skill level. For example, this can be something like how they use a sewing machine or their theoretical knowledge.

Next, they reported how they *assess the quality of a sewed product*, e.g., they check the seam quality or the inner finish of the end product.

We also gained some insights about how the participants think *sewing skills are organized*, if it is rather hierarchically or project-driven organized.

3.2.1 Assess Sewing Skills of a Person

The Participants reported what actions, movements, processes, and other characteristics they look for when assessing the student's abilities while observing them.

As subthemes to `Assess sewing skills of a person`, we coded the following themes in the transcripts of the interview:

```
THEME: Assessing sewing skills of a person
   SUBTHEME: Working with a Sewing Machine
   SUBTHEME: Knowledge
   SUBTHEME: Working Independently
   SUBTHEME: Logical Skills
   SUBTHEME: Manual Dexterity
   SUBTHEME: Soft Factors
```

In the following, we consider this themes to be *factors* to assess sewing skills of a person.

Working with a Sewing Machine

Handling the sewing machine

Participant E1 recognizes if somebody has experience in how they approach the sewing machine. If they "start slowly, with much observation: what does this machine do?", they believe it is a beginner. Expert E8 would "first, basically distinguish that one can handle the machines."

Participants reported the importance of threading. If this is done wrong, the seams can be faulty.

> E3: "That sounds simple now, but if you forget an element there and if you do not know now why I should thread it through the thread take-up lever or the thread-tensioning device, and you forget that, you will not get a seam, the thread will break or is it just faulty stitches. The threading is elementary important."

In E3's class, they stop the time, and every student has 3 minutes to get their machine ready to work, including the threading. Working with the sewing machines also makes sounds, which E8 uses to recognize if an experienced person sews.

> E8: "I hear that in the way.... such a machine has a pedal for acceleration. [...] So I hear that in the way that pedal is operated."

This statement is similar to the one of another participant in that they can hear if someone inserted the needle incorrectly:

> E5: "Usually you can tell by the sound if the needle is not correct, but not always."

Different body postures while sewing

A participant explained they can determine how much experience someone has from the body posture of the students sitting at the sewing machine.

> E8: "You can tell by the way they sit. So you can see if someone is sitting tensely at the machine."

Knowledge

Written exams to assess the student's skills

Participants reported that they use written exams to assess the student's skills.

> E3: "We could, for example, confront them with a concrete sewing issue, and then the student must give solutions or recommendations."

E2 mentioned that students could only "judge [their own sewed products] if [they] know how to do it in theory."

Nevertheless, E2 said that sewing is learning by doing, and maybe "one knows theoretically (...) everything quite well, but when they make it, then it always looks different." This is similar to E4's statement, which says that "much simply has to do with ability and not at all with the knowledge how to do something. But simply how to touch things or how to smooth things out."

Knowing how to influence the result

For E4, the pure theory is not that important, but they say that someone needs to understand how processes influence the result or how to prevent errors.

> E4: "So the whole theory behind it, I think, is not so important. But it is important that you know that something like that influences the result and how you can change it."

Terminology to understand processes

In addition, participants reported why they think technical terms are essential to follow manuals.

> E9: "Because I would also like them to know these certain technical terms; otherwise, they will never be able to read and follow a manual."

Importance of theoretical knowledge

Theoretical knowledge that participants considered important for sewers were how a sewing machine works (E2, E3, E7, E8), determining the grain line of the pattern piece (E8) or how to identify error sources (E5). E5 outlines: "Theoretical knowledge (...) is very, very useful, it has to be said."

Working Independently

Working more independently is considered as learning progress

In the sewing workshop, some of the participants noticed how much help the students needed or if they worked, in general, more independently (E4, E5, E9). If the student worked independently, the instructor perceived this as learning progress and that the student gained experience.

> E4: "I notice that especially through the fact that they (...) do not ask any more questions, and that the result still looks nice."

Necessary aid to complete a task

However, working independently as a student is not only about asking questions but also about how much aid their use to accomplish their task. A participant describes that some of their students draw a line where they want to sew.

> E5: "That' s not something you do. That is one thing that would take way too long if you were to do it all the time. You sew parallel to an edge, and you do it by eye. And someone who always tends to draw themselves a line does not necessarily have a place in a more advanced class."

Choosing more challenging projects

E1 told us that when their students made something more challenging, they thought: "Oh, they had learned something!" For E7, it is not only about the handcraft itself but also the confidence of the students to aim at challenging projects.

> E7: "I name the progress that I see. (...) And they are often not in the straight seam or the

> perfect, but the increasingly progressing imagination, in what they are doing."

When we asked what is a sign of learning progress in sewing, a participant answered the following:

> E9: "Then we showed different things, and you could see how they evolved their own ideas suddenly."

Logical Skills

Thinking steps ahead

Participants reported that sewers need logical skills because sewers always need to think some steps ahead.

> E1: "I mean, sewing is, is a very, very complicated thing. It's a bit like playing chess. You always have to think one step ahead. You have to put everything the other way around, and then you see it afterward in the finished state. So face on face means I always see the wrong side and so on. The classic mistake is putting it under the machine the way it will look later. Yes, this whole thinking ahead and everything the other way around."

Spatial visualization ability

Another participant also mentioned the importance of spatial visualization ability and described that this is needed to develop the pattern on paper first and to turn it into a physical product.

> E7: "I looked at how someone sets up a pattern, that is, what the patterns look like. And whether they have been done with heart and soul, so whether I can already see in it what kind of spatial understanding someone has."

Manual Dexterity

Participants think manual dexterity is important

Participants reported that manual dexterity is fundamental to sewing.

> E1: "It is a fundamental thing whether you have a little bit of dexterity or whether you realize a little bit of what is already apparent in theory."

Participant E3 specifies this more precisely, that one needs a "feeling for the pedal of the sewing machine, what it does and how all the elements move." E7 paid attention to "how does someone touch something, how do they hold it in their hand?" Once, E9 had a student who did not have any manual dexterity, and they were not able to make a stitch by hand. Sometime later, E9 told us, the student gave up on sewing.

Soft Factors

For some sewing is about having joy

E7 explained to us that they evaluate the ambition of their students. Someone needed to show assiduity, and the students needed to participate in their course continuously. For them, it is "not important, that it is perfect. It must give joy." (E7)

Sewers need a good eye for detail

Another participant claimed that sewers need a good eye for detail.

> E5: "In principle, they have an eye for it and know that you must pay attention to certain things. Moreover, it is not just about: is it together or still apart?"

3.2.2 Assess a Sewed Product

As subthemes to `Assessing quality of a sewed product`, we coded the following themes in the transcripts

of the interview:

```
THEME: Assessing quality of a sewed product
   SUBTHEME: Processing
   SUBTHEME: Seams
   SUBTHEME: Used Materials
   SUBTHEME: Functionality
   SUBTHEME: Overall Impression
```

In the following, we consider this themes to be *factors* to assess a sewed product.

Working samples are assessed

Participants described that students must submit a finished sewn product:

> E3: "During the semester, a product is manufactured, for example, a shirt (...), which, of course, was explained beforehand in the processing. This shirt is then evaluated."

Some of the participants explained they test their students in a practical exam on-site, in which the students have 90 minutes to complete a sewing task. The "(...) final result (can be) used to see whether these processing steps have been done correctly." (E3)

For assessing the final product, a participant explained they use an evaluation scheme.

> E2: "So if the overall part has 100 points (...) and I give for an outer processing 40 (points), interior-processing 40 (points), and 20 (points) for fabric choice or something. So if you divide it, it is also a little fairer than if I simply say it is a grade C or D."

In such an evaluation scheme, there could be the following properties to be considered.

Processing

Commenting on our work sample (see Figure 3.1), the participants demonstrated what details they look for when assessing the processing of a sewing piece.

Evaluate the the processing

On that bag, a participant commented, for example, that we should have processed it otherwise, so the tape from our sewn-in zipper should not be visible in the inner of the little bag. In our work sample, attention was also paid to whether the corners were processed correctly. Further, participants criticized that we have sewn the corners in two steps and not in one step all the way around and we skipped ironing the seams. Participants also criticized that we did not cut off the seam allowances. They also considered whether the selvages were precisely aligned.

Assessing details, like button holes or letter corners

In general, we observed that the instructors paid much attention to details, e.g., E2 considered whether students have sewed on buttons correctly or how they finished the waistband, if slits, letter corners, and lapels are present, and how the students worked them. E6 stated that they also consider whether the finished garment is symmetrical. Lock techniques (such as buttons or zippers) are a feature that participants mentioned more frequently (E2, E6, E8, E9), and some of them emphasized that for example: "(...) sewing in a zipper is not as easy as it seems". (E8) Further, a participant also said that they notice when people intentionally use basic techniques in their classes:

> E8: "In the pants class, we are practicing sewing zippers or button panels (...), he did not practice it, and I get (as a result) snap fasteners instead."

Cutting process is important for the quality

The cutting process is also of utmost importance. Thus, the participants paid attention to whether the students did the cutting to match the thread line (E2). Besides, E4 emphasizes that "if it is only cut in any way," it "does not look nice in the end."

Participants assessed the inside of the bag

The participants looked at the inside of our bag to assess

our work sample. Many noticed visible inside seams and criticized that they did not disappear between the outer fabric and the lining (E1, E2, E6, E9). Thus, E6 also emphasized that most of the time, someone can recognize a high-quality garment if one places their hand in the pocket and can not feel any seam allowances. In addition, some participants checked if we serged the selvages (E2, E5) and that no protruding threads were visible (E5).

Ironing is eminently important for sewing

In every interview, the participants mentioned ironing.

> E8: "Ironing is eminently important. Sewing does the machine, and the quality of the work result depends directly on the ironing."

In addition, E2 stated that a sewing piece is of higher quality or can still be improved if somebody knows how to iron it. If someone does not master ironing, they could get the sewing piece "in pretty bad shape."

Seams

Quality of seams as assessing factor

In our work sample, the participants often examined the seams (E1, E2, E5, E8). Among other things, they paid attention to whether the seams were straight and neat, the overall appearance, and whether they were evenly and durable.

For judging seam quality, quite a few criteria exist, as E3 described. Among others, these are:

- Seam width
- Stitch lengths
- Stitch length variations
- Seam crimping

E3 also said that for all the roundabout 100 sewing stitch types, there is a standard that determines the specific thread position in the sewing material. In addition, the participants often examined the seam allowance, which was too short in our work sample, which noticed E1, E2, E5, E8, and E9.

Appropriate seam allowance

As a rule of thumb, a participant described an appropriate seam allowance like this:

> E5: "A seam is a good seam if it runs parallel to the selvage and if there is about a centimeter of seam allowance, so that if I pull on it a slightly, it does not open up again right away."

Used Materials

Not all fabrics are suitable for all experience levels

When we asked if beginners or advanced should use a specific material, some participants answered that beginners should use cotton or "simple" fabrics first (E1, E2, E4, E6) because the fabric has a "good adhesion" (E2) and "forgives more mistakes" (E4). Advanced sewers can also tackle materials such as velvet, plush, corduroy, leather, silk (E2), or jersey (E1). E6 mentioned that someone with much experience knows what fabric to choose for a project. On the other hand, people who do not have a clue would not worry about it but buy something "that looks nice." About using materials, one participant commented that experienced sewers can work with more difficult materials.

> E3: "I think someone with experience definitely copes better with more difficult materials. (...) It is also always a matter of taste and [...] it is also the machine that should then be suitable [...]."

Functionality

The sewed product should have a good fit

With accessories like bags or similar things, there is no

property of dimensional accuracy, but with "clothing, we have this big field of fit" (E8). The garments sewn by students in the workshops, the pieces are pulled onto a tailor's bust to see if the garment has the right fit (E2). Thus, E1 also notices when people have made mistakes.

> E1: "A student asked me if I had a tailor's bust in size 32, then I asked and... I already know that is a little bit mean... I said, 'You did not make a seam allowance. That is why your dress is now not a size 34 and 36, but a size 32.'"

E2 also emphasized that it increases difficulty sewing a very tight-fitting garment because, if someone sewed "some oversized piece, it is not so dramatic. However, if you are sewing a sheath dress or a pair of trousers, for example, then it becomes challenging."

Sewed product must be functional

Some also paid attention to whether our work sample is functional, i.e., whether it fulfills the function (E4). E8 found it was enough in some cases: "If it should fulfill the pure function, then it does not matter if the seam is crooked and if threads are visible inside."

Overall Impression

Participants evaluated the overall impression

Participants also paid attention to the overall impression of a sewing piece.

> E5: "Things like that make such an overall impression, and then you can look: Has it been completely worked, or is there a corner missing somehow. Does it open up again, or can you still see the thread in a blind stitch, or can you not see it? So has it been worked correctly? Has it been sewed evenly? Has it been ironed? It is known that we can see if it is done so crudely with sweaty fingers, or if someone has made an effort or is trying."

Besides uniformity, E8 also includes dimensional accuracy in the overall impression.

3.2.3 Approaches to Organize Sewing Skills

To understand how sewing skills are organized, we asked whether it is *hierarchical* or *project-driven*. We explained to the participants that *hierarchical* means an organization similar to learning to write in school. First, students learn the ABC, then words and sentences. We explained the other *project-driven option* to be like an entirely different toolbox, with things that sewers can learn interdependently.

Both organizational forms are possible

Some participants (E1, E5, and E8) believed both structures are possible, and E5 explained that with the *project-driven* approach, a sewer may fail more often, but they still learn something.

Hierarchical Approach

Starting from scratch

A participant described how a hierarchical way of learning sewing worked in practice and explained that they were starting from scratch.

> E6: "We start really from scratch: Explaining sewing machine, machine threading, ironing machines, and all this machine stuff (...) and then it goes step by step."

Learning step by step

Further, E6 explained that the students practice sewing on paper, then simple seams on fabric and different techniques until they sew a collarless shirt, and finally more complex garments. They said that all basic techniques would repeat themselves in every new piece their student sew. In this approach, E1 explained, the students start with something easy, create a knowledge base, and then gain a routine by practicing the techniques. A participant also emphasized that this approach has the advantage that it is good to motivate people to have a sense of achievement.

> E2: "A small part where you can also see the success, and you know that it goes well. They will manage that as a beginner."

Easier but more time consuming

E5 compared the approach to the *project-driven*, claiming that it is "a little easier" to learn sewing hierarchically but takes more time.

Regulations for vocational training

Further, E8 explained that vocational training regulates the order students learn sewing: In the first apprenticeship year, they sew something like a skirt, with an easy cutting technique whereby the realization is simple. Following, they sew more challenging pieces like a blouse or pants. In the last year, students were finally able to sew a blazer. Furthermore, E8 concludes: "And that is indeed the way it is, because different garments need different techniques, that you need to master, to be able to continue (with more challenging projects)."

Practicing sewing on paper

Practicing sewing on paper before starting with fabrics was mentioned by E1 and E3 as well. E1 described that the first thing is to sew a line on paper, then sew next to the line and with distance to the edge of the sheet, as it would be done later in actual sewing when adding a seam allowance.

> E1: "So we first sew a line on a sheet of paper. There are also templates from companies where you first practice sewing on this line. Then to sew next to the line, then if there are only strokes, and then to sew next to the line, where there is no line at all, that I get a wide presser foot wide spacing, then to sew over a dart, how I get this swing."

In contrast, another participant strictly rejects this approach.

> E7: "Sewing on paper, I hold no brief for that. That is completely nonsensical and idiotic because we do not sew costumes out of paper. For costumes, we sew garments out of fabric."

Project-Driven Approach

Approach sewing projects without learning the basics step by step

One participant explains that they think sewers have a specific piece in mind that they want to sew and that it brings no joy to do the basics step by step before approaching the project.

> E4: "Usually: You have an idea, and then you want to realize it. It is rarely the case that you say, 'Okay, for my trousers, I will leave out the pockets. I will do the trousers this time' This is somehow not fun."

Further, they said that they had observed students attempting challenging techniques without having all the prior knowledge. E1 has also observed students who do not work from patterns but drape the fabric on a dressmaker's bust and see how it looks. The garment is then fastened with needles and sewn in place.

This approach might lead to failed projects

However, another participant objected to this approach and stressed that failed projects also come with wasted fabric and money.

> E1: "If they say, "I do not want these basics. I am just going to dive right in." I can not imagine that makes them happy in the end. After all, if it looks crooked and skewed, they say, "Well, it is all crooked now. Next time, it will be fine". First of all, it costs fabric and a lot of money. "

3.3 Discussion

Using the results from Section 3.2, we can answer RQ1 and describe which *factors* professionals and instructors use to assess the quality of sewing work or the handcrafting skills of a sewing person. We listed the identified factors in the interviews in Table 3.2, and in the following, we will discuss what the factors mean for the further course of this work.

ID	Factor
	Assessing skills of a person
F1	Working with a sewing machine
F1.1	Sound
F1.2	Speed
F1.3	Handling the sewing machine
F1.4	Body Posture
F2	Knowledge
F2.1	Identify source of errors
F2.2	Theoretical knowledge
F3	Working independently
F3.1	Required help
F3.2	Choosing more challenging projects
F3.3	Generating own ideas
F4	Logical skills
F4.1	General logical skills
F4.2	Spatial visualization ability
F5	Manual dexterity
F6	Soft factors
F6.1	Motivation
F6.2	Talent
	Assessing quality of a sewed product
F7	Processing
F7.1	Details
F7.2	Cutting
F7.3	Inner finish
F7.4	Ironed
F8	Seams
F8.1	Seam quality
F8.2	Seam allowance
F8.3	Trimmed seams
F9	Used materials
F10	Functionality
F10.1	Fit
F10.2	Pure functionality
F11	Overall impression

Table 3.2: Overview of identified factors in the expert interviews.

Assess how people work with sewing machines

We have discovered that participants can determine the person's skills by observing or testing the person and evaluating a finished sewn product. To assess a person's sewing skills, we could evaluate how a sewer works with their sewing machine. Therefore, we could use sensors to track the sewing machine's *sound* [F1.1] or *speed* [F1.2]. To this end, we would have to explore methods to map the sound to skills, such as using a deep learning approach for acoustics (Bianco et al. [2019]). For the factor *speed* [F1.2], we could evaluate the time they need to sew a specific garment, or we use the machine speed to conclude their skills. For *handling the sewing machine* [F1.3], we could assess if the person knows everything about using it and how to thread it. The *body posture* [F1.4] could be detectable with the methods that Roh et al. [2018] described. We assume we could implement a classifier with training data about different skilled sewers and their body posture on the machine.

Testing theoretical knowledge

Some participants explained that they use written exams to evaluate their student's skills. We could assess the *knowledge* [F2] with a test. For this purpose, we could develop a questionnaire with sewing experts that maps sewing skills correctly.

Determine independent working

The factor of *working independently* [F3] offers the possibility to evaluate which other tutorials are user consumed to see, for example, if they search for other explaining tutorials during their current project to see if they *required help* [F3.1]. From the search history of tutorials, we could also make conclusions if users try to approach more *challenging projects* [F3.2]. As *generating own ideas* [F3.3] might conclude that they will not interact with the tutorial system, we currently do not see an approach to use this assessment method.

Assess logical skills

As there is already research about assessing *logical skills* [F4] with a gamification approach by Yilmaz and Kayali [2016], we assume that a computer system can assess a user's logical skills, but this would also include figuring out how to map logical skills to sewing skills.

Measure dexterity

To measure *dexterity* [F5], we could use the computer

vision-based analysis proposed by Funke et al. [2019b] to analyze how the person works.

Unclear how to measure soft factors

We understand that *talent* [F6.1] and *motivation* [F6.2] also play a significant role in sewing. However, we are unclear on how to measure these *soft factors* [F6].

Assess quality of a sewed product

For assessing the *quality of a sewed product*, we expect that it is possible to use image classification (see Li et al. [2019]) to measure the quality of the factors *Processing* [F7] and *Seams* [S8] on a sewn product. For factors of *Processing* [F7], we could also let the user criteria-based self-assess Andrade and Valtcheva [2009] which experience they already have with different processing techniques.

Assess working sample sewed on paper

As some participants mentioned that they instruct students to practice sewing on paper, we also expect that it is possible to calculate the difference between the sewn line and a preprinted line, which would assess the *seam quality* [F8.1] and *seam allowance* [F8.2].

Ask user about used materials

For assessing the factor *Used Materials* [F9], we think it is possible to ask the user which materials they already used. For this, we would need to examine how the different materials map to a skill level.

Unclear how to measure functionality and overall impression

As the factor *Functionality* [F10] has a wide field of fit, we are not aware of an option for how to measure this automatically. The *overall impression* [F11] is a generic factor, so we are currently unaware of how we could teach a system to assess this.

Chapter 4

Derived Concepts for Automated Assessment of Practical Skills

With the expert interview, we found factors that instructors use to assess someone's sewing skills or the quality of a sewed piece. In Section 3.3, we have already discussed which factors are promising to evaluate sewing skills in an automated way.

Derived assessment mechanisms

From these findings, we used the factors to develop concepts with *assessment mechanisms* that can assess a user's skills. We selected some factors for which we expect the *assessment mechanism* is realizable with a computer system.

Embedded assessment mechanisms to prototypes

We developed a clickable prototype for each conceptualized assessment mechanism that shows the user interaction exemplary. After the prototypical assessment, the prototype will present the determined *recommendations* to the user. These *recommendations* are the tutorials the system evaluated matching the user's skills. The technical conception and evaluation of feasibility are not part of this work.

In Table 4.1, we present the prototypes and the used assessment mechanic and the respective factors.

ID	Name	Assessment Mechanic	Factor
PT1	Sewing on Paper	System prompts the user to print a file with preprinted lines and to sew the lines onto the paper to evaluate the seam quality	F8.1, F8.2
PT2	Carla's Sewing Studio	System prompts the user to go through a quiz to evaluate knowledge using a gamification approach	F2
PT3	Projects Done	System prompts the user to enter completed projects in the category selection and, in the next step, the respective complexity level of the project to evaluate the experience of the user	F7

Table 4.1: Overview of prototypes and respective assessment mechanics.

The factors that we will not consider in the proposed concepts are not excluded because they are not feasible or not promising but because the additional conception of them would exceed the scope of this paper.

4.1 Sewing On Paper (PT1)

Assess seam quality of sewing exercise and match it to required seam quality of the tutorial

For the prototype *Sewing On Paper*, we want to use the skill assessment factor of seam quality [F8.1] and seam allowance [F8.2] to evaluate the user's skills. Some participants mentioned that they instruct their students to practice their seam quality on paper. It helps to learn following the lines and teaches the basics of handling a sewing machine. So the basic idea is to give users a sewing exercise template (see Figure 4.1) and ask them to sew the *preprinted lines*. We expect that it is feasible to assess automatically

how much the *sewed line* on paper deviates from a *preprinted line*.

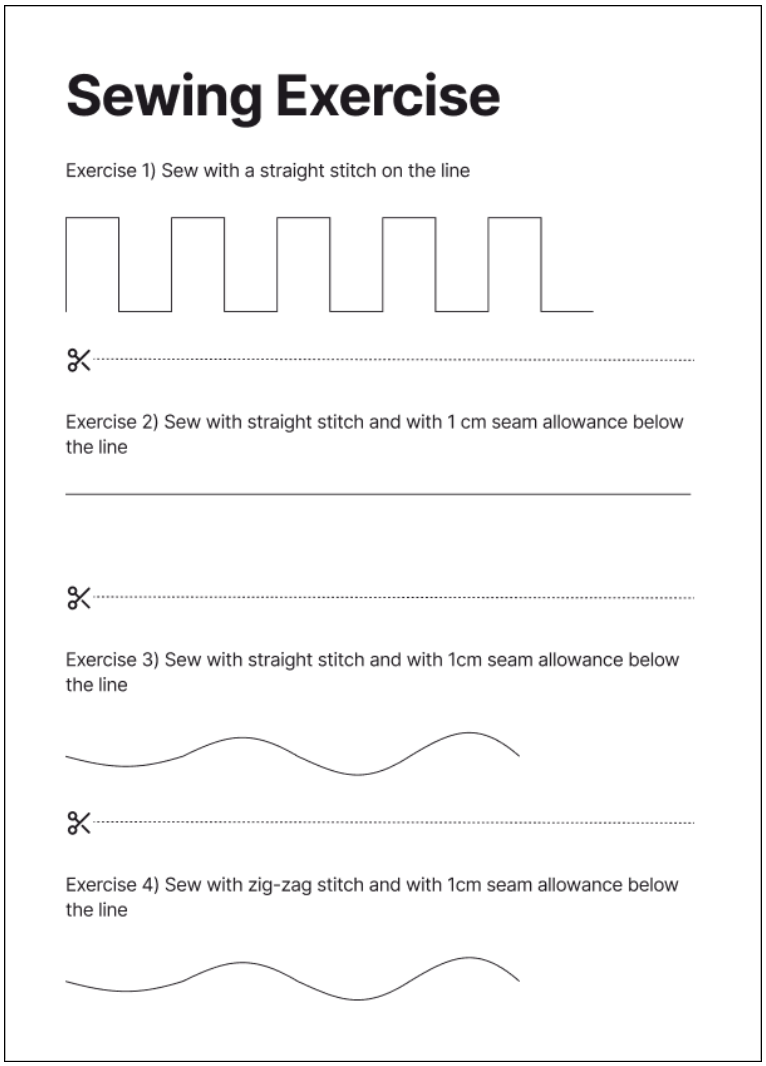

Figure 4.1: Example instructions for sewing exercises with the *preprinted lines*.

With this information, we can filter the videos, matching how well they can follow lines in sewing. So, if the sewing of users is not yet so precise, we could suggest projects where it is not of high importance to sew exactly straight seams. For example, an oversized garment is more forgiving of mistakes than one very close to the body. If users sewed very close to the prescribed seam, we could suggest detailed projects requiring good handcrafting skills. On the

system side, this prototype requires metadata for each tutorial about the complexity of the seam quality in tutorials.

Requirement of having some hardware and sewing material

Besides the device where users consume the tutorials, we require them to have a sewing machine, a printer, and a mobile phone. Furthermore, they need some dark thread, scissors and a sheet of white paper. We expect it takes about 20 minutes for the assessment process.

Figure 4.2: Scanning the sewing exercise result.

Users need to print out the exercise, sew it and scan and upload the results with their phones

First, we ask the user if they want to set up an assistant for selecting tutorials. After they confirm, they will be informed about the process and all the requirements. In the next step, we prompt the user to download and print the sewing exercise file (see Figure 4.1). Next, the users take the printout to their sewing machine and do the exercises described in the document.

To do this, the users have to sew according to the instructions and the preprinted lines. For uploading their results, the users have to scan a QR-Code with their phone (see Figure 4.2), which will lead to a web-based online app for the phone to scan the exercises. After the users scan all exercises, they can upload them, and the users return to the web browser.

Next, we inform the user that we highlight our recommendations with a green frame and a symbol. The background of this is that we want to ensure that the user is in control, so, by default, we show all available tutorials, but we give a visual cue to the ones recommended based on their skill assessment. Users interested in understanding the evaluation can examine the deviation's graphical representation (see Figure 4.3).

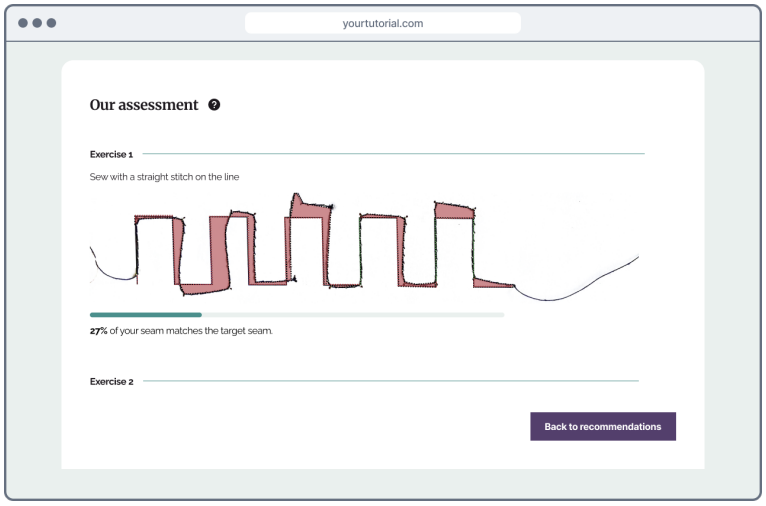

Figure 4.3: Presenting the difference between sewed line and preprinted line.

Back in the overview screen with all tutorials, the users can apply filters to see only the tutorials we recommend based on the skill assessment.

We present all screenshots of this prototype in the Appendix A.2.

4.2 Carla's Sewing Studio (PT2)

We assess the knowledge of the user and match the determined skills with the skills that are required for the tutorial

The factor *knowledge* [F2] is the basis for the prototype *Sewing On Paper*. We learned in the interviews that some courses of the universities of applied science test sewing-related topics in written exams. This type of examination inspired us to develop a quiz game concept. The questions in the quiz assess knowledge about handling sewing machines, characteristics of fabrics, and some tailoring topics. Developing the questions is not part of this work and should be carefully designed by people proficiently in sewing. The questions should map some practical sewing skills and express what a person is capable of sewing. To implement the system, we need to add meta-data to each tutorial which contains the required skills to complete the project. The quiz results are mapped to specified sewing skills, so we get a skill profile of the user. Finally, we can filter the tutorials based on the user's knowledge.

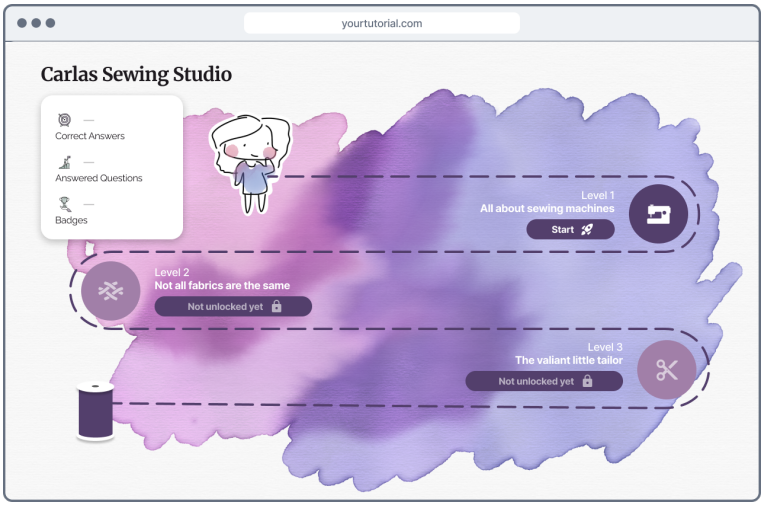

Figure 4.4: Users go through different levels with quiz questions covering different topics.

We used gamification elements to motivate the user completing the questionnaire

To motivate users to do the questionnaire, we used some gamification elements. Deci and Ryan [2012] explained in the *Self Determination Theory*, the fulfillment of the three basic psychological needs for autonomy, competence and relatedness are crucial for human's motivation. To motivate people doing the quiz, we use, like Sailer et al. [2017] de-

scribed, badges to fulfill the user's need for competence and added a story with the fictional character *Carla* to address the need for autonomy and relatedness.

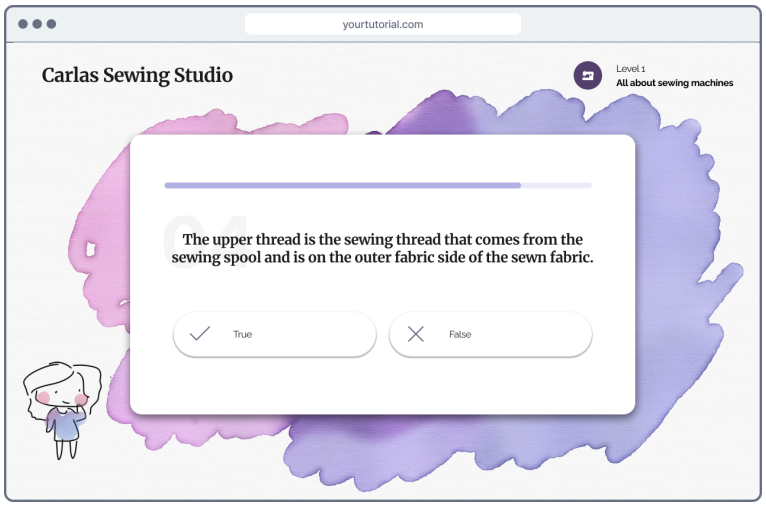

Figure 4.5: Example question to find out whether the user is familiar with using a sewing machine.

No additional hardware is needed, and users can do the quiz game in any browser.

Users answer the questions of each level

As shown in Figure 4.4, Users play various levels each dealing with a different topic. In the game, the user has to answer questions (see 4.5. For each level, the user earns a badge. After completing all levels, the flow is similar to the one described in Chapter 4.1. The system presents the highlighting of the tutorials, and the user can proceed to the tutorial overview, which now can be filtered by the user's skill level.

We present all screenshots of this prototype in the Appendix A.3.

4.3 Already Done Projects (PT3)

Gather user's completed projects and match tutorials which requires similar skills.

Based on the idea that sewing skills are comparable to a toolbox as in the project-driven approach, we designed a prototype, which evaluates the available skills with the projects a user completed. With this prototype, the user has to do a criteria-based self-assessment of their own sewed work, focusing on the processing of their completed projects. The criteria of this self-assessment is the complexity level of the processing. With this information, we can derive the skills the user must have had to complete the project. We consider this prototype as guidance for a improved self-estimation of a user's skills, as this requires the user to evaluate their success with completed projects.

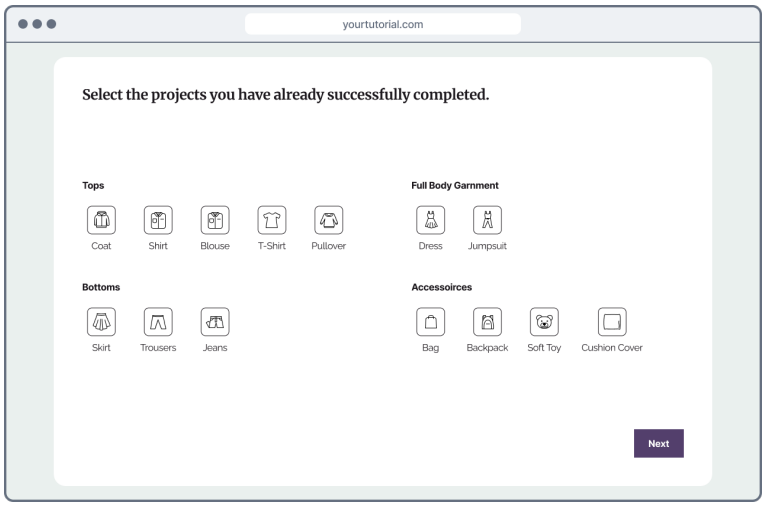

Figure 4.6: A selection of project categories from which users can choose the ones that they already have realized.

Except for a device that can run a browser, the user does not need additional hardware.

Users enter categories of completed projects and specify their complexity

Figure 4.6 shows how the system first prompts users to enter all categories of projects they have already completed. For example, this can be Bags, Trousers, or Blouses. Following, they have to specify the complexity of the items they sewed for each category. These complexity levels are following:

Easy. A project with fabrics that are easy to handle, e.g., cotton, doesn't have any accessories.

Medium. A project with fabrics that are easy to sew with accessories such as zippers.

Hard. A project with challenging fabrics, like leather or silk, and accessories, like zippers or buckles.

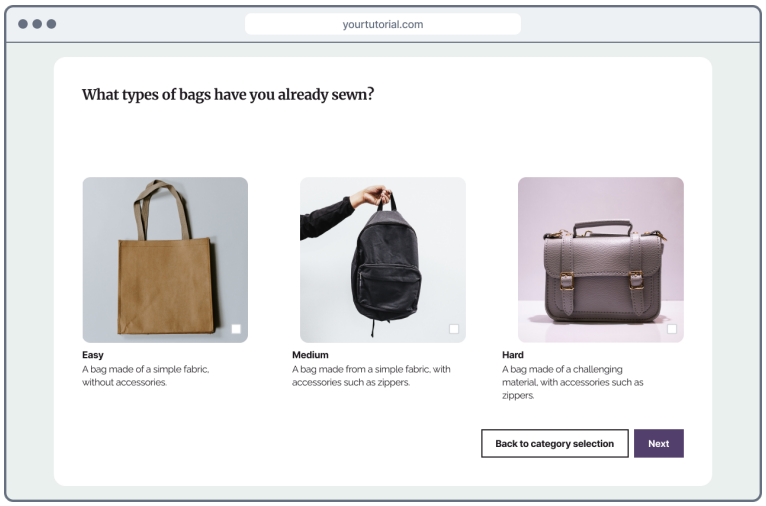

Figure 4.7: Specifying the complexity level of a completed project.

As presented in Figure 4.7, we provide some visual guidance with representative pictures to simplify it for the user to decide in which category their projects belong. After users assess the complexity of all project categories, the user flow will be the same as described in Chapter 4.1. First, the system explains the recommendations. Then, the user can browse through the highlighted or filtered tutorials.

We present all screenshots of this prototype in the Appendix A.4.

Chapter 5

User Study to Evaluate the User Acceptance

To find out if users would integrate our proposed prototypes into their sewing hobby, we tested them for user acceptance. In specfic, we want to answer following research question:

RQ2.1 Which of the prototypes is significantly preferred or rejected by hobby sewers?

RQ2.2 To what extent would hobby sewers integrate the prototypes into their hobby?

5.1 Methodology

To answer these questions, we set up a mixed-methods study in a within-subjects study design, where we let hobby sewers interact with the prototypes. In a follow-up, we will conduct a survey and a semi-structured interview.

5.1.1 Participants

We recruited participants who are hobby-sewers or who can imagine to learn sewing

We recruited N=12 participants (nine female, two male, and one non-binary person, aged from 18 to 34 years, see Table 5.1) through viva voce and a call for participants in maker communities. We interviewed two persons living in the USA and ten living in Germany. We translate citations from German participants into English. Our target group consists of persons who are hobby sewers, or persons who can imagine learning to sew. To know which skill level the participants have, respectively, we asked them to categorize themselves as *Novice*, *Advanced Beginner*, *Competence*, *Proficient*, or *Expert*. One interviewed person is a *Beginner*, three are *Advanced Beginners*, six of the participants estimate themselves to be at the *Competence* category, and two identify themselves as *Proficient*. None of the participants categorized itself as *Expert*.

5.1.2 Procedure

Participants interact with the click-dummy prototypes

During the 60 minutes remote-study, we provided the prototypes as click-dummies via remote control. As each participant will test all three prototypes, we expected the first discussed one to receive more feedback, so we rearranged the order for each participant. We asked the participants to run through an exemplary assessment of their sewing skills for each proposed concept. During this, we observed the participants and used a mixture of the retrospective and concurrent *Think Aloud* method, which is described in Ericsson and Simon [1993].

We gathered quantitative data with a questionnaire

Subsequently, we presented a questionnaire, which the participants filled out independently. The questions cover different criteria in order to evaluate if hobby sewers would use the proposed systems. For evaluating user acceptance, we used elements of the framework *Technology Acceptance Model* (TAM) by Davis [1989] which is designed for testing live systems. As Steinberg and Riggle [1995] indicate that prototypes can be validated with TAM we include questions about the perceived usefulness and ease of use of the TAM in our questionnaire. Additionally, we ask the partic-

Participant ID	Tutorial usage frequency	Sewing competence level	Frequency of noticing during a tutorial that it does not match own skills	How long the person is sewing	Age	Gender
P1	Never	Competence	Never	1–2 years	25–34	Male
P2	Never	Novice	Sometimes	Less than one year	18–24	Female
P3	Sometimes	Competence	Sometimes	More than 3 years	25–34	Female
P4	Sometimes	Proficient	Never	More than 3 years	25–34	Female
P5	Sometimes	Competence	Sometimes	More than 3 years	25–34	Female
P6	Always	Advanced Beginner	Sometimes	1–2 years	25–34	Female
P7	Rarely	Advanced Beginner	Sometimes	More than 3 years	25–34	Female
P8	Rarely	Competence	Sometimes	1–2 years	25–34	Female
P9	Often	Proficient	Rarely	More than 3 years	25–34	Non-binary
P10	Rarely	Competence	Sometimes	More than 3 years	25–34	Female
P11	Sometimes	Competence	Often	More than 3 years	25–34	Male
P12	Often	Advanced Beginner	Often	1–2 years	25–34	Female

Table 5.1: Participants of user acceptance study.

ipants to what extent they think the system is sound in assessing their practical skills. Next, we gathered data about the participant's trust in the skill assessment and the resulting recommendations. We built up the questionnaire with statements about the measured variables (see Table 5.2) and provided evaluating options on a bipolar five-point Likert scale, ranging from *strongly disagree* to *strongly agree*.

After the participant interacted with a prototype, we asked them to answer the questions.

In a subsequent interview we asked about their impressions and addressed findings of the questionnaire

Once the participant had evaluated the prototype with the questionnaire, we conducted a semi-structured interview. We asked about their first thoughts about the system and to what extent they think the displayed results will match their skills. Additionally, we addressed some questionnaire answers, and we specifically asked about responses that were outliers. After evaluating all prototypes, we asked for general data like their sewing proficiency, tutorial usage, and demographical data.

The complete study design, including all questions of the questionnaire and the guideline for the semi-structured interview, is attached to the work in Appendix B.2.

5.1.3 Analysis

Qualitative and quantitative analysis

We will proceed quantitatively and qualitatively to investigate the above questions. To answer the question **RQ2.1**, we want to quantitatively analyze whether there are differences based on the given answers in the questionnaire. We will proceed in two stages to find an answer to the question **RQ2.2**. We will quantitatively analyze questionnaire responses. For analyzing the interview, we choose a qualitative approach.

For descriptive statistics and to calculate the median as central tendency, we mapped the answers options of the questionnaire to the range of $[-2; 2]$ where *strongly disagree* $\mapsto -2$, and *strongly agree* $\mapsto 2$.

Evaluation Criteria to Assess the Prototypes		
ID	**Evaluation Criterion**	**Explanation**
	Perceived Usefulness (TAM)	
[C1]	Faster	Finding a matching tutorial faster
[C2]	Easier	Finding a matching tutorials easier
[C3]	Useful	Usefulness in sewing hobby
	Perceived Ease of Use (TAM)	
[C4]	Easy to learn	Using the system is easy to learn
[C5]	Expected behavior	System behaves as expected
[C6]	Understandable interaction	Interaction with the system is understandable
[C7]	Easy to use	System is easy to use
	Trusting the system	
[C8]	Practical skills	System can assess practical skills
[C9]	Trust	Trusting the system
[C10]	Prefer recommendations	Prefer the recommended tutorials
	Feelings	
[C11]	Annoying	Annoying to use the system

Table 5.2: In the survey, we included a question for each evaluation criteria, and participants answered on a 5-point Likert scale, between *strongly disagree* to *strongly agree*.

Study Variables

In the survey we controlled or measured following variables for each evaluation criterion given in Table 5.2.

Independent Variable

- **Shown Prototype.** The prototype which we show the participants, takes values from PT1–PT3

Dependent Variable

- **Agreement.** The degree of (dis-)agreement with the statement takes values according to the 5-point Likert scale

Analysis of Significant Differences in User Acceptance

Statistical test to detect significant differences in user acceptance

To answer **RQ2.1**, which is to find out if one concept is strongly preferred or rejected by the participants, we will conduct a quantitative analysis of the survey responses. We will conduct a statistical test for each evaluation criterion [C1–C11] (listed in Table 5.2) to see if the central tendencies differ for one of the criteria.

We have related samples and three treatments in each test because each participant answered the same question for each prototype. We have two variables, the participant's opinion, which depends on the shown prototype. According to Sullivan and Artino Jr [2013] we treat the values of Likert items as ordinally scaled data. With these properties the *Friedman test* is an option to examine **RQ2.1**.

With this test we want to verify the following hypothesis:

> H_1 : The distributions of the dependent variable are different.

Analysis of User Acceptance

Using Thematic Analysis to analyze interviews

First, we analyze the data quantitatively based on the frequency distributions of the participant's responses. In the next step, we use, similar to the previous expert study in Section 3.1.3, the *Thematic Analysis* by Clarke et al. [2015] inductively and in a semantic approach. After we transcribed the interviews, we coded the data. Because the interviews followed a semi-standardized protocol closely, we coded this data with *Structural Coding* by Saldana [2015].

5.2 Results of the User Study

First, we will give the results of our findings about if users prefer or reject one of the prototypes. Following, we will present the findings of the interviews and the surveys.

5.2.1 Do Users Prefer or Reject One of the Prototypes Significantly?

Following, we will evaluate if one of the prototypes performs significantly better or worse for the different evaluation criteria.

Testing for Normal Distribution

Shapiro-Wilk test to test for normal distribution

To demonstrate that we should perform a test that does not require normal distribution, we select the *Shapiro-Wilk* test as a preliminary test to check if the test results are normally distributed. According to Razali and Wah [2011] this test works well on small sample sizes like ours.

As expected (comparison Leung [2011]), there is evidence of normality only in some of the data sets. Since most data sets are not normally distributed at $\alpha = .05$, we can use this information to select a suitable test for a central tendency.

We list the results of this test in the appendix of this paper in Table C.1.

Detect Differences of the Prototypes for the Evaluation Criteria

Friedman test to analyze differences in the distributions of answers

Since we have three connected samples for each dependent variable in the form of the different prototypes, we decided to use the *Friedman test*, which is a non-parametric statistical test to examine differences in central tendency.

As described by Sheldon et al. [1996], the ordinally scaled data and not normal distributed meet the *Friedman Test* requirements, and we can use the test to analyze our data.

Friedman test shows that two criteria differ

The *Friedman test* shows that only in the criteria *Easy to use* [C7] ($\rho = .005, n = 12$) and *Annoying to use the system* [C11] ($\rho = .032, n = 12$) is a significant difference (at the significance level $\alpha = .05$) between the evaluated prototypes.

However, based on this test, it is impossible to determine which of the three prototypes differs. It would therefore be possible that only two of them differ significantly and that there are no significant differences between the others.

We attached the results to this work in the Appendix C.2.

Post hoc Test

Dunn-Bonferroni tests as Posthoc tests

As a posthoc test, we perform a pairwise comparison to evaluate which prototypes differ for the criteria *Easy to Use* [C7] and *Annoying To Use* [C11]. According to the *Dunn-Bonferroni* tests generated via the SPSS output, see Table 5.3, after *Bonferroni-Correction*, no statistically significant difference appears in the pairwise comparisons.

Pairwise Comparisons					
Sample 1 - Sample 2	**Test Statistic**	**Std. Error**	**Std. Test Statistic**	**Sign.**	**Adj. Sig.**[a]
Easy To Use					
PT1 - PT2	0.75	0.408	1.837	0.066	0.199
PT1 - PT3	0.875	0.408	2.143	0.032	0.096
PT2 - PT3	-0.125	0.408	-0.306	0.759	1
Using the system is annoying					
PT2 - PT3	0.625	0.408	1.531	0.126	0.377
PT3 - PT1	-0.75	0.408	-1.837	0.066	0.199
PT2 - PT1	-0.125	0.408	-0.306	0.759	1
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.					
Asymptotic significances (2-sided tests) are displayed. The significance level is $\alpha = 0.050$.					
a) Significance values have been adjusted by *Bonferroni-Correction* for multiple tests.					

Table 5.3: SPSS-output of pairwise comparisons for criteria *Easy to Use* and *Annoying to Use the System*.

No significant differences

Conforming to this, we can not reject the null hypothesis for the criteria *Easy to Use* and *Annoying to use the System*. In conclusion, we do not find a proof that H_0 is invalid for any dependent variables, we must assume that it still holds. Thus, our statement H_1 cannot be proven.

5.2.2 Findings in the Surveys and Interviews

Results of Thematic Analysis

In the following, we will describe the interviews and survey findings for the prototypes PT1–PT3. First, we will report the findings related to all of the prototypes. Following, we will first present for each prototype the descriptive statistics of the survey results and the results of the qualitative analysis according to the *Thematic Analysis* by Clarke et al. [2015].

Below, we list the codes that we were able to identify in the TA. Some of these correspond to the evaluation criteria we noted at the beginning in Table 5.2.

```
THEME: Perceived usefulness
      SUBTHEME: Effort
```

```
            SUBSUBTHEME: Faster [C1]
            SUBSUBTHEME: Easier [C2]
      SUBTHEME: Useful [C3]
THEME: Perceived ease of use
      SUBTHEME: Easy to learn [C4]
      SUBTHEME: Expected behavior [C5]
      SUBTHEME: Understandable interaction [C6]
      SUBTHEME: Easy to use [C7]
      SUBTHEME: Understanding the Assessment
THEME: Expressiveness of sewing skills
      SUBTHEME: Practical Skills [C8]
THEME: Trust
      SUBTHEME: Trust [C9]
      SUBTHEME: Prefer Recommendation [C10]
THEME: Feelings
      SUBTHEME: Annoying [C11]
```

Sewing on Paper (PT1)

First, we will give an overview of the descriptive statistics of the survey about *Sewing on Paper* (PT1). Second, we will present the results of the user interviews.

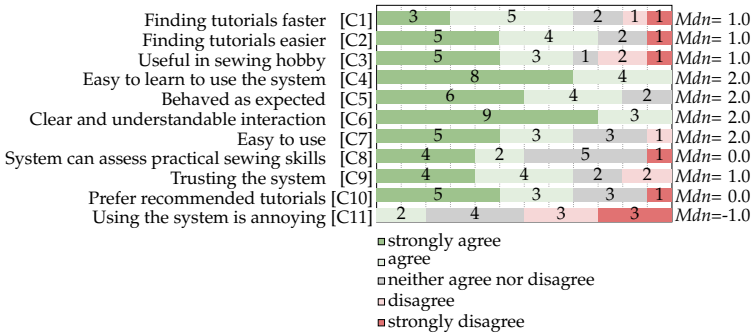

Figure 5.1: Distribution of answers for prototype *Sewing on Paper* (in total numbers).

The statements were rated at least over $\sim 66\%$ for nine of eleven evaluation criteria in the prototype's advantage

Descriptive Statistics Figure 5.1 shows the distribution of answers in the survey for the *Sewing on Paper* prototype for each evaluation criteria. In the appendix, we list the

explicit statements for these criteria in Table B.2.7. The participants answered on a 5-point Likert scale ranging from *Strongly Agree* to *Strongly Disagree*. The numbers given on each bar show the responses in total numbers, respectively. We did this survey with all of the 12 participants.

For criteria C1–C7, at least 50% responded with *Strongly Agree* or *Agree*. It is noticeable that two participants disagreed, and one participant strongly disagreed with the usefulness [C3] of the prototype in their sewing hobby. No participants (strongly) disagreed with the statements in criteria *Easy to Learn* [C4], *Behaved as Expected* [C5], and C*lear and Understandable Interaction* [C6]. One participant disagreed that the system is *Easy to use* [C7], whereas eight participants agreed or strongly agreed. Five participants could not decide whether the prototype evaluates *practical sewing skills* [C8] well; one person also stated that they strongly disagreed. Eight respondents *trusted the system* [C9] and *preferred to use the recommended tutorials* [C10]. In contrast, two people would rather not trust the system, and one strongly disagreed with preferring the recommended tutorials. Also, four people were neutral on whether it is *annoying to use the system* [C11], and two agreed it is annoying.

Findings from the Interviews

Expressiveness of Sewing Skills Participants have discussed how expressive the prototype's assessment mechanism is in evaluating their sewing skills [C8]. This mechanism evaluates sewed lines on paper, comparing the user's sewed line with the preprinted lines. Participants told us what they think about the user input, if straight sewed seams represent sewing skills, and if sewing on paper is comparable to sewing on a fabric.

The input in the system demands the user to perform a sewing exercise and upload it for the assessment. Five participants emphasized this type of input because it is purely practical.

Participants highlighted that they enter a practical work sample into the system

> P7: "I actually liked that the exact sewing skills, so the skills themselves are evaluated."

Because of this type of practical input, four participants liked the idea that the system evaluates sewing skills very objectively. They felt that this is a "safe method" (P6) because it works with "numbers, data, facts, and you feel like there is (...) something happening in the background." (P6)

Producing straight seams is not representative of general sewing skills

However, participants noted that perfect seams are only a tiny part of sewing skills and do not represent general sewing skills.

> P9: "I feel like sewing seams are only a small part of the things that I struggle with in sewing. (...) So I think basically to me, the hardest parts of sewing are, for instance, if you have a curved seam and you have to attach those two together, and then the curves are opposite, right? Like we have to attach the fabric together and try to be as straight as possible and not get the fabric to buckle."

Moreover, another participant added that they think cutting, ironing, and how someone treats the sewing machine are also a part of sewing skills.

> P5: "That you know your sewing machine, that you know, for example, which presser foot to use and when. That you can somehow cut properly and that you can iron properly. Ironing is actually a relatively large part of sewing. Yes, all in all, I do not know; I think spatial thinking is also often important."

Hobby sewers endeavor different quality standards of their seams

Other participants view the evaluation mechanism with skepticism, as they assume that not all hobby sewers aspire to produce straight seams.

> P3: "You are still proud when you have sewn something, even if it is not one hundred percent straight or something. Because you made it yourself, and just because the corner is a little crooked or something, it does not really make any difference to you."

On the other hand, others claimed that there are tutorials where it is essential to be able to sew straight.

> P7: "Especially because there are these different instructions or there are different sewing patterns for which it is important that this is sewn exactly."

Users might put in extra effort to make it correct or just have good manual dexterity

Another opinion participants formed about this assessing mechanism was that users might also put extra effort into completing the tasks exceptionally well.

> P7: "I can imagine that there are also people who try to work as precisely as possible when it comes to the seam allowance, which they might not do otherwise. Especially when you have a longer sewing project, I always find it exhausting to work very long, very precisely. And then I can imagine that maybe on the paper, you either draw in the seam allowance or measure it or something."

This opinion resembles the statement of another participant, who reported they accomplished such a sewing exercise as a beginner in a sewing course very well, without having any sewing skills. Based on this experience, the participant claimed that this prototype is more suited to test manual dexterity than sewing skills.

> P12: "I am a dead beginner sewer, and I still have my paper because we started with paper exercises like these from day one of my sewing class. And they were very good. But I was a

> beginner. And so I knew nothing. I knew what the parts of the sewing machine were. And I was good with my hands. But I was not anything except a beginner. And so I think that people who are, you know, have good hand-eye coordination and understand how to, like, rotate the crank on their sewing machines and put the needle down and pivot like they are going to do great. They are going to do great on this. Especially people who have, you know, like, seen their little magnetic plates that mark your seam. (...) So I don't think you will necessarily get accurate assessments of people's practical skills."

The system does not take sewing knowledge into account

Another issue users raised was that this system does not ask for knowledge, which could lead to the system suggesting tutorials for which users lack an understanding of terms.

> P3: "It can possibly assess the practical a bit better. But now it can happen to me that I get a video where they use all the professional terms I've never heard. But I can sew super nicely straight."

Concerns about how much sewing on paper represents sewing with fabrics

Regarding users having to sew on paper instead of fabric, five respondents expressed concern that paper behaves differently than fabric.

> P9: "I worry about how much it can generalize to other materials used in sewing because most of it is about using different materials."

Further, P9 described that processing leather, silk, or muslin leads to own unique challenges, respectively. In addition, P1 explained that paper would not slip or snag in the sewing machine.

Participants had different opinions about usefulness

Perceived Usefulness Some participants considered the

system useful [C3] in their sewing hobby. One participant explained that it added more value than assessing the straightness of the seams by themselves.

> P6: "Just that you can actually put your practical skills into a tool like that that is not subjectively assessed, 'What do I think I can sew right now and how good are my stitches?' But I just made it as good as I could. And then I upload it, and it just gets looked at how straight they are. And definitely gives the feeling of, 'Okay, there's really something being checked,' and gives me added value than I just assessed, "Yeah, I can sew straight already."

Other participants opposed the system's usefulness because they estimated that the result was inaccurate.

> P9: "I ranked that the one that I like the least because I do not know if it would give me an accurate outcome."

Participants rated the effort differently

Participants rated the effort [C1][C2] to operate the system and make and upload the sewing exercise differently.

Two participants thought they could do it quickly, and P11 phrased this as not having to "produce anything gigantic, big."

In contrast, five participants criticized that it is much effort to do the exercises and upload them.

> P5: "I think it is quite a lot of effort to do that because you do not get that much-added value out of it in the end."

In addition, a participant remarked that doing the assessment is a hurdle to getting to the tutorials.

> P3: "So I just can not now quickly look for a tutorial if I have to do it before. Well, I suppose you do not have to do it repeatedly. But yes, it is already a hurdle."

Interaction with the prototype was clear and understandable

Perceived Ease of Use For users P1 and P8–P12, there were no problems interacting [C6][C7] with the prototype; the user flow was "super understandable" (P1) and behaved as expected [C5].

Regarding the process that the users scan the sewing samples with the smartphone and upload them directly, some of the participants said the "seamless integration is enjoyable to use" (P12)[C7], or they liked the fact that they did not have to use a physical scanner device (P10)[C7].

However, two participants view the process of uploading the sewing exercise with their smartphones with skepticism. One participant finds the user flow unusual, and another assumes that there could be people who would have trouble using it [C7].

For another person, the user flow is confusing. In their opinion, it was unclear what would happen, and they did not expect [C5] that the system would ask for a sewing sample.

Participants criticize the required devices

Seven participants criticize the requirement of having a printer.

> P7: "I would consider it a disadvantage that you need three devices. (...) It is the device with which I look at [the website]. Then I need a printer, and I have to scan it with the smartphone".

However, another participant argues that people who sew and use tutorials tend to print out patterns and, therefore, have probably access to a printer. Since P11 did not have a sewing machine, they would have sewed it by hand. However, P3 noted that sewing by hand on paper does not work

well because the needle does not go through the paper easily and it sometimes tears.

Not every participant could understand the assessment

We noted some statements about how participants understood the assessment in the interviews.

Two participants mentioned that they liked the visualization to show the deviation of the sewing result to the *preprinted lines*. Two participants liked the comment on the tutorial describing why the it is (un-)suitable for their sewing result. One participant was surprised that the system evaluates the seam in the actual sewed line differently than they had seen it by eye.

Participants had issues understanding why a tutorial is (not) recommended as matching to their sewing skills.

> P3: "Well, I guess just because I do not understand the connection with the perfect straight seams and why I cannot sew the dog bag. (...) It does not say, 'Okay, you cannot sew the bags because you do not know how to sew in a zipper or how to sew these pockets on,' or anything like that. It just tells me, 'Your seams were not straight enough right now.' And that is the way it is. That wouldn't stop me from trying it anyway. Because if I want to sew them now, then I will just sew them crooked."

More participants reported trusting the system than there were reports of distrust.

Trust Trusting the system is about whether the participants generally trust it [C9] and if they prefer the tutorials recommended based on the skill assessment [C10].

One participant has the impression of being graded by a professional.

> P8: "Making the sewing exercise, I find itself somewhere good because you had the feeling that maybe a professional sits on the other side, which then looks at your sewing exercises, for example, and evaluates."

Some participants would prefer the recommended tutorials, others not

Three participants said they prefer to watch the recommended tutorials [C10] as the result of the skill assessment is considered as a proof for insufficient sewing skills:

> P2: "I probably would prefer to look at the marked tutorials because I was proven to need a little more practice right away."

A participant told us they would choose the recommended tutorial from similar ones that give an introduction to the same piece of clothing.

> P1: "But if you imagine now, for example, I am looking for a tutorial (...), which would be about the same piece of clothing. Then there is then also an evaluation of it. So I would definitely take the tutorial, so to speak, where the system would recommend that to me, based on the prototype, first."

However, another participant does not prefer the recommended tutorials because the level of complexity is not a filter criterion for them.

> P4: "Would you prefer to watch the highlighted tutorials? Nope. Well, but I would not do it by difficulty."

Feelings Participants reported how they felt using the prototypes.

Some felt supported, others felt demotivated

A participant emphasized that they felt very supported by the system because

> P2: "With tutorials, you do not interact with anyone (...), and the [system] gives you a bit of a feeling of this, that I am really being helped now."

On the other hand, a participant noted that the system could be demotivating for people who do not have the demand for perfect seams.

> P3: "So I can imagine that it may also be a bit demotivating, especially as a beginner or advanced but not yet a professional... You are still proud of it when you have sewn something, even if it is not one hundred percent straight or something because you made it yourself, and just because the corner is a little crooked or something, it does not make any difference to you, actually."

In some cases, the prototype would be considered annoying

In the questionnaire we asked, if participants consider the prototype as annoying. Some of the participants commented on the question: "I do not find it annoying at all" (P11). In contrast, P4 points out that not everyone would "get the sewing machine and start sewing" but wants to explore the tutorial website first and then decides to do the assessment. For P9, the system would be annoying if it would force them to use the assessment before they get access to a tutorial.

Carla's Sewing Studio (PT2)

In the following, we will review the descriptive statistics on the survey analysis of *Carla's Sewing Studio* (PT2). Then we will report the results of the qualitative analysis of the interviews.

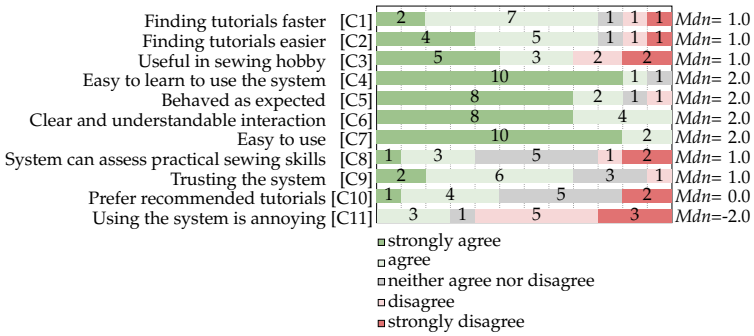

Figure 5.2: Distribution of answers for prototype *Carla's Sewing Studio* (in total numbers).

The statements were rated at least over $\sim 66\%$ for nine of eleven evaluation criteria in the prototype's advantage

Descriptive Statistics A distribution of the given answers to the questionnaire on *Carla's Sewing Studio* is shown in Figure 5.2. The figure has the same structure as for PT1 and shows the frequencies of the answers for each evaluation criterion given in total numbers. For criteria C1–C7, at least eight participants agree or strongly agree with the evaluated statements. Four participants do not consider the *system useful in their sewing hobby* [C3]. For the evaluation criteria *Clear and Understandable Interaction* [C6] and *Easy To Use* [C7], all participants (strongly) agree. Five participants neither agree nor disagree whether the system can assess *practical sewing skills* [C8], whereas three participants (strongly) disagree and four (strongly) agree. Similarly, for the *Prefer Recommended Tutorials* [C10] criterion, five participants neither agree nor disagree, while five participants (strongly) agree and two strongly disagree. Eight participants disagree with the statement that the *system is annoying* [C11], whereas three participants consider the system as annoying to use.

Findings from the Interviews

Expressiveness of Sewing Skills We asked users about their agreement with the statement *The system can assess practical skills* [C8]. Based on this question, participants discussed whether the theoretical knowledge about sewing, which PT2 assesses, represents having sewing skills.

Criticism that the system can not evaluate practical skills

In the interview, participants pointed out that the prototype evaluates practical skills inaccurately. People could be theoretically proficient in knowledge about sewing but have deficits in applying theory into practice.

> P1: "About 'I think the system is good at evaluating my practical skills': Of course not! Because I have not done anything practical. So I can also read through a Wikipedia article and then know everything, so to speak. But it has nothing to do with whether I can operate a sewing machine or not. That's why I would say that practical ability is of course, very difficult. So then I would say, yes, I cannot get anything out of the system, so to speak, as far as practical skills are concerned. "

In addition, users were wondering about how questions can be formulated so that they map practical skills:

> P7: "I also do not know how you can meaningfully map the skills, that is, what you learn when you sew, how you can map that in questions like that. So I do not think that is an appropriate type of exam for that."

Suggestion to add questions that ask for mastered techniques

However, another participant proposed to formulate questions in such a way that the system can find out which techniques people already mastered, and the system could use this information as a filter criterion:

> P5: "Well, I mean, it was only two questions, but I think from this theoretical knowledge, you can estimate what someone has already done. So especially if you ask good questions about certain techniques or something like that, you can find out if the person has already done that. And then you could even not only somehow level beginners, advanced, professional or so, but can even say: "Okay, here is a zipper sewed, you have never done: This could be a challenge for you." That you really filter it by different techniques or so."

Terminology is important for understanding tutorials

On the other hand, however, users explained that sewing knowledge is also essential. For example, it often takes an understanding of terms to rework tutorials.

> P4: "And it is mostly about terminology in our sewing tutorials, whether it is a video or a text. Terms are important because it is about communicating (...) and that is not just visually."

Users could guess (correct) answers

Two participants also pointed out that it is possible to guess answers. More specifically, in the questionnaire there was no option *I do not know* given which would tell the system that the participant has no knowledge about a topic. Therefore, some participants guessed their answer.

> P6: "Well, I guessed the two questions blindly, to be honest, because I do not know either, and I do not know where the upper thread comes out at the top or the bottom, but I know how to thread (...) the sewing machine and I have no idea where it comes out, but in any case, a straight seam comes out at the end. So a little bit: I think I would just rate my practical skills higher than what I know theoretically. And with such guessed questions, especially if it is true/false, you have a fifty-fifty chance...It is the question of whether that then reflects your knowledge in this case."

Perceived Usefulness We asked several questions about the *Perceived Usefulness* of the survey [C1–C3], and the participants reported their opinions about the system's usefulness and the effort to use the system.

Helps user to find out what sewing level their skills are

One participant mentioned that understanding and mapping their sewing skills to a difficulty level of a tutorial on conventional tutorial websites is difficult, and our system would give guidance to understand what level they can be assigned to.

> P5: "So I found the idea quite cool, because if you are otherwise on a website, then you have there a difficulty of, no idea, 1 to 5 stars but no relation to it. So if I do not know — have already struggled at some point to make a coat and have already worked out a few things, then I would be better guided by such a question system because I already know certain things, like when you say that is for beginners: 'Yes, am I still a beginner or not?' I find that difficult. So I find something like this, that certain things are asked, quite cool."

System is entertaining

For other users, however, the system offers no added value, but they would use it for entertainment.

> P6: "I feel that it does not bring me that much-added value, that it is such a fun factor. I do a quiz and find out what percentage I can already sew, not really as support."

Can be used educational

Users pointed out that they found the system's learning effect to be good, as the system evaluates the answers to questions and provides feedback.

> P9: "It is like if I do not know this, now I know this. Like now I have improved my knowledge in the sense of like, you know, what is relevant to be able to make a shirt successfully."

Participants discussed the effort to use the system

Further, participants gave their estimations on how effortful it is to use the system. One person noted that it is convenient that only an internet browser is needed for the evaluation.

> P9: "Like it is easy in that you do not have the context switch and go to a sewing machine."

Another participant noted that it would end up taking a lot of time to go through all the questions:

> P7: "Yes, and maybe also if that per category, if you then have the three categories or more and per category ten questions is at least not so super little. It would take a relatively large amount of time again."

None of the participants mentioned that it was annoying to use the system.

Perceived Ease of Use With criteria C4–C7, we asked about *Perceived Usefulness* in the survey. Participants reported in the interview whether they found the interaction with the system clear and understandable, and whether they understood the assessment by the system.

Interaction mainly clear and understandable

Participants emphasized that the interaction with the system was clear and concise. One person would not have expected to receive feedback directly.

> P12: "I do not think I expected immediate feedback on whether a question was answered correctly or not."

Yet, participants expected an analysis of the results at the end along with a classification of their skills into a level.

> P12: "And I also expected to be given some kind of analysis of the questionnaire results. Like based on your answers to these questions, your skill level in sewing machine knowledge is ranked as beginner, intermediate, advanced or whatever it is."

Understanding the assessment

One person had the impression that the assessment was arbitrary and pointed out that for some of the presented tutorials, the explanation of why the video is unsuitable for the user does not make sense in terms of sewing logic.

> P9: "Like, it seems sort of arbitrary what the judgment is. So. Yeah. And then, okay, another thing is like, how does this map like why is flared skirt like the sewing machine category while like dog bag is like tailoring like I actually feel like you do no tailoring with the bag because you do not have a body to fit it in. It is just sort of like a bag with set dimensions. So I would feel like this would be more about like fabrics because it is probably a heavier fabric than like a t shirt. But then the t shirt has like the fabrics category result. So I think that is also something that is slightly confusing to me. Same with the cushion cover is like you do not tailor a cushion cover like this would probably be more for like sewing machine basics."

Trust For C9 and C10 we evaluated the user's trust in the system and the recommendations, and participants gave reasons for their rating in the interview.

Participants would give it a try and evaluate the result in a real system

Participants did not directly specify whether they would trust the system. Two participants stated that they would try out the real system and then compare whether the tutorials the system recommends fit their skill level. Then they are able to say more precisely whether they would trust the system.

> P11: "I am generally, first of all, of the opinion that someone has probably thought of something. I can try it out. If it does not work, I still have the option to click on all the other things. I can simply ignore the recommendation. But since someone will probably have put work into it to give me something useful, I will try that first, because in case of doubt, it makes my work easier, so why not? In so far, if it [the tutorial] is just roughly what I want to do, just click first. "

Distrust because of missing practical skill assessment

The concept does not convince other users because it does not consider practical skills.

> P6: "For me personally, this: 'Would I trust what comes out' — Rather not, because sewing is already something practical, and it is so now the question whether that comes out so well with such theory questions then."

Feelings In addition to the question we provided in the survey, whether it is annoying to use the system, some users also reported what they felt when interacting with the prototype. A topic that occurred repeatedly was the gamification used in the prototype.

Getting recommendations motivates filling out questionnaires

One of the participants stated that they generally enjoyed filling out questionnaires and getting recommendations:

> P2: "So, I mean, who does not like that, filling something out and then getting recommendations. I think that is always cool."

Gamification is evaluated both positively and negatively

We observed that some highlighted some benefits of the gamification concept of the system, whereas another participant found the concept unnecessary. Starting with the positive characteristics of the quiz, eight of the twelve respondents spoke positively about the gamified character of the questionnaire. Some indicated that they found the

Carla character "cute." Participants felt a personal connection to the character.

> P8: "What I definitely liked was quasi with Carla. I found that quite cute somehow because you have the feeling that you have a personal relationship, so to speak, because you simply saw this face of hers, (...) even if it was only a drawing. But it had something personal, and you had the feeling that more attention was paid to you. So it was not so — I would say anonymized, like now, for example, in the prototype one."

Additionally, others stated that it was fun to interact with the prototype.

> P5: "But I just thought the workshop was kind of cute, and it was fun. (...) I think the sewing workshop takes a bit longer, but somehow I thought it was funnier."

The same person stated that they were happy about the badge they received when they completed a level.

> P5: "I am kind of happy about that. I received the sewing machine badge. I am happy about that!"

However, one person also noted that they first had the impression it was addressed to younger people and then felt an ambivalence between the cute presentation and the complexity of the questions:

> P7: "I found that the design was very cute with the pretty colors, and so I found that so cute. I also just meant, (...) that I had thought in the first moment: Okay, there you can somehow put younger people or sewing beginners. But then I

> had the feeling that (...) the questions were very complicated or very technical."

In contrast to the participants who liked the approach, one person felt that the presentation as a game is unnecessary and that the questionnaire alone is sufficient.

> P4: "I just find that misleadingly expressed. I think it can be kept much simpler. And would not have to build it into a story like that. And also the three levels, I think that seems like a lot of effort somehow, does not it? Well, I do not want to play a computer game, and levels to me sound a bit like a learning app or a computer game or something. And I just want to quickly assess my sewing skills. And I think that is too many levels again — The Carla level and this level. I do not think it is necessary. And if the questionnaire consists of only nine questions and not only three. And you do not have any levels. Then you have clicked through it quickly anyway."

Possibility of performance pressure on users

Three participants expressed that they felt performance pressure when they took the test. However, two of them assumed they felt this because the interviewer observed them.

> P5: "So, what I also thought for a moment, maybe that is because of the situation now. If you watch me, but now already very briefly this performance pressure moment. Okay, now you have to answer correctly. But I think that is something else again when you are doing it alone. So I do not know. I know that when I do it alone, no one sees that. I do not know if every person who clicks through is aware that it is only evaluated in front of the computer and not by a person. Then I think it would not matter. But anyway, because you are watching me now, I had a short, stressful thought: 'You have to do this correct now.' "

Some are motivated by the system, for others it has the opposite effect

We noted that some participants had feelings about whether the system was motivating or not. There were different opinions on this.

First, the presented progress in the level overview, which the system showed, motivated participants to gain better skills.

> P2: "And I found that actually quite cool because then you can also recognize such progress. Of course, it motivates you to get better."

Besides that, a participant described that the system motivated them to answer all questions.

> P6: "So I think it looked nice, it was easy to use and motivated to just click through the questions, where otherwise, if you, I say if this 'Choose a dress, skirt, pants' has two or three more levels, you might get tired at some point and think: 'Oh no, come on, I will not do it.' And there is just like, there is more motivation to play through that completely and click through."

However, the system can be demotivating by pointing out wrong answers, mainly because the solution to a question comes right after submitting the answer.

> P12: "I almost wonder if it might be better rather than flashing red and green for the correct and incorrect answers to collect the data and ascribe a skill level once the person has answered the question and then give them the option to go back and look at their right or wrong answers. Because sometimes, that instant negative feedback, if you get the question wrong, can be very off-putting."

Further, we observed that one person who identified themselves as a beginner made negative comments about their own sewing skills while they were filling out the questions:

> P2: "You can tell how little I know about sewing".

Additionally, using the system could discourage people with no answers to the questions because they think the following content would not suit them.

> P7: "I think that the first question — I do not know, to what extent something like that is just built in prototype-wise or if that is actually a question that you would ask then. I do not know if a question like that about a type of sewing machine would not actually deter me from using the site. Because as I understand it now, it is its own type of sewing machine. And most people only have one. And then when people ask directly for different kinds of sewing machines, I wonder if people do not say, 'Okay, this whole site is above my demands and above what I can deliver. Maybe I will not do it after all.'"

User has to think about questions

One participant found that this is not a system they can thoughtlessly click through, as they have to think about the answers to the questions.

> P9: "Like it is easy in that you do not have the context switch and go to a sewing machine. But it is hard in that you actually have to think as opposed to just checking boxes of things that you have made."

Projects Done (PT3)

As with the previous prototypes, we will first examine the descriptive statistics and then present the findings from the interviews about the prototype *Projects Done* (PT3).

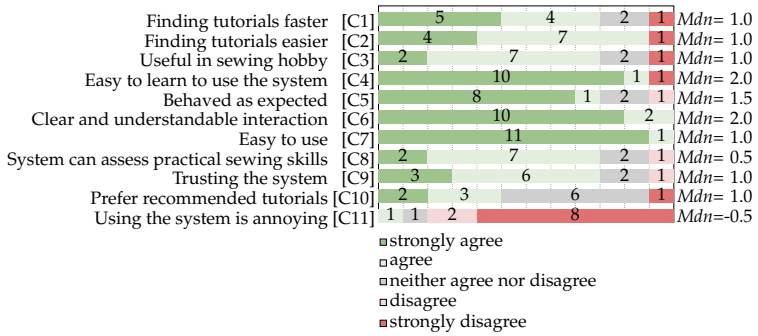

Figure 5.3: Distribution of answers for prototype *Project Done* (in total numbers).

Descriptive Statistics Figure 5.3 illustrates the distribution of answers on *(Strongly) Agree* to *(Strongly) Disagree* with the given statements for each evaluation criterion for prototype *Projects Done*. The figure shows that eight of the participants (strongly) agree with criteria [C1–C3] on *Perceived Usefulness*. One person strongly disagrees with the statements for each of these criteria. For the evaluation criteria of *Perceived Ease of Use* [C4–C7], at least nine participants (strongly) agree, and all participants report a *clear and understandable interaction* [C6] with the system and that the system is *easy to use* [C7]. One person reports that the system is not good at *assessing practical skills* [C8]. Also, one disagrees on *trusting the system* [C9]. Five respondents (strongly) agree that they would *use the recommended tutorials* [C10], whereas six people abstain, and one strongly disagrees. Ten participants (strongly) disagree that *using the system is annoying* [C11], whereas one agrees that it is annoying.

The statements were rated at least over $\sim 66\%$ for ten of eleven evaluation criteria in the prototype's advantage

Findings from the Interviews

Expressiveness of Sewing Skills In the survey, we asked participants whether the system is good at assessing practical sewing skills [C8]. In the following, participants discussed whether the query of already completed projects and their corresponding complexity level can represent sewing skills.

Skill Assessment depends on user's self-assessment of quality and complexity of sewn products

The estimation of this skills depends on how the user estimates their projects' quality level and complexity of sewn products by themselves. This self-estimation can lead to a wrong result because it is difficult for people to estimate themselves.

> P12: "I also think that it relies a lot more subjectively on what somebody considers to be a successful completion, and the lack of objectivity may lead to some inaccurate assessments of people. From my understanding of psychology, people tend to rate themselves higher than their actual skill level. We are not good at being objective about ourselves. We are not! No, nobody really is. Or almost nobody is. It is very difficult to be objective about your own skills and shortcomings. So I suspect that I might be rated at a higher skill level with this assessment system than I would in a more objective way of measuring."

Assigning completed projects to given complexity level is challenging

We also noticed that participants had difficulties assigning their sewn projects to the given complexity levels.

> P10: "I think it is difficult to somehow assign one of these three examples. It was displayed below: 'Yes, you once sewed a zipper into a bag?' wasn't it? Well, you can do what applies to that, but if you have to make a bag that somehow does not fit in with anything, I am not sure what you would rather choose: Whether you overestimate or underestimate it or whether it somehow does not fit at all. However, I think, such a first assessment, to say roughly: 'Okay, a jute bag is different from sewing a backpack.'"

Users might lie

Participants reported that users might not truthfully indicate which projects they have already completed.

> P9: "I could also lie to the system or say I made something I did not make."

Not all clothes in category list available

Another challenge arose for the participants when a sewed item did not match one of the system's clothing categories or if the item was uncommon.

> P3: "At the beginning, with the categories you have to tick, so T-shirt, Jacket, etc... I had the problem that I once sewed a very simple Kimono, and I did not quite know how to classify it."

When we asked how the users dealt with the fact that some projects were not listed, we learned that they tried to map it into a category on their own:

> P9: "I kind of mapped it. For instance, most recently, I sewed a pair of like wings that were kind of long, complicated, fancy sleeves. So I sort of like mapped that to the blouse thing. So I basically tried to make my best sense of the things that I had sewn before and give it to the system that way."

Participants do not expect an exact assessment of their skills

Participants think that it is difficult for the system to distinguish between detecting details in a wide range of evaluation possibilities, i.e., between absolute beginner and professional.

> P6: "So if the range is from 'I can sew a square with a sewing machine' to 'I have sewn some haute couture dress' it is probably harder to get the differences in between."

Another user stated that the system does not ask for enough details for a meaningful assessment.

> P9: "It does not have enough detailed information to be able to have that assessment"

For example, another user missed that the system does not ask how often they have sewn something.

> P7: I imagine it would tend to underestimate them. Not explicitly related to me, but people generally — by the fact that one has sewn some things more often and others less often.

Theoretical knowledge not taken into account

One user explained that the system lacked determining which sewing terminology a user could understand. Thus, the user expected that they get difficulties reworking recommended tutorials.

> P3: "Just because I can sew a cool bag, but I do not know the terms of the things at all. However, if I then watch a video that also sews a complicated bag, she probably uses these terms. Then I have a problem again because I do not know them. I can do it, but, mhm, difficult."

Participants think that similar projects are recommended.

Participants assumed that the system evaluated the experience gained in completed projects. Based on this assessment, they thought the system recommends projects that are similar to the ones they already did.

> P1: "And practically, one can already see: 'Okay, what projects has he/she already started or has he/she already completed?' And depending on that, there is an experience in the domain. Means somehow, yes, if I, for example, want to sew the same again — so I want to sew a new bag (...) and I want to sew that kind of bag from a tutorial or search for it. Then the system can — if I already think about what kind of bag I have already sewn? — of course, estimate a little bit: Okay, what level of experience is he coming from?"

Perceived Usefulness With criteria C1–C3, we evaluated the perceived usefulness of the system. Based on this questions, the participants discussed how much the system would be useful in their sewing hobby and rated the effort using it.

Participants evaluated the perceived usefulness differently. Some participants would like to integrate or try the system, whereas others stated that the system would not offer them any added value. One participant said that there is no skill level indication on a tutorial website they frequently use. They would appreciate using the system to find a tutorial matching the difficulty they can handle.

Helps to find tutorials that match skill level

> P9: "Like, again, in a sense, (...) cosplay-tutorial.com is a website that I use a lot. And then there are a bunch of different tutorials for things you want to make. And then I like to try to find the closest ones, but of course, they have no sort of like skill level indication. So I can totally imagine that this specific flow being integrated with that user flow in terms of me being like: Oh, I want to do some sort of like head armor, and then like being able to sort the tutorials by like difficulty or something and being: We think you are at this difficulty, so check out these nearby tutorials and stuff."

Other participants would at least try it out and then compare if the system works well, but tend to also look into tutorials that the system does not explicitly recommend to them.

Need to try it out, but a tendency to use not recommended tutorials

> P5: "I think most: 'Just try it out, and if it is not working out, then that is the way it is.' (...) All tutorials will continue to be displayed. But I would also not be afraid of making tutorials where it says: "Yes, that is actually too hard for you." But on the other hand, I was using many discussion boards already, which concern sewing and Facebook groups and stuff. And I know that there are many people who, I

> think, would be very happy if there was something like that. Because they always have a hard time estimating their level and are quickly disappointed when something does not work out. And therefore, I believe that such a system can be very useful."

No added value on proficient skill levels

One participant who identified themselves as proficient expressed concern that the system would not add value to finding tutorials. They assumed that if they rated every category with the highest rating, all tutorials would be marked as recommended.

> P4: "I think this prioritization with the green does not do anything for me. Because that — well, I am all about faster grasping. And then, I do not have that. So there are still hundreds of things and possibilities. So I scroll through it and look at it very briefly, and then I know whether it is something awesome or not. So the density of exciting things for me does not change. And I think I scrolled through it quickly with or without a green frame. And then, we have already talked about how it would make sense if everything was green at a more advanced level. And then it still does not bring anything."

Does not require much effort

Eight out of twelve interviewees emphasized in the interview that the system does not require much effort nor any additional hardware.

> P11: "This is, of course, relatively quick to do because I do not have to print anything out, and I do not have to have a machine there. In case of doubt, I can click through it on a tablet and watch the videos while sitting on the train or whatever."

Other participants said they wished they knew what level the recommended tutorials are.

> P12: "I found it a little bit difficult to see what the difficulty of the project was, but the tutorial level, other than, you know, a lot of them had boxes and this little icon saying that they were recommended for me."

Perceived Ease of Use We asked for the *Perceived Ease of Use* [C4–C7], and participants reported issues regarding expected behavior and rated their understanding in the assessment.

Complexity level assessment was not expected

Two participants expressed that they did not expect an assessment of the complexity level to come in the second step after the clothes category selection.

> P11: "That is not what I did expect to come at that moment. Then I thought it was good that it was there, but it was not like, I thought: 'Okay, you pick something, then the following category comes: What have you seen in seams or something?'"

The participant who tried to classify a kimono to the existing clothing categories described that it would have been easier for them to classify the kimono if they had already seen the classification of complexity in the category selection.

> P3: "Well, I do not know how the subcategories for the jacket would have been. Maybe the first level would have been a jacket that you just cannot close or something. Maybe that would have been the first level, and then it would have actually fit in well."

Understanding recommendations based on the assessment is rated contrary

One participant mentioned in the interview that they understood the assessment.

> P9: "Yeah, for sure. So, again like, I thought the flow was great, and it made a lot of sense.

> And I feel like it was a little more clear on how they recommended the tutorials to me this time around."

Another participant did not understand why the system did not recommend some tutorials. They also did not understand why the system did not transfer the performance in a completed category to other categories that were not completed yet, thus, recommending tutorials from other categories as well.

> P3: "I think it was a jumpsuit video, which was not recommended to me because I need more skills or something. I do not know what level this Jumpsuit is; I have no idea. But as I said, if I just say I can sew a medium complicated dress, then I think I can sew a jumpsuit or at least try. So there, I think I just did not understand why that is not recommended. But I think that is just because of this lack of transfer between the different projects."

Opposing opinions about trusting the system

Trust We asked to what extent participants would trust the recommended tutorials and the system. Participants were of opposing views on this topic. One participant stated that they would trust the system. The participant mentioned they would search for a project they had already marked as done in the initial project selection.

> P1: "I would then again trust the system's assessment, of course. As a matter of course, the system can match: 'Hey, he has already sewn a bag before. I have a video here showing how to sew a bag in the difficulty level he has already sewn. So this could fit for him.' "

However, others view the system with more skepticism. At least they would watch the recommended tutorials, but they would also consider other tutorials which are not recommended.

> P10: "So I think with the last system, I would most likely look at those that were not suggested to me because, as I said, I find the assessment a bit difficult. I think I would also prefer to first try out the ones that are recommended to me, but then I would still have the feeling that I would have to take a look at the others as well to see whether I might be able to manage them as well or whether I might understand them nevertheless."

Others stated that they would not distrust the recommendation in principle, but would verify whether the tutorial matches their skills and if they also had all the materials and equipment available.

> P6: "In principle, I do not think anything will come out of it that does not fit at all. But such an intermediate step, as a human being, to read over again, to see if there might be something else? Or simply a technique that I cannot do? It can happen suddenly in between, a specific sewing machine is required that I would not have needed for my skirt, but now I need it. Just like that: 'Do I have materials, techniques, and so on available?' "

Feelings In the interviews, we also noted that participants talked about how they felt interacting with the system.

A participant declared the system as flattering

One participant described it as a sense of accomplishment to check off the things they have already made, and that it boosts their self-confidence in what they can sew.

> P4: "Well, it's also a system that flatters you. Because you can tick off what you have already done, and no matter how much you tick off, you are just always, well, you are always sort of forced to think again about what you have

already done. And even if it is only two things, it is also quasi - if you want to strengthen people's sewing self-confidence, it is smart to ask about it because everyone thinks about it themselves. Like: 'Oh yeah, that is right: I sewed that and the bag for my grandma, too. And the dress and something else. (...)' So I think that gives everyone a good feeling."

5.3 Discussion

In this section, we will evaluate the results of the user study described in Section 5.2. First, we will discuss which prototype is strongly preferred or rejected by the participants. Therefore, we consider the statistical results of the *Friedman test* applied in Section 5.2.1. In the following, we investigate the results of the user interviews and the survey presented in Section 5.2.2 to evaluate to what extent users would integrate the systems into their sewing hobby.

Discussion: Differences in User Acceptance

There are no statistically significant differences in the ratings for different criteria

In the results of our statistical tests, we found evidence for differences in the central tendency with the *Friedman test* for two of the eleven evaluation criteria. For these two evaluation criteria, *Easy to Use* [C7] and *Annoying* [C11], we performed a post hoc test to analyze which of the prototypes differed in the central tendency. Using the *Dunn-Bonferroni-Test* for pairwise comparisons, we could not verify a statistically significant difference between the prototypes for these two evaluation criteria. The fact that the *Friedman test* indicated a difference, which was not detectable in the post hoc test, makes these results somewhat difficult to interpret. Considering the central tendency for both evaluation criteria, we can see that *Sewing on Paper* (PT1) has a median of 1.0, where PT2 $(Mdn = 2.0)$ and PT3 $(Mdn = 2.0)$ score better. We think this is because the user flow in PT1 requires a context switch where one has to move from the active device to a sewing machine and then to a mobile de-

vice, whereas the other two prototypes do not require a context switch. In the results of the survey investigating whether participants think it is annoying to use the prototype [C11] before they can access the tutorials, *Projects Done* (PT3) has a median of -2.0, followed by *Carla's Sewing Studio* (PT2) with $Mdn = -1.0$ and *Sewing on Paper* (PT1) with $Mdn = -0.5$. In contrast to the other evaluation criteria, the lowest value, in this case, is the one with the lowest median, which means that participants evaluate *Projects Done* (PT3) better than *Sewing on Paper* (PT1) for the criterion *Annoying* [C11]. This evaluation result may be partially due to the required effort, as participants criticized that PT1 required much effort, whereas, for PT3, participants often mentioned that it required little effort.

Since we only find evidence of a difference in the central tendency for two of the eleven evaluation criteria, and these are not proven to be significant in further tests, we conclude that the participants evaluate the prototypes similarly, and that they are accepted equally by the participants. However, it is possible that significant differences will be found for sample sizes greater than N=12. We will discuss the extent of this user acceptance in the next section.

Discussion: User Acceptance

In general, we could detect a tendency indicating that the users accept all prototypes PT1–PT3. In the survey results, we count how many evaluation criteria the participants evaluated positively with a threshold of ∼66% votes to determine what the majority of participants indicated. In this context, positive responses mean for criteria C1–C10 *Strongly Agree* or *Agree* and for C11 *Strongly Disagree* or *Agree*. For PT1 and PT2, the majority of the participants rates nine out of eleven evaluation criteria positively. For PT3, it is ten out of eleven positively rated criteria. These results reveal that the users accept all systems but also that the systems have some shortcomings for specific evaluation criteria.

Sewing on Paper (PT1) For *Sewing on Paper*, the criteria *System can assess practical skills* [C8] and *Annoying to use* [C11] do not reach the threshold of ∼66% positive answers in the survey. Participants discussed these criteria topics in the interviews. In assessing practical skills, participants criticized the relevance of straight seams, and we can conclude that sewing skills must include much more than the correctness of a sewed seam on paper. The system lacks in detecting how well a user handles connecting fabrics, which is one of the main tasks in sewing.

Sewing is more than making straight seams

Using the system involves much effort

It is not surprising that this system is more annoying to use compared to the others because it involves the user sewing something followed by an upload of the result. In addition, participants told us that they needed many devices like a sewing machine and a printer to do it. We conclude that these device requirements are a drawback of this system. However, we can assume that people who use sewing tutorials have access to a sewing machine and a printer: printing sewing patterns is usually required for the sewing hobby in general as sewing can not be practiced properly without sewing patterns.

Supports user instead of self-assessing own skills

System useful for beginners

In the interviews, participants acknowledged that the prototype is objective and no self-assessment is involved. This learning underlines that humans have difficulties in self-assessing their skills, and thus, this challenges users to find a matching tutorial. In our research, we did not plan to find out which prototype fits which target group, but the participants gave insights about for whom they think it is suitable. We can assume that *Sewing on Paper* can help beginners filter tutorials by evaluating manual dexterity with the sewing machine for doing their first steps in the new sewing hobby. More advanced users might not be challenged by doing straight seams, so we can assume that this filtering mechanism would not add much value for them.

Carla's Sewing Studio (PT2) The system *Carla's Sewing Studio* (PT2) shows deficiencies for the evaluation criteria *Practical Skills* [C8] and *Preferring the recommendations* [C10].

Sewing is more than theoretical knowledge

Participants argued that the system does not consider prac-

tical skills and that sewing is more than just theoretical knowledge. We can conclude that theoretical knowledge does not express practical skills.

Used terminology can help to filter matching tutorials

However, the prototype concept is still helpful in filtering tutorials because the terminology used in the tutorials is essential for understanding and reworking the tutorial.For preferring the recommendations, we learned that participants would at least try it out and evaluate if the recommendation matched their skills. So we can assume they would use the system if the recommendations matched their skills.

Gamfication motivated to answer all questions

In the interviews, we found that participants liked the Gamification approach and thought it was fun to interact with the system. They reported that the gamification logic would help them to stay motivated to answer all questions, with which we reached our goal to keep the user motivated.

Detect differences between beginners and advanced

Some participants suggested that this prototype might help find the difference between beginners and advanced, but only related to the theoretical knowledge. As total beginners might not have much experience, we assume they might be overwhelmed or demotivated by these questions. We assume that the prototypes help people who know how to handle the sewing machine and have already had some experience with processing fabrics. For very experienced people, it might not help filter the tutorials because if they know everything, the system would recommend all tutorials.

Projects Done (PT3) For *Projects Done* (PT3), trusting the recommendations [C10] did not reach the threshold of ∼66% (strongly) agreement.

Mental Model and Conceptual Model differed

Some people indicated they would at least try the recommendations in a real system. If the tutorial matches the user's skills, we can assume that the user would use the assessment to filter tutorials. We did not find direct evidence why the participants would be skeptical about the recommended tutorials. However, we also noticed that this system's mental and conceptual models differ. We

planned that the system would include "transitive" experiences. Transitive means, for example, if someone sewed a jacket on the highest complexity level, we assume they can also sew a blouse on a medium level, even if they have never sewn a blouse before. Participants thought they only got recommended similar things they already have done before. Since hobby sewers are also likely to want to sew different pieces, we can imagine that assuming that only similar projects are displayed, participants consider this filtering useless and thus would not prefer the recommended tutorials.

Low hurdle to use the system

In the survey and the interviews, we can observe that people highlighted the low effort of using this prototype. We think this influences whether users would use the system because a short-winded assessment that does not need additional devices can be done better in between.

Deficiency in filtering for proficient users

We realized in the interviews that the system would recommend all tutorials for users who rated every clothing category in the highest complexity level. In this case, the system would not help them to find a matching tutorial, as they would still have the same selection, and we think that very experienced users would not benefit from using the system.

Suggestion to combine the systems

Further Considerations In the interviews, we saw that the respective systems have different advantages and disadvantages and that participants evaluated the systems differently. Some participants argued that we should combine the prototypes to combine their strengths. For example, we could combine *Sewing on Paper* (PT1) and *Carla's Sewing Studio* (PT2) to combine users' manual dexterity and theoretical knowledge to get a more detailed skill profile.

Some participants have indicated what they would still be missing in these systems. With this information, we can conclude some issues that still can be improvement.

Ensure learning progress

First, the system needs to ensure that users can have learning progress and do not stay at a level. To this end, the system needs to update the skill profile of the user from time

to time after the initial assessment. *Sewing on Paper* (PT1) is not capable of this reevaluation because sewing more difficult figures will not disclose more about the user's skills. We think this feature can be embedded only in PT2 and PT3. For example, the character Carla could accompany the user over a long period, and there are new levels to play from time to time. For *Projects Done* (PT3), a user could evaluate after they consumed a tutorial if they were successful with it and the system updates the profile respectively. However, users should be able to switch between a challenging or relaxing mode. The challenging mode is for filtering tutorials that are challenging so that users can learn techniques and skills. The relaxing option is to have tutorials that do not require learning something new.

Assessing proficient users not solved yet

Until now, we did not evaluate a solution on how to filter for matching tutorials for proficient users. A proficient participant explained that the required skills are nothing they filter the tutorials for but rather about how inspirational or innovative the patterns are. We conclude that another solution to filter tutorials should be found for proficient users and leave this open for further research.

Limitations Participants did not experience the assessment of the real system. Interacting with the prototypes is rather like having a walk-through of the concepts. Participants could not experience if the recommended tutorials match their skill set and if they could rework the tutorial. Further participants did not sew the exercise for PT1 and did not go through all questions in PT2, and did not enter all complexity levels in PT3. We expect that this influences the results because interacting with the real system might change the opinions about the evaluation criteria.

Chapter 6

Summary and future work

6.1 Summary

In summary, this thesis explored automated user skill evaluation methods to provide matching DIY tutorials.

We identified factors which professionals use to assess student's skills

To create these methods, we first identified factors that professionals and instructors use to assess the quality of sewing work or the handcrafting skills of a sewing person [RQ1]. In interviews with instructors and professionals, we learned that they evaluate the quality of the sewn product and the students' handcrafting skills. For both factors, we could identify subfactors for a closer examination of the quality or skills. We learned that for assessing a student's skill, instructors examine how they work with a sewing machine and how independent they work. Further, instructors consider the student's knowledge, manual dexterity, logical skills, and some soft factors, like talent. If the instructors assess a working sample, they assess the processing, the seams, used materials, the functionality, and the overall impression.

Development of prototypes to demonstrate how the system assess user's skills based on the identified factors

Based on these results, we developed three concepts we implemented in prototypes to evaluate if users would accept these concepts [RQ2]. We decided on the approach *Sewing on Paper* (PT1), which assesses the seam quality and seam allowances, the approach *Carla's Sewing Studio* (PT2) to assess the user's knowledge, and *Projects Done* (PT3) to evaluate the processing of done projects. Evaluating the user study in which we tested the user acceptance of these three prototypes, a statistical investigation showed that the prototypes are equally accepted. We can conclude that users would accept all prototypes from the survey and the interviews we did in the user study. However, we discovered that the prototypes are not for all hobby sewers suitable, as we realized that the system does not add value to filter tutorials for very experienced sewers.

Limitations of this thesis

A limitation in this study is that we only tested the user acceptance with prototypes that are not delivering an assessment of the user's skills, as this is only exemplary. This exemplary system might have influenced the participant's opinions. For the statistical analysis of the survey, the sample size was too small to get significant results. Besides that, we assumed that the systems were technically feasible but did not further investigate.

6.2 Future work

Technical Feasibility

Future research should consider the feasibility of implementing such systems, as we just expected that the proposed prototype should be realizable based on related research.

Assessment for proficients

Besides that, future work could investigate how to identify quality and skill differences of highly experienced people.

Validity for other handcrafting hobbies

One topic that we did not address in this work is what implications we can draw from the domain of sewing to general handcrafts, so we leave that open for future research as well.

Combinations

Participants indicated they would prefer a combination of the prototypes in the user study. Future work could

investigate combining the systems and evaluate if a combined assessment mechanism increases the expressiveness of sewing skills.

Other factors

Since we could not consider all identified factors in this work in the concepts, it could also be promising to conceptualize them.

Appendix A

Prototypes of the Concepts

A.1 General

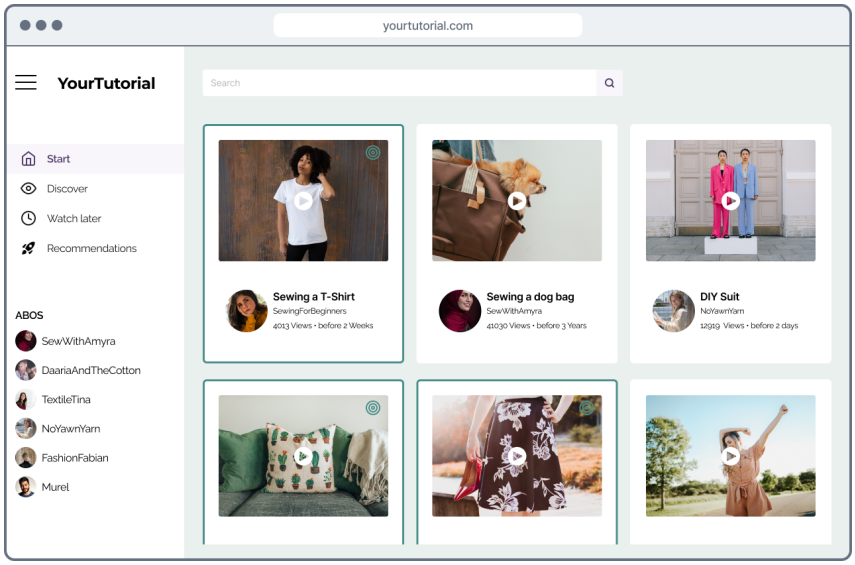

Figure A.1: Recommendations are highlighted with green frame and circle

A.2 Sewing on Paper (PT1)

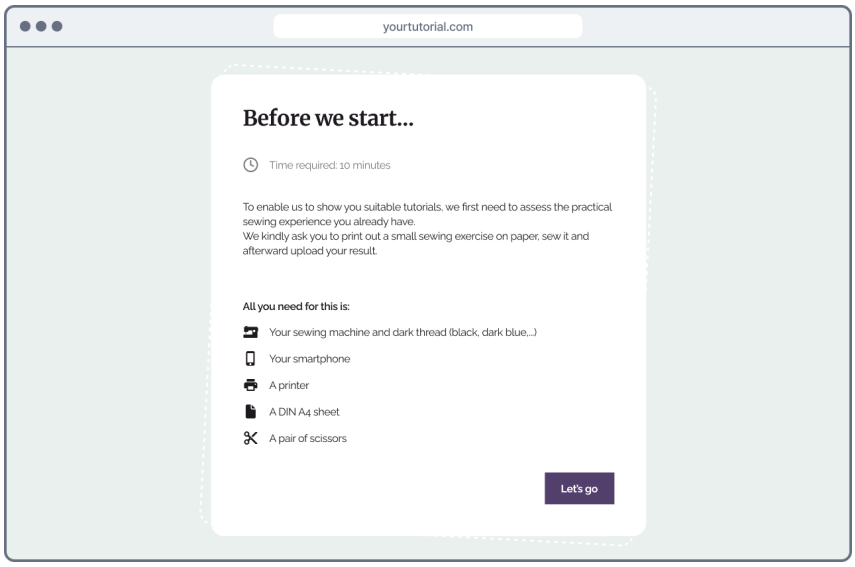

Figure A.2: Information about the process.

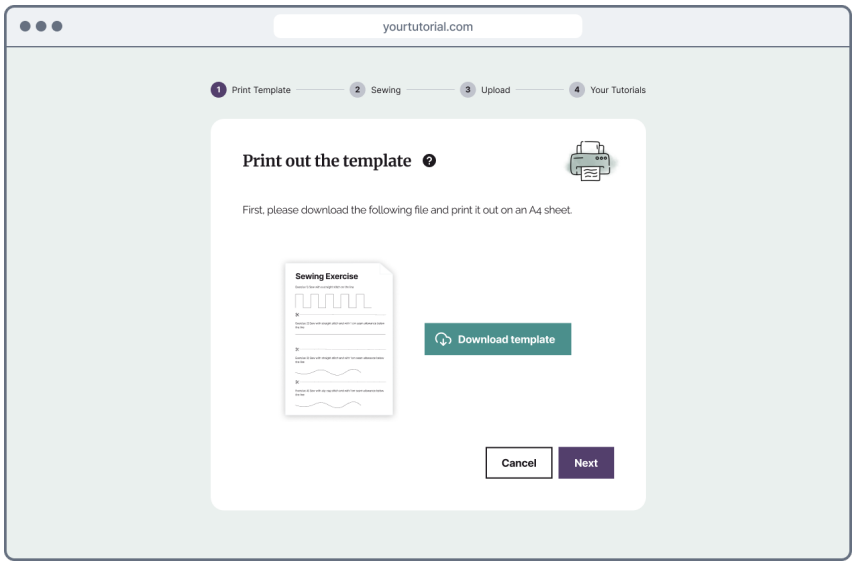

Figure A.3: System prompts user to download sewing exercise.

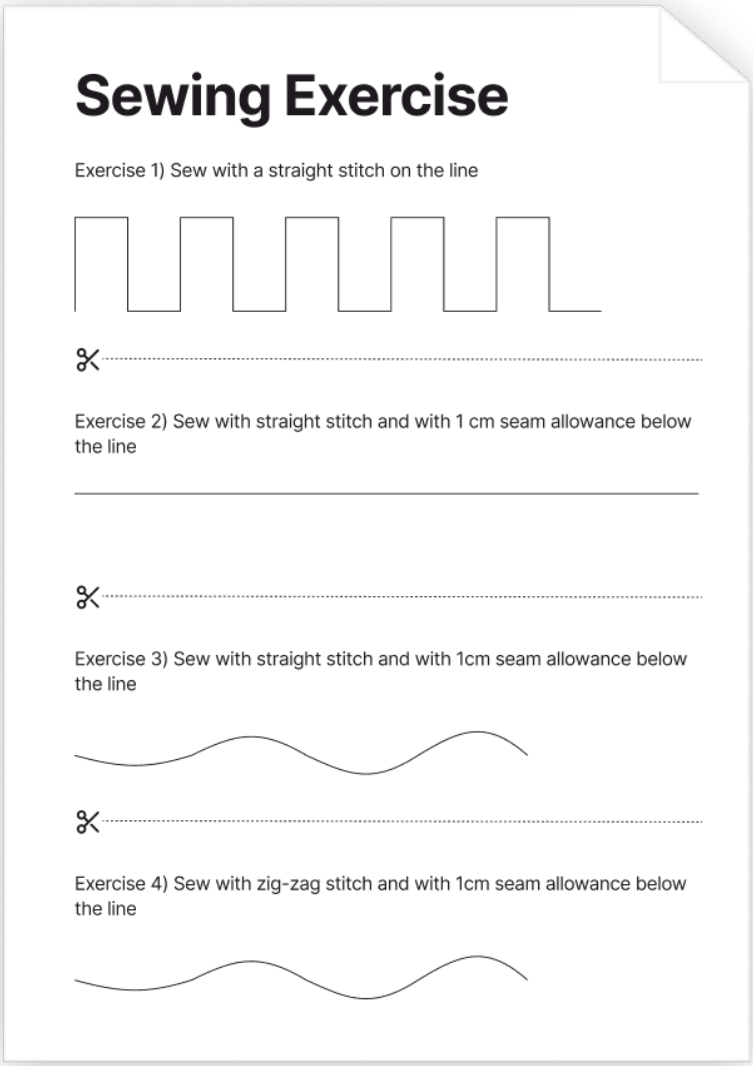

Figure A.4: Exemplary sewing exercise.

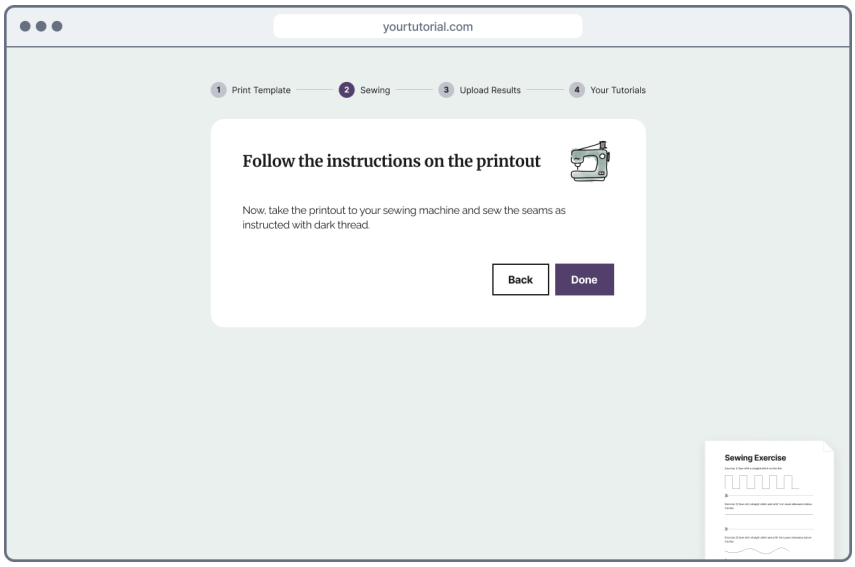

Figure A.5: System instructs user to do the sewing exercise.

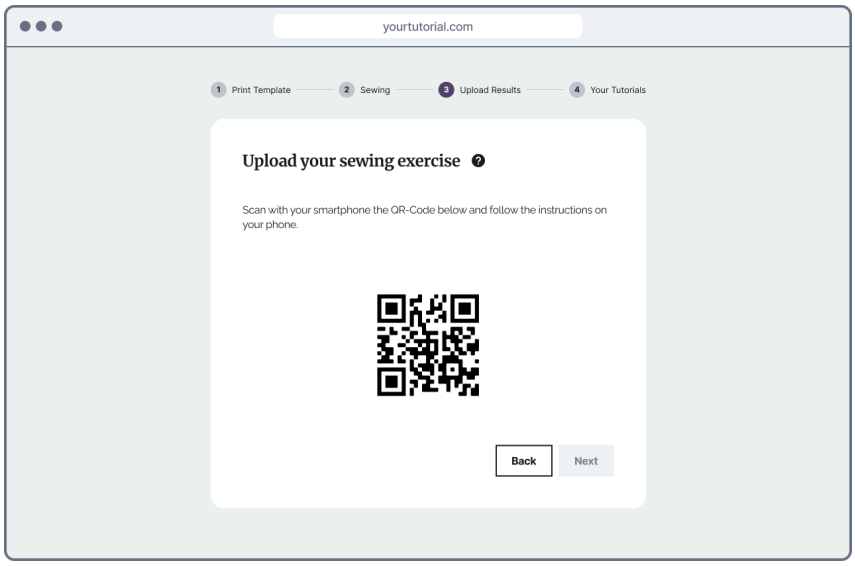

Figure A.6: User can scan QR-Code with smartphone.

Figure A.7: System explains the process of uploading.

Figure A.8: User scans the sewing example.

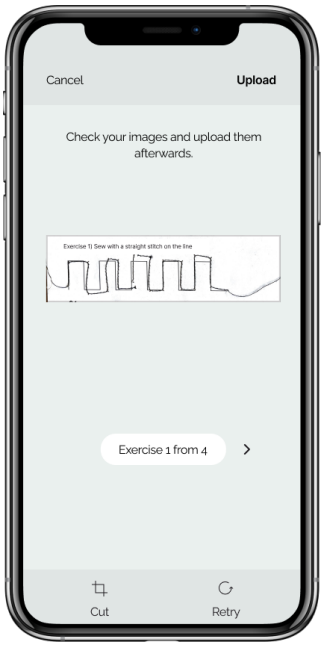

Figure A.9: System asks user to confirm upload.

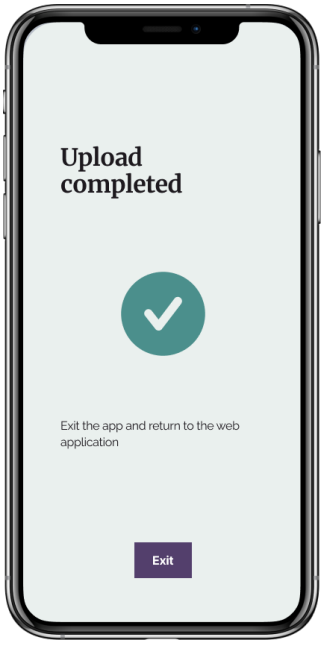

Figure A.10: Upload is completed.

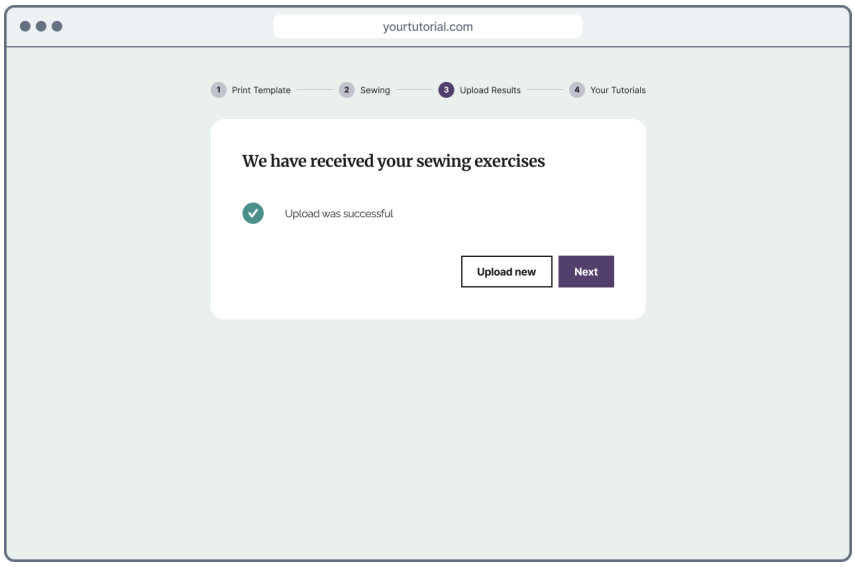

Figure A.11: System received the upload.

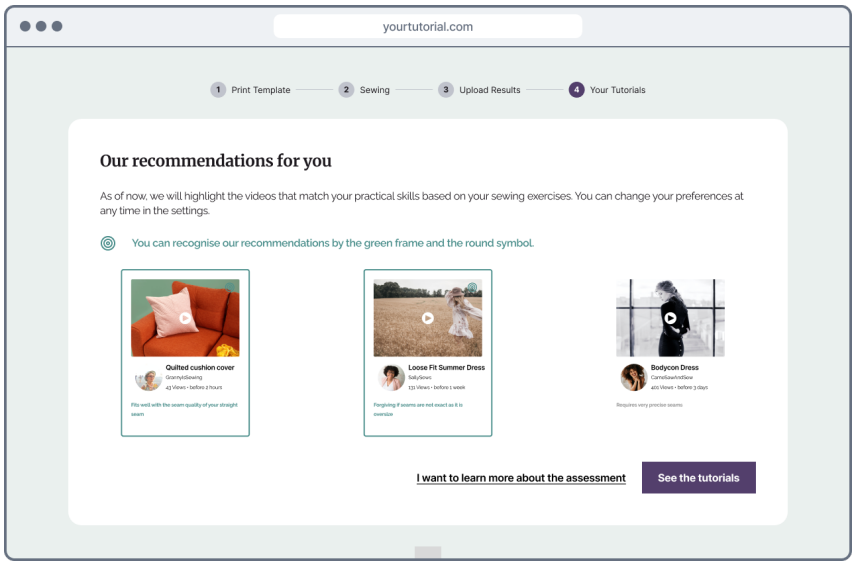

Figure A.12: System explains how recommendations will look like.

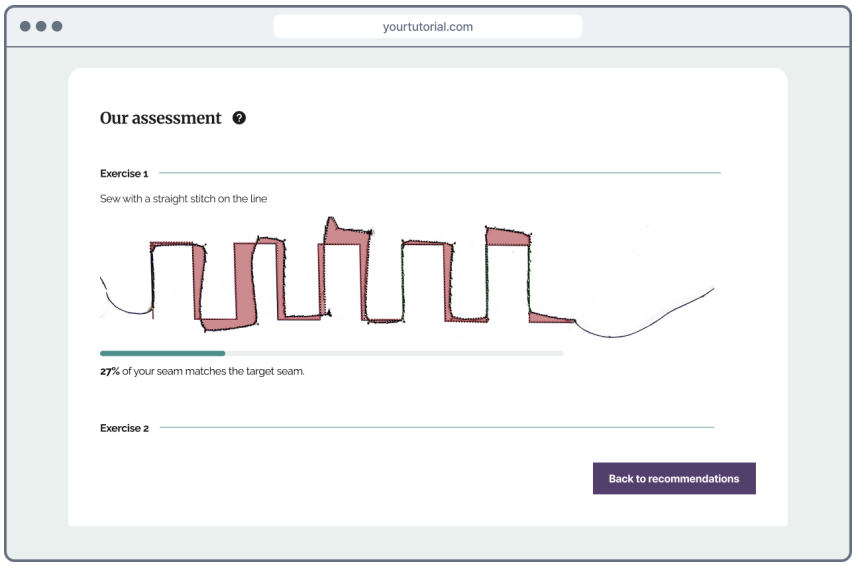

Figure A.13: System shows the results of the sewing exercise.

A.3 Carla's Sewing Studio (PT2)

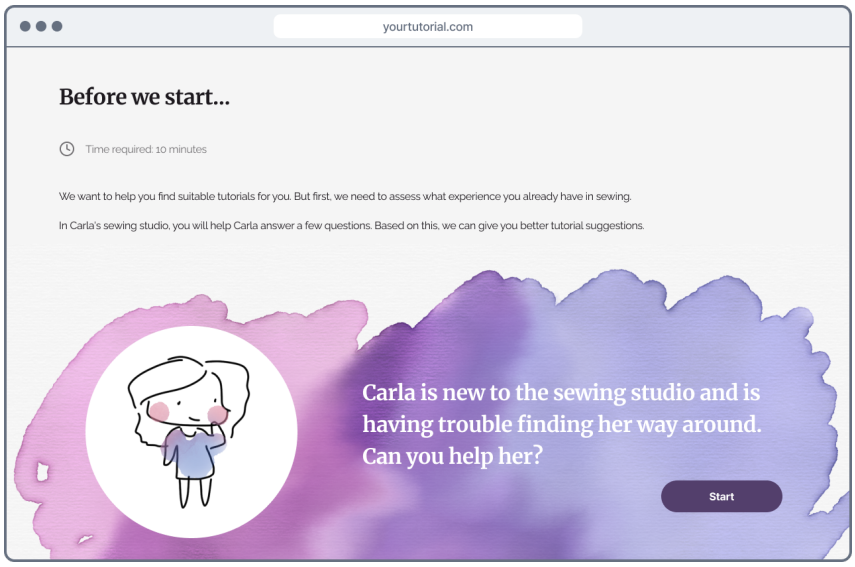

Figure A.14: Information about the process.

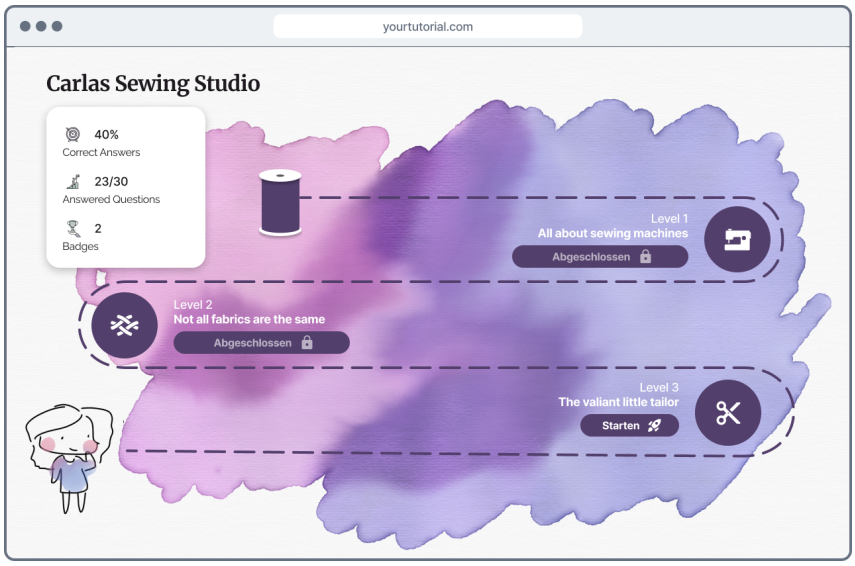

Figure A.15: System displays level overview.

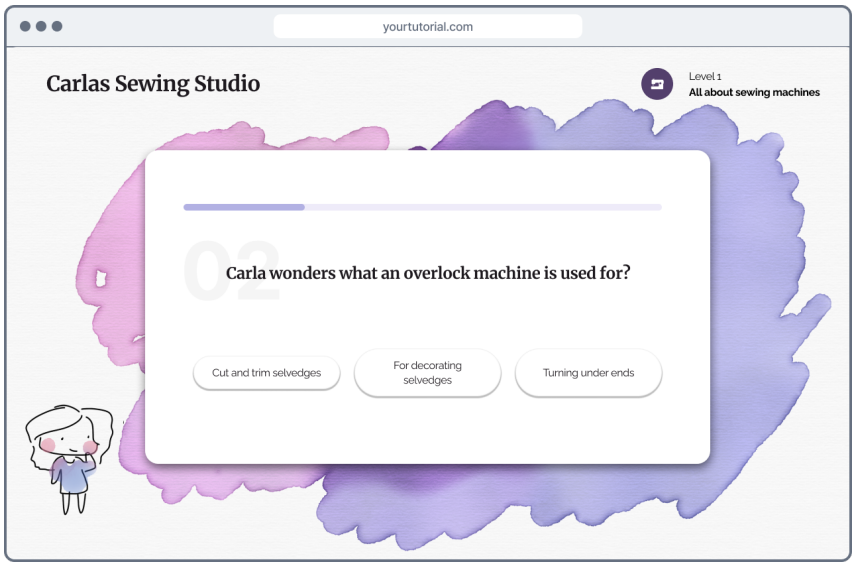

Figure A.16: System asks user a question.

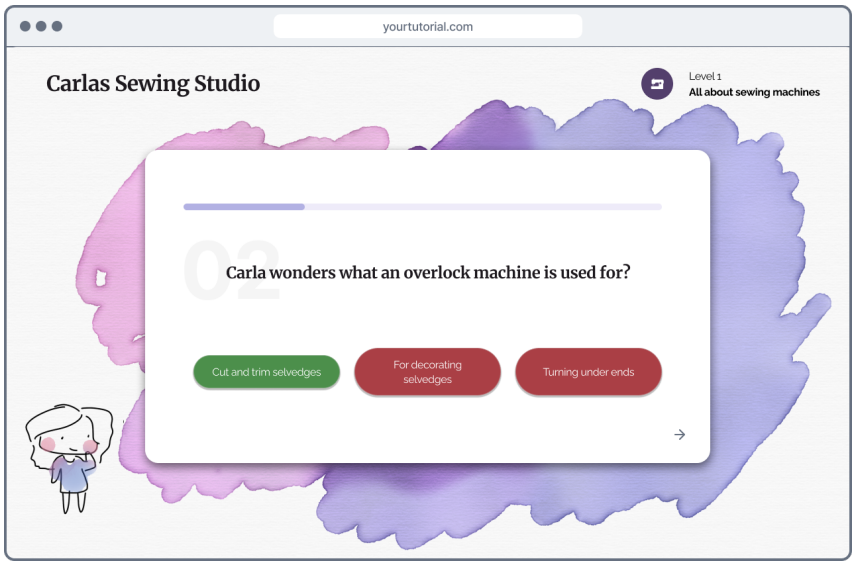

Figure A.17: System evaluates the user's answer.

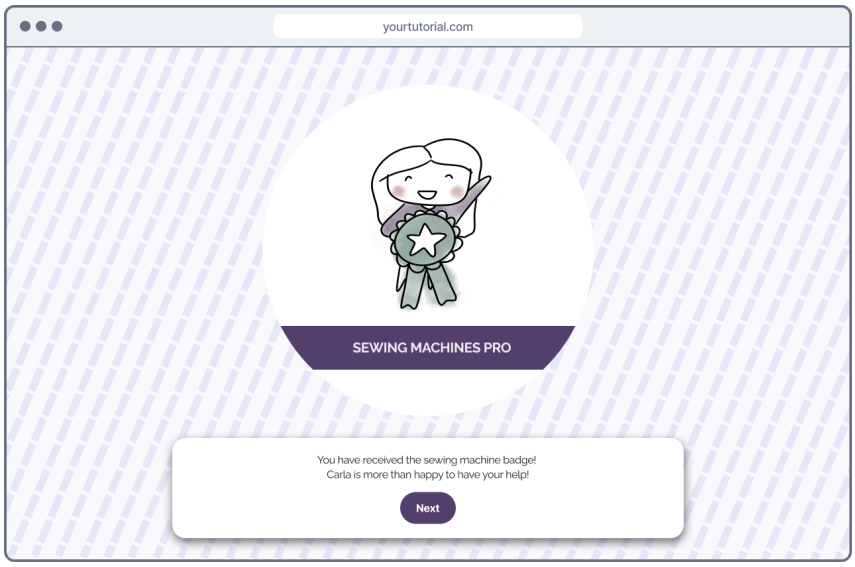

Figure A.18: User receives a badge.

A.4 Projects Done (PT3)

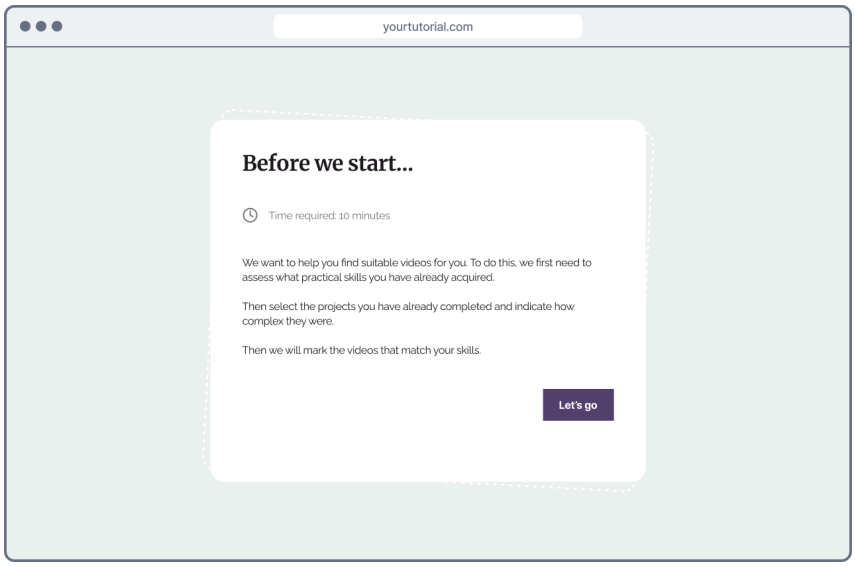

Figure A.19: Information about the process.

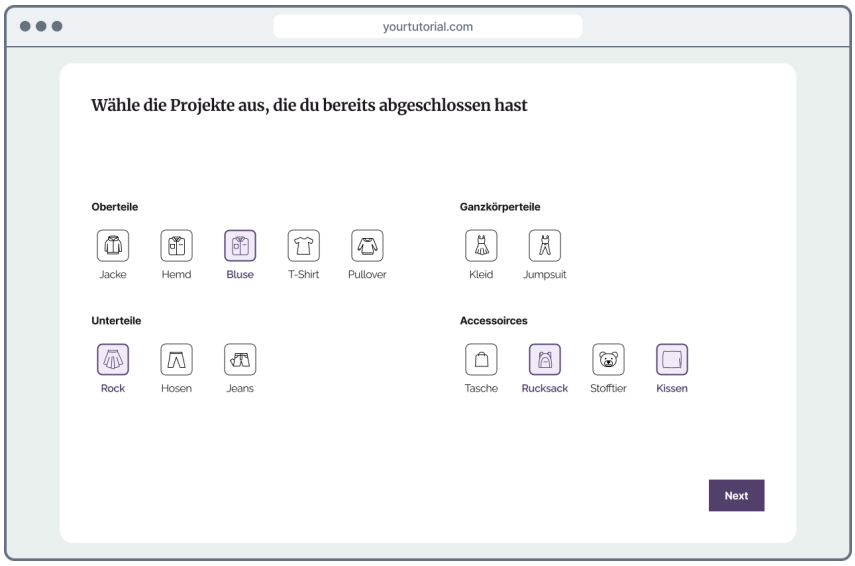

Figure A.20: System prompts user to enter categories they already did.

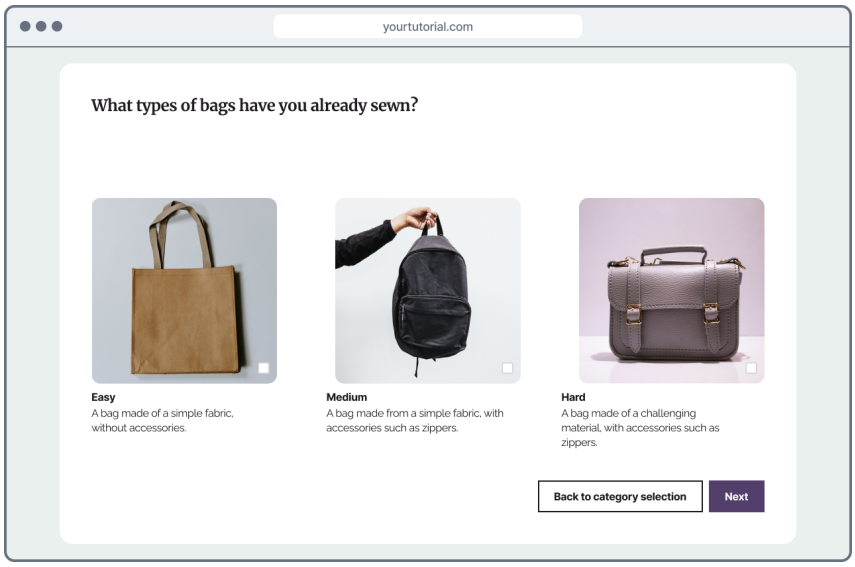

Figure A.21: System asks for complexity level of projects in a category

Appendix B

Study Designs

B.1 Expert Interview: Identify Factors to Assess Quality of Sewing Work

B.1.1 Subject of the study

To explore possibilities to enable a system to assess a user's sewing skills, we first want to find out by which factors professionals and instructors in the discipline of sewing technology assess the sewing work or the skills of their students.

B.1.2 Target Group

- Number of participants: around 10
- Participants are professionals or instructors in the discipline of sewing technology in vocational schools, universities of applied sciences, universities, and sewing courses.

B.1.3 The goal of the Study

The following study will examine factors to assess sewing skills:

RQ1 What are factors that professionals and instructors use to assess the quality of sewing work or the handcrafting skills of a sewing person?

B.1.4 Data Collection Method

We will collect the data qualitatively. For this purpose, we will conduct a semi-structured interview.

B.1.5 Setup of the Study

Approx. 1 week before the Study:

- E-Mail consent form to participants

- Create and send a Zoom link, one Zoom session per participant
- *Duration:* approximately 30 minutes

Just before the session

- Have the guideline with questions for the interview ready
- Have the pictures of the sewed piece for evaluation ready

Conducting the Study

- The director of studies welcomes the participant
  - *General aim:* find out how they assess the sewing skills of their students
  - Go through the consent form together, and explain the different points, especially about the use of personal data:
    1. Recording the voice
- Semi-Structured Interview Questions:
  1. As an instructor/teacher/professor, how is your occupational routine with the students?
     - If it is about teaching sewing:
       (a) Do you divide your students into different levels?
       (b) What do you use to determine whether someone has experience or not? At which things do you look?
       (c) How do you observe a learning process?
  2. Do you examine students?
     (a) How is this typically done?
     (b) What forms of testing do you use?
  3. In the pandemic, we all switched to distance learning. What problems did you encounter?
     (a) How did you solve the problems?
  4. Please imagine that I'm one of your students, and I would show you this sewed piece and ask for feedback:
     (a) Please describe what exactly you look at to evaluate whether I did this well or poorly?

5. Would you assign a certain level of experience to certain materials/projects?
   (a) What would you consider beginner or more advanced typically?
6. How do you think sewing skills are organized?
   (a) Is it hierarchical like in elementary school, where you first learn ABC, then write words and essays? *or*
   (b) Are sewing skills more organized like a diverse toolbox, with skills that one can learn independently of each other?

B.2 User Study: Testing User Acceptance

B.2.1 Subject of the study

To identify factors that can be used to evaluate the practical skills of users, the first step was a qualitative study with experts and teachers in the discipline of sewing technology. As a result of this study, we were able to identify, among others, three factors for the evaluation of sewing techniques, which we theoretically expect to be testable in an automated way:

(A1) Sew predefined lines (on paper), comparison between target and actual seam.

(A2) Conclusions on practical knowledge by testing theoretical questions

(A3) Conclusions on practical knowledge through already gained experience with projects/used materials

Based on these three factors, we developed the following prototypes:

(PT1) Users print out a sewing specification on paper and sew along the drawn lines. Photos of the different sewing samples are then uploaded, and the system calculates the deviation between the target and actual seams. .

(PT2) The users are guided through a quiz game to answer theoretical questions about sewing technology.

(PT3) Users select which sewing projects they have already completed and indicate whether these projects belong to the categories ”easy”, ”medium”, and ”difficult”.

B.2.2 Target Group

- Number of Participants: at least 10
- Participants already have sewing experience or can imagine learning to sew.

B.2.3 Goal of the Study

The following study will examine the acceptance of the prototypes by users.

RQ2 To what extent are our proposed concepts, to assess the user's skills, accepted by users?

B.2.4 Data Collection Method

The data will be collected qualitatively and quantitatively. For this purpose, the following will be performed for each prototype PT1-PT3:

1. Remote usability test with evaluation strategy: mixture of elements of "Observation" and "Think Aloud" (M1) structured
2. Mixture of questionnaire survey (M2) with a Likert Scale and a semi-standardized interview (M3)

The order of the prototypes is permuted for each study session so that the learning effect, which occurs from the second prototype onward, is distributed equally across all evaluations of the prototypes.

B.2.5 Preparation of the Study

Approx. 1 Week before the Study

- Email consent form to participants.
- Create and send a Zoom link; one Zoom session per participant.

- Advise participants that remote sessions must be conducted using Zoom Desktop Client for Windows, Mac or Linux, or iPad for participants to take control (to interact with the prototype and complete the questionnaire)
- Duration: approximately 60 minutes

Just before the Session

- Open Figma file with prototypes
- Open the website with the survey
- Open consent form (PDF)

B.2.6 Conducting the Study

1. The director of studies welcomes the participant
   - General aim: to help users find *matching tutorials* more easily.
   - *Matching tutorial:* a tutorial that you can do with your current level of knowledge without further help.
   - We developed three different concepts for participants to test out
   - Go through the consent form together, and explain the different points, especially about the use of personal data:
     (a) Recording the voice
     (b) Recording the webcam video
2. The director of studies explains what will happen in the following:
   (a) Participants will see different prototypes, they can take over the control and try out the prototype
   (b) The Participants are encouraged to say what they think about the system
   (c) Meanwhile, the study leader will observe how the person interacts with the prototype.
   (d) After each prototype, the participant is asked to fill out a questionnaire and answer some questions
   (e) After the evaluation of all prototypes some general data will be taken (e.g. sewing experience, age, etc.)
3. For all prototypes (PT1–PT3) the study leader will conduct following:

(M1) The software prototype presented via remote session in Zoom. The participants grant control over the prototype, so they can interact with the system, while the study leader observes the participant. The study leader takes notes on the person's interaction with the prototype. The participant is encouraged to directly express thoughts about the prototypes shown. *Exception for (PT1):* From the download of the print template, the user flow foresees that a sewing exercise is carried out. It is not necessary for participants to print the paper and do the sewing exercise. At this point, participants can have a look at the paper and continue without uploading the exercise. Also the QR code does not need to be scanned, by clicking on the QR code participants will be redirected to the upload process. This should be announced.

(M2) Questionnaire, see below

(M3) Structure of semi-standardized interview:

(I1) What are your first thoughts about the system? (Why?)

(I2) What did you like about the concept? (Why?)

(I3) Where did something not worked as you excepted? (Why?)

(I4) For (PT1-PT3):

(P1) If the system rates how well you have sewn on paper, to what extent do you think the tutorials that are displayed based on the result match your skills?

(P2) To what extent do you think the tutorials displayed based on a test of your theoretical knowledge of sewing techniques match your practical skills?

(P3) To what extent do you think the tutorials displayed, based on your previous completed projects and their difficulty levels, match your practical skills?

(I5) Address observations (see notes), and clarify *"Why?"* if necessary. (The goal is to understand if the user had problems with something or is enthusiastic (not enthusiastic) about something.)

(I6) If relevant, refer to questions in the questionnaire, address particularly negative/positive points, clarify *"Why?"*

4. Gathering general and demographic data (see below).

Evaluation Criteria	Statement
C1 Faster	I expect to find matching tutorials faster with the use of the system
C2 Easier	I expect to find matching tutorials easier with the use of the system
C3 Usefulness	I think the system would be useful in my sewing hobby
C4 Learning Easy	Learning to use the system would be easy for me
C5 Behaved as expected	The system behaved as I expected
C6 Interaction clear and understandable	The interaction with the system is clear and understandable of for me
C7 Easy to use	I think the system is easy to use
C8 Assessing practical skills	I believe the system is good at assessing my practical skills
C9 Trust	I would trust the system's assessment
C10 Prefer to watch recommended tutorials	I would prefer to watch the highlights tutorials
C11 Annoying to use system	I think it's annoying to use the system before I get access to the tutorials

Table B.1: Participants evaluate this statements on a 5-point Likert scale (strongly agree - strongly disagree

B.2.7 Questionnaire

Evaluation of The Prototypes

The assessment in the Table B.2.7 is done for every shown prototype PT1–PT3. The answer options for each statement are:

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neither agree, nor disagree
- ◯ Disagree
- ◯ Strongly Disagree

Experience with Sewing Techniques & Sewing Tutorials

How often do you use sewing tutorials?

◯ Never
◯ Rarely
◯ Sometimes
◯ Often
◯ Always

How do you rate your own sewing skills?

◯ Novice
◯ Advanced Beginner
◯ Competence
◯ Proficient
◯ Expert

How often does it happen that you watch a tutorial and then realize that you lack practical skills to perform the content shown?

◯ Never
◯ Rarely
◯ Sometimes
◯ Often
◯ Always

How long have you been sewing?

◯ Not started, but plan to start

◯ Not started, and don't plan to start

◯ Less than one year

◯ 1-2 years

◯ More than 3 years

Demographic Data

How old are you?

◯ under 18

◯ 18–24

◯ 25–35

◯ 35–50

◯ over 50

Your Gender

◯ Female

◯ Male

◯ Non-Binary

Appendix C

User Study: Statistics

		Normality Test			
		Shapiro-Wilk			Normal
		Statistic	df	Significance	Distributed
	Sewing on Paper (PT1)				
[C1]	Faster	.873	12	.072	Yes
[C2]	Easier	.799	12	.009	No
[C3]	Usefulness	.831	12	.022	No
[C4]	Learning Easy	.608	12	.000	No
[C5]	Behaved as expected	.777	12	.005	No
[C6]	Interaction clear and understandable	.552	12	.000	No
[C7]	Easy to use	.846	12	.033	No
[C8]	Assessing practical skills	.850	12	.037	No
[C9]	Trust	.859	12	.048	No
[C10]	Prefer to watch recommended tutorials	.824	12	.018	No
[C11]	Annoying to use system	.890	12	.118	Yes
	Carla's Sewing Studio (PT2)				
[C1]	Faster	.798	12	.009	No
[C2]	Easier	.824	12	.018	No
[C3]	Usefulness	.794	12	.008	No
[C4]	Learning Easy	.479	12	.000	No
[C5]	Behaved as expected	.668	12	.000	No
[C6]	Interaction clear and understandable	.608	12	.000	No
[C7]	Easy to use	.465	12	.000	No
[C8]	Assessing practical skills	.910	12	.212	Yes
[C9]	Trust	.884	12	.099	Yes
[C10]	Prefer to watch recommended tutorials	.861	12	.051	Yes
[C11]	Annoying to use system	.836	12	.025	No
	Projects Done (PT3)				
[C1]	Faster	.799	12	.009	No
[C2]	Easier	.658	12	.000	No
[C3]	Usefulness	.779	12	.005	No
[C4]	Learning Easy	.427	12	.000	No
[C5]	Behaved as expected	.680	12	.001	No
[C6]	Interaction clear and understandable	.465	12	.000	No
[C7]	Easy to use	.327	12	.000	No
[C8]	Assessing practical skills	.843	12	.030	No
[C9]	Trust	.865	12	.056	Yes
[C10]	Prefer to watch recommended tutorials	.864	12	.055	Yes
[C11]	Annoying to use system	.668	12	.000	No

Table C.1: Normality Test with Shapiro Wilk at significane level .05.

		Related-Samples Friedman's Two-Way Analysis of Variance by Ranks					
		Hypothesis Test Summary[a]		Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary			
		Sig.[b, c]	Decision	Total N	Test Statistic	Degree of Freedom	Asymptotic Sig. (2-sided test)
[C1]	Faster	0.431	Retain H_0	12	1.684	2	0.431
[C2]	Easier	0.786	Retain H_0	12	0.483	2	0.786
[C3]	Usefulness	0.879	Retain H_0	12	0.258	2	0.879
[C4]	Learning Easy	0.584	Retain H_0	12	1.077	2	0.584
[C5]	Behaved as expected	0.727	Retain H_0	12	0.636	2	0.727
[C6]	Interaction clear and understandable	0.549	Retain H_0	12	1.200	2	0.549
[C7]	Easy to use	0.005	Reject H_0	12	10.750	2	0.005
[C8]	Assessing practical skills	0.128	Retain H_0	12	4.108	2	0.128
[C9]	Trust	0.697	Retain H_0	12	0.722	2	0.697
[C10]	Prefer to watch recommended tutorials	0.066	Retain H_0	12	5.429	2	0.066
[C11]	Annoying to use system	0.032	Reject H_0	12	6.889	2	0.032

[a] The distributions of PT1, PT2, and PT3 are the same.
[b] The significance level is 0.050.
[c] Asymptotic significance is displayed.

Table C.2: Results of *Friedman test*.

Bibliography

Heidi Andrade and Anna Valtcheva. Promoting learning and achievement through self-assessment. *Theory into practice*, 48(1):12–19, 2009.

Bundesinstitut Berufsbildung. Ausbildung gestalten maßschneider/maßschneiderin, 2008. URL `https://www.bibb.de/dienst/veroeffentlichungen/de/publication/show/8986`. Accessed on 27.07.2022.

Michael J Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A Roch, Sharon Gannot, and Charles-Alban Deledalle. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628, 2019.

Janet A Blood and Terri D Owens. Authentic assessment for basic clothing construction: The practical final exam. In *International Textile and Apparel Association Annual Conference Proceedings*, volume 72. Iowa State University Digital Press, 2015.

Robin E Clark. Performance assessment in the arts. *Kappa Delta Pi Record*, 39(1):29–32, 2002.

Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 222(2015):248, 2015.

Tom Cole and Marco Gillies. More than a bit of coding:(un-) grounded (non-) theory in hci. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–11, 2022.

Fred D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319, September 1989. ISSN 02767783. doi: 10.2307/249008. URL https://www.jstor.org/stable/249008?origin=crossref.

Edward L Deci and Richard M Ryan. Self-determination theory. 2012.

Audrey Desjardins and Ron Wakkary. Manifestations of everyday design: Guiding goals and motivations. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, C&C '13, page 253–262, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450321501. doi: 10.1145/2466627.2466643. URL https://doi.org/10.1145/2466627.2466643.

K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data*. The MIT Press, April 1993. ISBN 978-0-262-27239-1. doi: 10.7551/mitpress/5657.001.0001. URL https://doi.org/10.7551/mitpress/5657.001.0001.

Isabel Funke, Sören Torge Mees, Jürgen Weitz, and Stefanie Speidel. Video-based surgical skill assessment using 3d convolutional neural networks. *International journal of computer assisted radiology and surgery*, 14(7):1217–1225, 2019a.

Isabel Funke, Sören Torge Mees, Jürgen Weitz, and Stefanie Speidel. Video-based surgical skill assessment using 3D convolutional neural networks. *Int J CARS*, 14 (7):1217–1225, July 2019b. ISSN 1861-6410, 1861-6429. doi: 10.1007/s11548-019-01995-1. URL http://link.springer.com/10.1007/s11548-019-01995-1.

Jun Gong, Fraser Anderson, George Fitzmaurice, and Tovi Grossman. Instrumenting and analyzing fabrication activities, users, and expertise. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

Kay Hendrickson, Janet Hutchinson Hiller, and Nancy Mordhorst. Measuring up: quality standards for sewn items/projects. 2004.

Amy Hurst, Scott E Hudson, and Jennifer Mankoff. Dynamic detection of novice vs. skilled use without a task model. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–280, 2007.

Hendrickson Kay, Jan Hiller, and Nancy Mordhorst. Sewing skills checklist - washington state university, May 2000. URL `https://s3.wp.wsu.edu/uploads/sites/2086/2014/09/Sewing-Skills-Checklist.pdf`.

Justin Kruger and David Dunning. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121, 1999.

Shing-On Leung. A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales. *Journal of social service research*, 37(4):412–421, 2011.

Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019.

Janet Metcalfe, Arthur P Shimamura, et al. *Metacognition: Knowing about knowing*. MIT press, 1994.

Anna-Zoe Patterns. Levelling up – thoughts on pattern difficulty ratings, 2018. URL `https://www.annazoepatterns.com/levelling-thoughts-pattern-difficulty-ratings/`. Accessed on 26.07.2022.

Grant A Pignatiello, Richard J Martin, and Ronald L Hickman Jr. Decision fatigue: A conceptual analysis. *Journal of health psychology*, 25(1):123–135, 2020.

Nornadiah Mohd Razali and Yap Bee Wah. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33, 2011.

Jongryun Roh, Hyeong-jun Park, Kwang Jin Lee, Joonho Hyeong, Sayup Kim, and Boreom Lee. Sitting posture

monitoring system based on a low-cost load cell using machine learning. *Sensors*, 18(1):208, 2018.

Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69:371–380, 2017. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2016.12.033. URL https://www.sciencedirect.com/science/article/pii/S074756321630855X.

J. Saldana. *The Coding Manual for Qualitative Researchers*. SAGE Publications, 2015. ISBN 9781473902497. URL https://books.google.de/books?id=N8pOrgEACAAJ.

Joachim Schubert. *Fachwörterbuch Textil: deutsch-englisch, english-german*. Edition Textil. Deutscher Fachverl, Frankfurt am Main, 7., völlig überarb. und erw. aufl edition, 2005. ISBN 978-3-87150-894-3.

Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, 1(4):221–228, 1996.

Geoffrey Steinberg and Charlene Riggle. Early Validation of Information System Prototypes Using the Technology Acceptance Model. 1995.

Burda Style. Kreationen nach level, 2022a. URL https://www.burdastyle.de/kreationen-nach-level. Accessed on 26.07.2022.

Burda Style. Wir über uns, 2022b. URL https://www.burdastyle.de/wir-ueber-uns. Accessed on 26.07.2022.

Gail M Sullivan and Anthony R Artino Jr. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542, 2013.

Murat Yilmaz and Şahin Kayali. An exploratory study to assess analytical and logical thinking skills of the

software practitioners using a gamification perspective. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 21(1):178–189, 2016.

Index

Typeset July 27, 2022