Software-supported Participatory Design: Design and Evaluation of the Tool PDot

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

UNIVERSITY OF LEICESTER

by
Matthias Heintz
Department of Informatics
University of Leicester

March, 2017

Abstract

Participatory Design (PD) is a common software development approach that actively includes end-users in the design process. This ensures tailored results and can lead to a strong feeling of ownership and overall empowers end-users. Commonly applied paper-based PD approaches have several shortcomings. A prototype presented on paper is not interactive for the end-user to experience it. Preparing PD ideas captured as physical artefacts (e.g. sketches on acetates) for further data analysis can be unduly time consuming. Using software tools to conduct PD activities instead of relying on paper-based methods can address these shortcomings. The author has been motivated to design, develop, and evaluate two such tools - PDotCapturer and PDotAnalyser. PDotCapturer is used by end-users participating in PD activities to create new designs from scratch or express (re-)design ideas. PDotAnalyser is used by designers to work with and further analyse the ideas captured. PDotCapturer is compared with similar paper-based approaches to evaluate the relative effectiveness of tool-based and paper-based PD activities in terms of quantity and quality of design ideas elicited. To perform this comparison, the coding scheme CAt+ (Categories plus Attributes) to rate the quality of PD ideas is developed. CAt+ can also be used to filter and aggregate PD ideas to support designers in making sense of as well as addressing such ideas for re-design. Results of the comparisons of paper-based and tool-based approaches show that paper is advantageous in some regards (e.g. number of ideas gathered), but the tool is comparable or in some regards outperforms paper (e.g. user preference). Given the additional advantages tool-usage can bring (e.g. automated analysis support), the context where paper-based or tool-based PD approaches suit better is discussed. For future work the use of PDotCapturer and PDotAnalyser in diverse and distributed settings will be explored.

Acknowledgements

First of all I want to thank my supervisor Dr. Effie Law for her helpful support. Her feedback and input in many fruitful meetings helped me to shape my research and improve my scientific as well as writing skills.

I want to thank the Go-Lab project for making this PhD possible through funding my research with a studentship as well as all the project partners for making Go-Lab such an interesting and rewarding project to work for. Thanks to all the participants in the PD activities, not only for shaping the Go-Lab results but also for using and providing direct or indirect feedback on PDotCapturer. I want to especially thank the technical team at EPFL (Ecole polytechnique fédérale de Lausanne) for their feedback regarding my tool ideas, and my colleagues at the University of Leicester team (Effie, Jan, Rob, and Samaneh) for making working for the project a pleasure.

I want to thank the people in the Department of Informatics of the University of Leicester for providing such a nice working environment. Special thanks go to my second supervisor Prof. Thomas Erlebach, for useful discussions of my research, and Gilbert Laycock, for his technical support. Also to all the PhD students with whom I could discuss work-related matters as well as share breaks from our research endeavours.

Finally I want to thank my family and friends, especially my parents, Christina and Norbert, and my sister Stephanie for their support and help throughout the time of my PhD. Their interest in my work led to interesting discussions of my research. Without them writing this thesis would have been a lot harder.

Contents

List of Tables

List of Figures

Abbreviations

3D three-dimensional.

ACM Association for Computing Machinery.

ACTA Anti-Counterfeiting Trade Agreement.

CHI ACM CHI Conference on Human Factors in Computing Systems.

CoPDA International Workshop on Cultures of Participation in the Digital Age: From "Have to" to "Want to" Participate.

DR Designer Requirement.

EDUCON IEEE Global Engineering Education Conference.

GD General Design.

GUI Graphical User Interface.

GWT Google Web Toolkit.

HCI Human-Computer Interaction.

HTML HyperText Markup Language.

IBF Informant, Balanced, Facilitated design.

IEEE Institute of Electrical and Electronics Engineers.

INTERACT International Conference on Human-Computer Interaction.

N/A Not Applicable.

NordiCHI Nordic Conference on Human-Computer Interaction.

NV Not Violated.

PD Participatory Design.

R Requirement.

RAId Rapid Analysis of design Ideas.

RE Requirements Engineering.

RQ Research Question.

SD Standard Deviation.

SMEs Small and Medium-Sized Enterprises.

STCSN Special Technical Community on Social Networking.

SUS System Usability Scale.

UCD User-Centred Design.

UMUX Usability Metric for User Experience.

UR User Requirement.

UX User Experience.

V Violated.

VDTs Visual Display Terminals.

VR Virtual Reality.

Glossary

CAt+ (Categories plus Attributes) is a newly created coding scheme to rate PD ideas regarding webapps to compare the results of different approaches (e.g. paper-based and tool-based) and to support designers in their idea analysis work.

Designers are researchers as well as practitioners having design experience (e.g. designing user interfaces) in their job (i.e. the term is used in its broadest sense).

Distributed Participatory Design is a software development approach that actively includes end-users as design partners which is applied in a setting where end-users and designers are not in the same physical location.

End-User is a person using a product or service, specifically a webapp.

Go-Lab (Global Online Science Labs Inquiry Learning at School) was a European project with the goal to enhance science lessons by integrating online labs.

Layered Elaboration is a PD method where an initial prototype design on paper is overlaid with acetates on which end-users annotate the initial design with comments and re-design suggestions.

Participatory Design a software development approach that actively includes end-users as design partners.

PDot (Participatory Design online tool) is the name of a newly created, dedicated Participatory Design online tool consisting of PDotCapturer and PDotAnalyser.

PDotAnalyser is the PDot GUI aimed at designers: it presents the gathered PD ideas in an aggregated way and offers additional visualisation and filter options.

PDotCapturer is the PDot GUI aimed at end-users: it is focusing on the idea gathering aspect of PD.

Webapps is a collective term used to refer to a variety of web-based applications, websites, and online portals.

Chapter 1

Introduction

This chapter motivates the research described in this thesis by presenting and highlighting gaps in the current Participatory Design (PD) research literature, mainly the missing comparison of paper-based and tool-based PD approaches. This general observation leads to specific research questions and the description of the approach that is applied to address those questions. The Go-Lab project, aiming at improving science lessons in schools by including online laboratories, is presented as the context of this work. The main research contributions and the publications derived are listed and an outline of this thesis is given.

1.1 Motivation

Let 'all voices be heard' (Fischer, 2013) is the tenet of Participatory Design (PD) (Muller, 2007; Schuler and Namioka, 1993). The goal is to gather insights and input, especially design ideas and suggestions, from end-users. This is achieved by including prospective end-users in the design and development process for various physical as well as digital systems, products, and services (Greenbaum and Kyng, 1991; Halskov and Hansen, 2015; Muller, 2007; Schuler and Namioka, 1993).

By actively including end-users it is ensured that the resulting product is not only tailored to their needs but also shaped by their input. This can, for example, result in a strong feeling of ownership (Carroll et al., 2000) and high acceptance of the end-product (Abras et al., 2004; Read et al., 2016; Scariot et al., 2012). It empowers end-users in the (software) design process as they can actively shape the result (Clement and Van den Besselaar, 1993).

Supporting end-users to express their design ideas is certainly very important for eliciting and capturing their quality input. But it is also relevant to support people responsible for the subsequent idea analysis process to maximise the value of PD activities. In software development projects the people responsible for working with end-user ideas can be graphic designers, software developers, or human-computer interaction (HCI)/PD specialists. Nonetheless, subsequently the term 'designer' is used in its broadest sense, referring to researchers as well as practitioners having design experience (e.g. designing user interfaces) in their job[1].

To empower end-users, a proliferation of PD approaches, methods, techniques, and tools has been developed, for example: an expert designer co-creates with a single end-user to create a 3D mock-up from scratch using clays; a designer elicits ideas on a simple 2D paper mock-up from a group of end-users using coloured pens (see Sanders et al., 2010). The approaches developed and applied to actively include end-users as design partners are as varied as the things that can be designed, including cities (Crewe, 2001; Forlano and Mathew, 2014), digital assistive technologies for people with autism spectrum disorders (Francis et al., 2009), innovative input methods (Kühnel et al., 2011) and digital family calendars (Neustaedter and Bernheim Brush, 2006).

Even when narrowing the scope of products to digital artefacts, the number of PD approaches that can be used is still high (e.g. Sanders et al., 2010). Paper-based approaches are widely used to present as well as capture design ideas in PD (Kelly et al., 2006; Muller, 1991). Examples of paper usage in PD activities are storyboards (Marois et al., 2010), 'Comicboarding' (Moraveji et al., 2007), sticky notes (Carmel et al., 1993; Druin, 2002; Svanaes and Seland, 2004), 'bags of stuff' (including paper among other materials, Druin et al., 2001), 'big paper' (Guha et al., 2004), and 'Layered Elaboration' (Walsh et al., 2010).

Although the use of paper to gather PD ideas has a long history and is well established in the PD research area, some researchers used software tools in PD activities due to several assumed beneficial qualities of tool over paper. Naghsh and Andy (2004) applied a software tool instead of paper as it could support the evaluation of dynamic behaviour in a prototype and Walsh et al. (2012) introduced a tool to help keeping track of PD ideas assigned to different versions of a constantly changing prototype (Walsh and Foss (2015) presented an improved version of the same tool,

[1]This inclusive use of the term designer is not arbitrary. Especially in small research projects and small and medium-sized enterprises (SMEs) the roles of graphic designer and developer are not strictly separate and instead combined in one person. For example for SMEs a reduced set of Rational Unified Process roles has been developed (Monteiro et al., 2012) in which the roles 'Designer', 'User-Interface Designer', and 'Graphic Artist' are all included in the role 'Implementer'.

supporting, for example, audio input). These software tools were all developed to support a distributed PD setting (Danielsson et al., 2008; Gumm, 2006; Lohmann et al., 2008; Walsh, 2011), where digital data simplified the exchange between end-users and designers.

However, Naghsh and Andy (2004); Walsh et al. (2012); Walsh and Foss (2015) did not compare their software tools with corresponding paper-based approaches. One of the open questions regarding paper-based approaches and software tools applied in PD is whether the quantity and quality of ideas gathered using a software tool are comparable with those elicited when using its paper-based counterpart. This comparability has been under-researched.

Such a comparison is essential to prove that the new tool-based method works and that using it yields similar or better results as compared with using paper. Only then it can be determined and advised in which context which method should be deployed. An example for such a comparison in an area closely related to PD is the evaluation of CanonSketch (Campos and Nunes, 2007), a design tool for User-Centred Design (UCD) activities. Contrary to the philosophy of PD, this tool did not include actual end-users in the design process. For CanonSketch it was shown that the tool outperformed paper, for the most of the tested qualities even significantly (Campos and Nunes, 2007). This motivated the author to conduct comparison studies with software-based and paper-based approaches in PD.

The selection of a particular PD approach depends on several interrelated factors: the nature and current state of the project (e.g. timeline), the characteristics of the end-users involved (e.g. design competence), the medium of the product (e.g. web-based), and constraints of the application context (e.g. classroom). With software tools as possible alternatives to paper-based approaches, guidelines for selecting which of the two approaches to use in PD are lacking. These gaps could be bridged by the development of a dedicated PD tool and by empirical findings of comparing paper-based and tool-based approaches.

Apart from the above mentioned literature review on PD work, general design research literature was reviewed for analysing strengths and weaknesses of paper- and tool-based approaches (Campos and Nunes, 2007; Guo et al., 2007; Landay and Myers, 1995, 2001; Van de Kant et al., 1998). Insights thus gained could inform the work of PD presented in this thesis.

To summarise the motivation of this research: Participatory Design empowers end-users by actively including them in the design process of any service or product, namely digital artefacts for the scope of this thesis. To do so, a proliferation of approaches has been developed, most of them paper-based, but some using software tools. Although essential, a comparison of paper- and tool-based approaches is mostly missing in the existing PD literature and therefore addressed.

1.2 Research Questions

From the general assumption that software support might benefit end-users and designers in PD activities and the observation that the comparison of paper- and tool-based PD methods is inadequate, the following research questions have been derived:

- To what extent can software tools **replace** current paper-based participatory design approaches
  - in **gathering** ideas from end-users (RQ1) and
  - in **analysing** the ideas of end-users by designers (RQ2)?
- To what extent is the **number of ideas** captured by tool-based PD activities different from that by their paper-based counterparts
  - using an **off-the-shelf tool**, (RQ3) and
  - using a newly created, dedicated **PD tool** (RQ4)?
- How are the ideas captured by tool-based PD activities **qualitatively different** from those by their paper-based counterparts
  - using an **off-the-shelf tool** (RQ5), and
  - using a newly created, dedicated **PD tool** (RQ6)?

1.3 Approach

As the existing PD tools are not available for testing (Walsh et al., 2012) or have specific hard- and software requirements (Naghsh and Andy, 2004), they could not be used for the research presented in this thesis. This motivated the evaluation of existing, available tools (Section 2.4) and, based on the results (Section 3.3), the development of a new software tool for PD activities.

The strengths and weaknesses identified for paper and tool (Section 2.2) were taken into consideration when developing the dedicated PD tool. An example for this was the decision to only include freehand drawing and not to provide pre-defined shapes. Although the latter might have the benefit to speed up the design progress, freehand drawing instead addressed several other points: The weakness of 'thinking in known patterns, less creatively' (Van de Kant et al., 1998) when using a tool offering pre-defined shapes could be avoided and the strength of using paper to 'focus on content and functionality (conceptual design) rather than details (alignment and colours)' (Landay and Myers, 2001; Lin et al., 2002) could be brought into the tool.

The new tool I developed was called *Participatory Design online tool (PDot)*. It consisted of two web-based Graphical User Interfaces (GUI) accessing the same data. One GUI was aimed at end-users - *PDotCapturer* - focusing on the idea gathering aspect of PD (Figure 1.1, Chapter 4). To further support the task of idea analysis, a second GUI was developed which was aimed at designers - *PDotAnalyser* - which not only presented the gathered ideas in an aggregated way but also offered additional visualisation and filter options (Figure 1.2, Chapter 5). PDotCapturer and PDotAnalyser have been developed in iterative circles of development and evaluations to ensure good usability, positive user experience, and appropriate functionality.

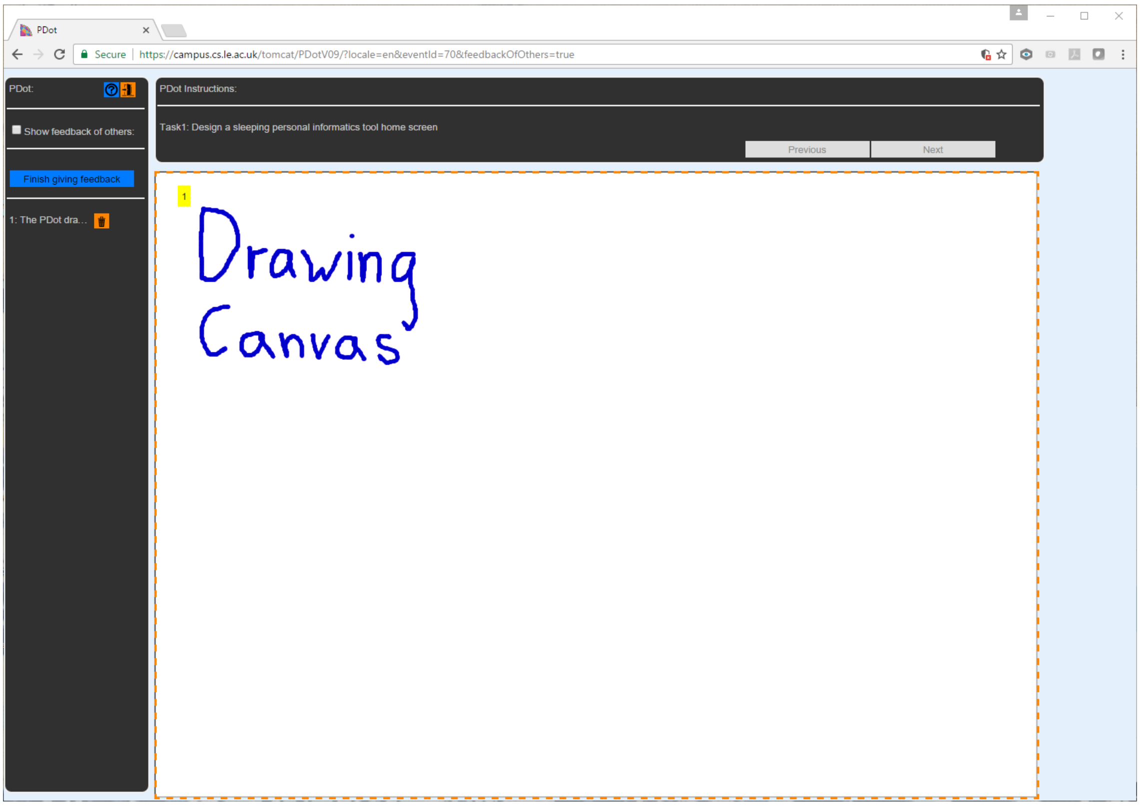


Figure 1.1: Screenshot of PDotCapturer.

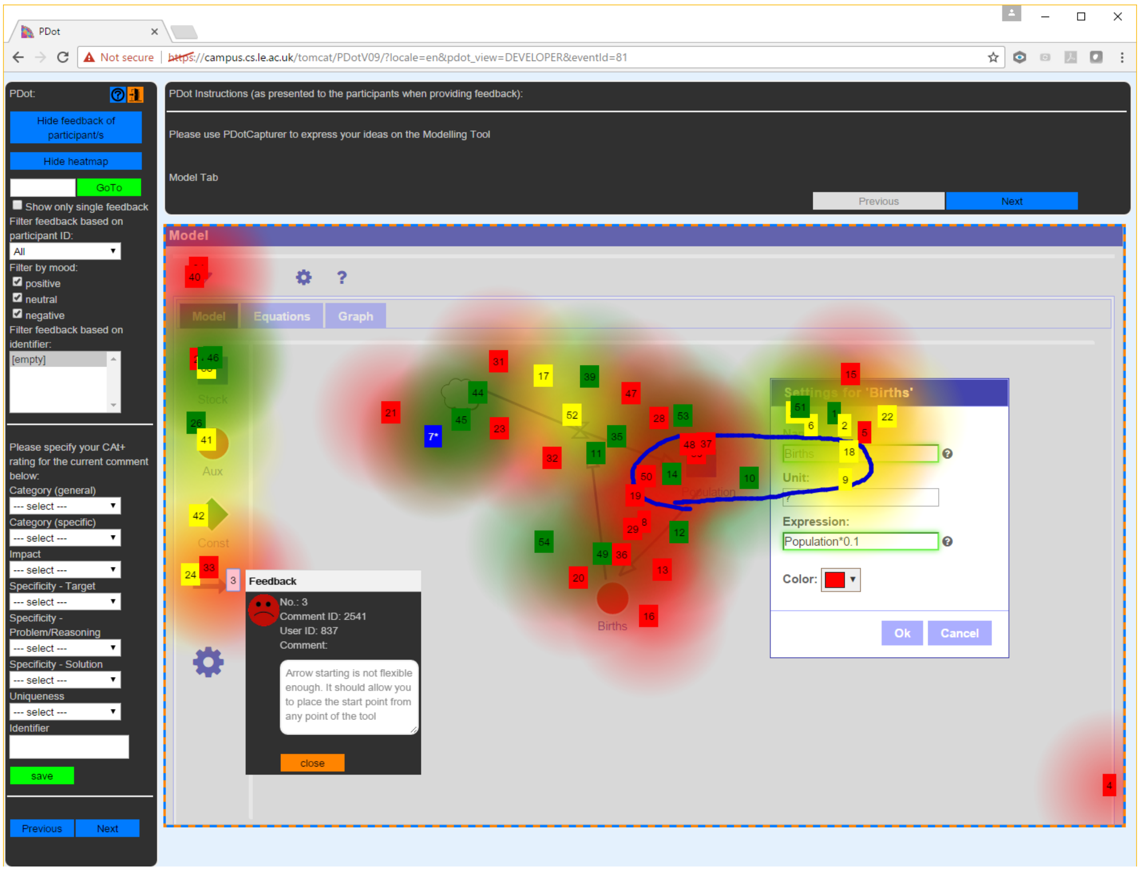


Figure 1.2: Screenshot of PDotAnalyser.

Paper-based approaches were compared with an off-the-shelf tool and different versions of the PDotCapturer tool in a co-located setting (to keep the number of variables that could not be controlled low) to find commonalities and differences between paper- and tool-based participatory design activities (Section 6). Most of the PD activities and studies presented in this thesis have been conducted in the context of the Go-Lab project which is presented in the following sections.

Goal and output of the Go-Lab project

Go-Lab (Global Online Science Labs Inquiry Learning at School) was a European project which provided the context of this research. Its goal was the enhancement of science lessons by integrating online labs (remote and virtual laboratories or scientific data available on the Internet, De Jong et al., 2014; Govaerts et al., 2013). This was done by enabling teachers to create online lessons containing their own learning content, extended through resources available on the Internet, digital scaffolding applications (which support learning) and an online lab.

The applications and online labs, together with example lessons created by project partners and fellow teachers could be found on a portal developed and filled with initial content by the project (Govaerts et al., 2013). This portal and its content were the webapps for which most of the PD activities presented in this thesis were conducted. They were developed for and with students and teachers, online lab providers, and researchers spread all over Europe.

The webapps (a collective term used to refer to a variety of web-based applications, websites, and online portals) of interest for this work were therefore software applications for enhancing the use of online learning resources. The end-users of these webapps were school children and their teachers using online labs and ancillary apps to learn and teach science subjects. This setting allowed for the evaluation and comparison of tools and paper with a variety of different prototypes (from small applications with a single purpose to large websites hosting a wide selection of content) and different end-users (school children, university students, teachers).

The Go-Lab project developed new and innovative online labs and scaffolding apps as well as collected existing ones in a repository website (Govaerts et al., 2013). To ensure compliance of the developed digital artefacts with the requirements and visions of the prospective users, a PD approach has been applied in the project.

PD activities in the Go-Lab project

In the Go-Lab project, the prototype development phase was kicked off more or less at the same time as the requirements engineering (RE) process involving end-users. This set-up was somewhat unusual, because typically, at least in the initial project phase, a user-based RE process takes place prior to implementation. However, in the context of Go-Lab, some mock-ups of the main features were created based on the legacy of related projects. These mock-ups were used as the base for the PD activities, where they have been further refined, instead of starting the design with end-users from scratch.

To some extent this could be considered as advantageous, given that bootstrapping a new system could have been very challenging and time-consuming, and it is always advisable to avoid reinventing wheels. Adapting existing designs with the ultimate goal of enhancing them could be more cost-effective than creating the design from scratch, especially as potential end-users had limited or even no experience with online labs. Nevertheless, ongoing involvement of end-users to improve the proposed design was indispensable. This complied with the principle of PD, whereby end-users could exert direct influence on the proposed design by voicing out their needs or preferences and by articulating their ideas that were seriously considered and addressed during the development process.

Like in any diverse setting, PD activities were very helpful in the Go-Lab project. To ensure the final webapps were applicable and usable in the varied usage scenarios given by the Europe-wide use of the project results, the ideas of many different people in their specific context should be captured (e.g. the requirements of a Cyprian student might differ a lot from those of a Dutch teacher). Thus PD methods and tools which support the participatory design of webapp prototypes to inform the software design in Go-Lab were of special interest.

To be able to offer a wide selection of labs and apps to teachers and students the Go-Lab project made use of existing resources as well as developing new ones. The overall goal of the PD activities was therefore twofold in the project: First, existing apps and labs had to be improved together with end-users, if necessary, to be added to the portal. Second, prototypes of the portal and newly developed apps and labs needed to be designed with prospective end-users using PD methods.

Noteworthy is that the Go-Lab project only provided the background and application options for the PDot tools to be designed, developed, and evaluated in the course of this research. Besides requirements and PD input and feedback, especially from the technical project partners, PDot was solely developed by the author of this thesis.

1.4 Main Contributions

Research literature is rich when it comes to supporting expert designers with software tools (e.g. Guo et al., 2007; Klemmer et al., 2000; Landay and Myers, 1995, 2001; Newman et al., 2003; Van de Kant et al., 1998) but sparse regarding **tool-support for participatory design activities** including end-users (e.g. Naghsh and Andy, 2004; Walsh et al., 2012). This imbalance is addressed through research questions targeting software-supported PD.

Existing PD software tools are either limited in their applicability (e.g. only work in combination with particular existing software or have specific hardware requirements like the need for graphic tablets as input devices, Naghsh and Andy, 2004), are not publicly accessible for testing and improvement or have software requirements, like plugins that are needed to run them (e.g. Walsh et al., 2012). This gap of a **flexibly usable software tool for PD activities** is addressed by developing PDotCapturer, a software tool that runs in any major browser with keyboard and mouse as commonly used input devices without any additional installation (of a plugin or the tool itself).

PD software tools are presented and evaluated in the literature (e.g. Naghsh and Andy, 2004; Walsh et al., 2012), but they are not compared with their paper-based counterparts. This lack of **comparison between paper and tools in PD** is addressed by performing and analysing several studies comparing different paper-based PD methods with tools offering similar functionality. The identification of commonalities and differences between paper and tool can help PD researchers and practitioners to decide which approach to use in a particular context.

A **coding scheme** is needed to compare paper and tool not only quantitatively but also qualitatively. Existing schemes are tailored for their use-cases and content coded (e.g. Kindred and Mohammed, 2005; Könings et al., 2010; Madden et al., 2013; Stumpf et al., 2007). This lack of a coding scheme for PD ideas gathered on webapps is addressed by developing the coding scheme CAt+.

Besides end-users also **designers could benefit from further support in PD activities** (Read et al., 2016). This is addressed by developing PDotAnalyser, a software tool partly automating and thus supporting PD idea analysis. CAt+ is integrated in the tool to evaluate how a coding scheme initially developed to compare paper and tool PD ideas can be applied to support designers in their analysis task.

1.5 Outline of the Thesis

1.5.1 Structure

Chapter 2: Participatory Design is a broad area. Therefore it overlaps with other approaches, like user-centred design. To clarify the scope of my research and its implications, the research area in which the research took place is defined and located in the Human-Computer Interaction (HCI) research field. The state of the art regarding software support for participatory design is presented. A lack of a proper software tool supporting gathering (and analysis) of PD ideas adequately is identified. As a first step to close this gap an evaluation of existing online tools that could support PD idea gathering is presented.

Chapter 3: To allow for the qualitative comparison of differences between paper-based and tool-based approaches the coding scheme CAt+ is developed and refined. A paper-based PD approach is compared with the off-the-shelf software tool selected for PD in the previous Chapter 2. The results lead to the development of a dedicated tool.

Chapter 4: The design, development, and evaluation of PDotCapturer, an online tool tailor-made to gather PD ideas from end-users, is presented. This tool provides either an empty page or allows for the integration of an existing (interactive) prototype on which end-users can express their PD ideas using freehand drawing, textual description, and other feedback modalities. PDotCapturer can also be used to access and visualise the ideas to support designers in their idea analysis task. But this is not the intended use-case, and it is tedious to do so, thereby calling for a dedicated tool supporting designers.

Chapter 5: The design, development, and evaluation of PDotAnalyser, an online tool to support designers in the task of PD idea analysis, is presented. It can either be used by end-users and designers together to facilitate their communication, or by designers on their own to support their task of making sense and addressing the ideas when applying the (re-)design to the prototype. The CAt+ coding scheme is integrated into PDotAnalyser to support designers in their analysis task, by first classifying and then filtering PD ideas using the categories and attributes defined by the coding scheme.

Chapter 6: Paper-based PD approaches are compared with PDotCapturer. This addresses the question if a dedicated tool can produce better results than the off-the-shelf tool used in Chapter 3. Besides this comparison of paper and tool a study letting end-users create designs from scratch, a common PD scenario, is presented to compare Layered Elaboration (paper-based) and PDotCapturer (tool-based approach) in this setting.

Chapter 7: Based on the findings presented in previous chapters specific implications regarding the research questions and general implications are drawn. Additionally general limitations of the research presented in this thesis are presented.

Chapter 8: In the final chapter concluding remarks and an outlook on possible future work are given.

A graphical representation of the content of the different chapters (besides Introduction, General Discussion, and Conclusion and Future Work) can be seen in Figure 1.3.

1.5.2 Publications

CHI 2014

The current state of the then work in progress of the development of a tool to support participatory design was described in a poster titled "Pdot: Participatory Design Online Tool" (Heintz et al., 2014a). It was submitted for the work in progress section of the ACM CHI Conference on Human Factors in Computing Systems. The poster was accepted and was presented at CHI 2014.

Special Issue IEEE STCSN e-letter vol. 2, no. 3 "Large-Scale Social Requirements Engineering"

For a special issue IEEE (Institute of Electrical and Electronics Engineers) STCSN (Special Technical Community on Social Networking) e-letter on large-scale social requirements engineering an article with the title "Review of Online Tools for Asynchronous Distribute Online Participatory Design" was submitted and accepted (Heintz et al., 2014b). It describes an evaluation of existing online tools, not explicitly designed for the use in participatory design, regarding their suitability and usability to be used for PD activities. The article can be found online at `http://stcsn.ieee.net/e-letter/stcsn-e-letter-vol-2-no-3/review-of-online-tools-for-asynchronous-distribute-online-participatory-design`.

EDUCON 2015

A full paper with the title "A Survey on the Usage of Online Labs in Science Education: Challenges and Implications" (Heintz et al., 2015a) about the usage patterns and experiences of teachers and students with online labs was presented at EDUCON 2015. The relation between this paper and this thesis was that the results of this survey also informed the design of the PD tools. Some requirements for a tool that

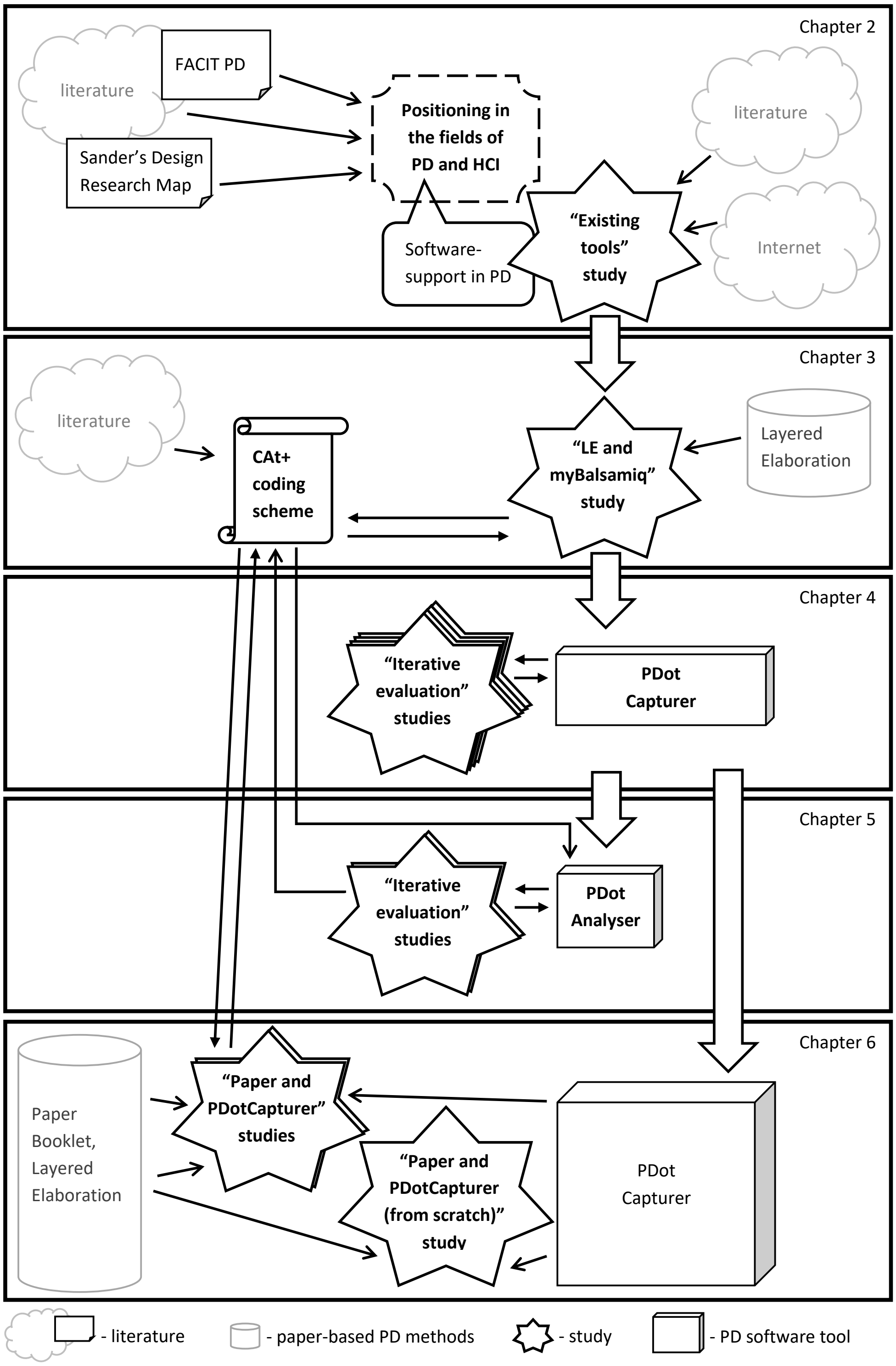


Figure 1.3: Overview of the content of the thesis.

worked in a school environment (e.g. tool must work in all major browsers and without plugins) were informed by the results of this survey.

INTERACT 2015

A full paper with the title “Paper or Pixel? Comparing Paper- and Tool-based Participatory Design Approaches” (Heintz et al., 2015b) about the comparison of paper-based and tool-based Participatory Design approaches and the development of the coding scheme CAt+ was accepted and presented at INTERACT 2015.

LUXReP 2015 Workshop at INTERACT 2015

A paper with the title “Solution-based Requirements Capture with PDot in an E-Learning Context” (Heintz and Law, 2015) containing preliminary usability and User Experience evaluation results for PDotCapturer was presented at the The Landscape of the UX Requirements Practices workshop at INTERACT 2015.

CoPDA 2016 Workshop at NORDICHI 2016

A paper with the title “Challenges and Resolutions for Engaging Teachers and Students in Participatory Design of Online Science Learning Resources” (Heintz and Law, 2016) has been presented at the fourth edition of the International Workshop on Cultures of Participation in the Digital Age: From “Have to” to “Want to” Participate, co-located with the 9th Nordic Conference on Human-Computer Interaction (NordiCHI 2016).

BRITISH HCI 2017

A full paper with the title “Comparing Paper and Software Tool for Participatory Design from Scratch” about the comparison of Layered Elaboration and PDotCapturer when creating a design from scratch was accepted and presented at BRITISH HCI 2017.

Chapter 2

State of the Art

This chapter presents the diversity of the PD research area and positions the research presented in this thesis within it. Advantages and disadvantages of paper-based and tool-based approaches are presented to motivate the usage of software in PD. An evaluation of existing tools to support participatory design leads to the selection of myBalsamiq for the comparison of paper with an existing tool.

2.1 Positioning in the Participatory Design Research Area

2.1.1 Diversity of the PD Research Area

Because participatory design is such a broad area (e.g. Halskov and Hansen, 2015), this section aims to position the tools and approaches applied and developed as part of this thesis in this area. This will give the reader an idea of what to expect and where to locate the research and findings presented.

The wide range of PD can be seen in the following two examples. In the first one the focus is on participation, in the second one the focus is on (co-)design/creation. Hendriks et al. (2014) point out that the "perception of what participatory design is, comes from how PD has traditionally been defined. Most visions on PD date back to the early tradition of the Scandinavian legacy of PD where partners of (relatively) equal cognitive and physical abilities participated in the PD sessions." (Hendriks et al., 2014, p. 34). Nevertheless they call their work with people with dementia, who lack many of the abilities to provide traditional PD input (e.g. verbal communication and visual thinking combined with abstraction), participatory design and present 7 challenges they encountered. Although their paper hints at calling methods that actively include any end-user as partner or co-designer, rather than object of study or informant, participatory design, Hendriks et al. (2014) do not give a specific definition of the term. De Angeli et al. (2014) define participatory development, derived from PD, in their paper as follows: “We reflect on a case of participatory development, which we interpret in the literal sense as leaving the development to volunteers in the community.” (De Angeli et al., 2014, p. 12). The reason for their literal interpretation of the participation term is that they are interested in sustainability and they argue that it can only be achieved if the community is not only involved in the design process but becomes the active part of the implementation. They therefore also acknowledge that their examples of PD and participatory development are an “... extreme case where users themselves develop the result of a PD study” (De Angeli et al., 2014, p. 17) and that they “... push the boundaries of participation to the physical assembly of the artefact.” (De Angeli et al., 2014, p. 18). For this thesis the definition of PD by Carroll et al. (2000, p. 239) is adopted: “Participatory design-also called cooperative design-is the inclusion of users or user representatives within a development team, such that they actively help in setting design goals and planning prototypes”.

The notion of participatory design has evolved since it started out in Europe as a Scandinavian movement in the 1960’s to democratize the workplace. From there it was more widely used to include future end-users in the design and development process not only of their workplace and working conditions but for various physical and digital environments and goods. This development over time and in different places (and therefore environments) in the world is one explanation for the diverse understanding and diffuse definition of PD. As the initial PD movement was adapted in different contexts (for example in the United States and by the HCI community) over time one strand emerged that mainly focussed on creating more user-friendly systems (Clement and Van den Besselaar, 1993; Gennari and Reddy, 2000; Greenbaum, 1993; Gregory, 2003; Muller et al., 1993; Porayska-Pomsta et al., 2012; Puri et al., 2004).

According to Kensing and Blomberg (1998) PD literature was dominated by three main issues: “(1) the politics of design, (2) the nature of participation, and (3) methods, tools and techniques for carrying out design projects.” (Kensing and Blomberg, 1998, p. 168). This research could add to the body of knowledge regarding the third issue by providing and evaluating tools to carry out design projects supporting end-user involvement.

Gartner and Wagner (1996) identified three arenas of participation with different granularity. Arena A is the 'individual project arena' which is concerned with the design of specific systems. Arena B is the 'company arena' which is concerned with designing organizational frameworks. Arena C is the 'national arena' which is concerned with the industrial relations context (Gartner and Wagner, 1996; Kensing and Blomberg, 1998). The research presented in this thesis focuses on Arena A: supporting the design of individual systems on a project level.

PD is one of the many different areas in the inherently heterogeneous field of Human-Computer Interaction (HCI), and it overlaps conceptually as well as practically with some related areas such as usability evaluation. To avoid any potential confusion, it is necessary to position the research work presented in this thesis. This is done by referencing it to a framework of scoping PD (Walsh et al., 2013a) and by locating it within the neighbouring areas of PD (Sanders, 2006).

2.1.2 Classification of this PD Research Work in FACIT PD

Walsh's FACIT PD framework (Walsh et al., 2013a) was identified as one way to classify the PD software tool PDotCapturer (see Chapter 4 for details) and subsequently the PD research and work presented in this thesis as, similar to Walsh et al. (2013a), the target groups included both children and adults. The resulting classification can be seen in Figure 2.1. Specifically, the position of PDotCapturer is indicated with reference to the eight dimensions of FACIT PD along the spectrum of 'less ... more'. For some dimensions this can be done more precise than for others, where PDotCapturer covers a range of possibilities. Only PDotCapturer is positioned in the framework, as FACIT PD was developed to classify idea gathering PD techniques. Nonetheless, PDotAnalyser (see Chapter 5), the software tool developed to analyse PD ideas, could be classified accordingly, as it works with the ideas gathered through PDotCapturer.

The *Partner experience* dimension of FACIT PD specifies the amount of design experience required for participation. For this dimension PDotCapturer is close to the 'no expertise' end as the end-users work on their own and express their ideas using the conventional means of text and freehand drawing. The main expertise needed is knowledge on how to operate the tool. Usability and UX studies with the target groups of the tool show that the usability is suitable for the purpose. Additional expertise needed to express PD ideas using PDotCapturer is covered by general PC and Internet expertise (maybe with a slight exception for 'drawing with a mouse', which might be unfamiliar, but intuitive).

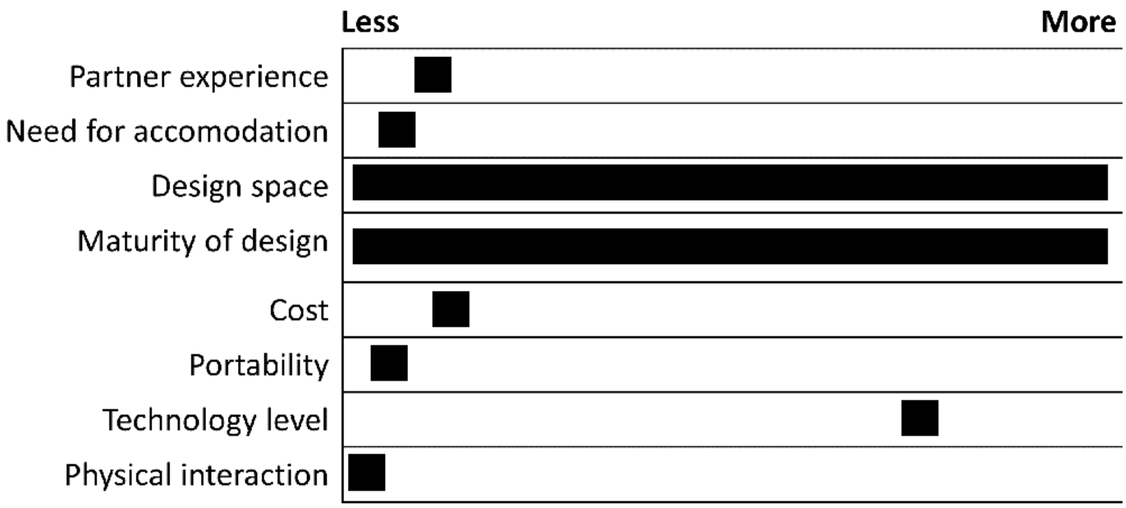


Figure 2.1: Classification of PDotCapturer using the FACIT PD framework (Walsh et al., 2013a).

The dimension *Need for accommodation* is concerned with the question if and how many modifications of the technique are needed based on the age and cognitive ability of the end-users. Accordingly, PDotCapturer can be considered as low, as it simply needs a standard mouse and a keyboard (or input devices providing the same input signal). In cases where end-users cannot interact via such input devices (e.g. people with impaired hand dexterity), a design team member, as proxy contributor, can help express their ideas using the tool.

Design space examines the design problem and how explicitly it is defined. PDot-Capturer is flexible in the design goal that can be specified. The design tasks it can support can range from generically asking for design ideas for a new concept (non-specific) to reviewing particular parts of existing prototypes (highly specified). In Go-Lab and therefore for this research work PDotCapturer has been mostly used on the 'highly specified' end of the design space dimension. Participants have mostly been asked to annotate a specific existing design based on a concrete usage scenario. But PDotCapturer was also used with the open task of designing interfaces from scratch.

This flexibility is also the reason why PDotCapturer covers the whole spectrum of the dimension *Maturity of design*. This dimension looks at the position of the design activity within the complete design process. PDotCapturer can be used early in the design process with a blank (virtual) slate to draw on (see Section 6.3 for details) as well as in later stages to inform the re-design of a developed prototype (see Section 6.2 and 6.1 for details). The scenario in which PDot has been mostly used in Go-

Lab is with a 'mid' maturity of design, being early to late prototypes, for which the participants provided PD design ideas.

The *Cost* dimension is concerned with the monetary value of the materials used. Under the assumption that PC equipment and a flat-rate Internet connection are already present (which is mostly commonplace nowadays) the usage of PDotCapturer is on the no-cost end of this dimension. The idea gathering process is digital and thus does not need any physical material which would cost money.

Portability is concerned with the physical movability of techniques and their outcomes. One could argue that with the tool and ideas expressed being accessible on the Internet from anywhere with a browser, the technique and especially the artefacts generated would be highly portable. But being "tethered by the need to use a computer lab" (Walsh et al., 2013a, p. 2897) is explicitly mentioned as one example for a non-portable technique. As the end-users are required to use a computer to access PDotCapturer it is thus put on the non-portable end of this dimension.

The *Technology* dimension considers how refined the technical equipment applied in a technique is. PDotCapturer, using PCs or laptops, is on the high-tech usage end of this dimension, although not at the very far end, as no special equipment (like cameras) is needed.

The dimension *Physical interaction* measures how much moving around for end-users is caused by a technique. As with Layered Elaboration (Walsh et al., 2010), the analogous technique from which the PDotCapturer interaction metaphor is derived, PDotCapturer is on the low movement end of this dimension. End-users sit at a PC to draw and provide textual comments and therefore do not physically move around when this technique is applied.

The main take-away point of the classification in the FACIT PD framework is on the one hand to get an impression of the properties of this research regarding eight important dimensions of PD. On the other hand it can be used to compare this research and PDotCapturer to other PD work, classified using the same framework.

2.1.3 Positioning of this PD Research Work in a Design Research Map

While it might be relatively straightforward to position PDotCapturer with regard to the FACIT PD framework, it proved more challenging to pin PDot down in a map of research areas of HCI. It is because the demarcation of such areas, like PD, is not clear (Bergvall-Kåreborn and Ståhlbrost, 2008). Nevertheless, I attempted to use Sander's design research map (Sanders, 2006, 2008) to position PDot (Figure 2.2).

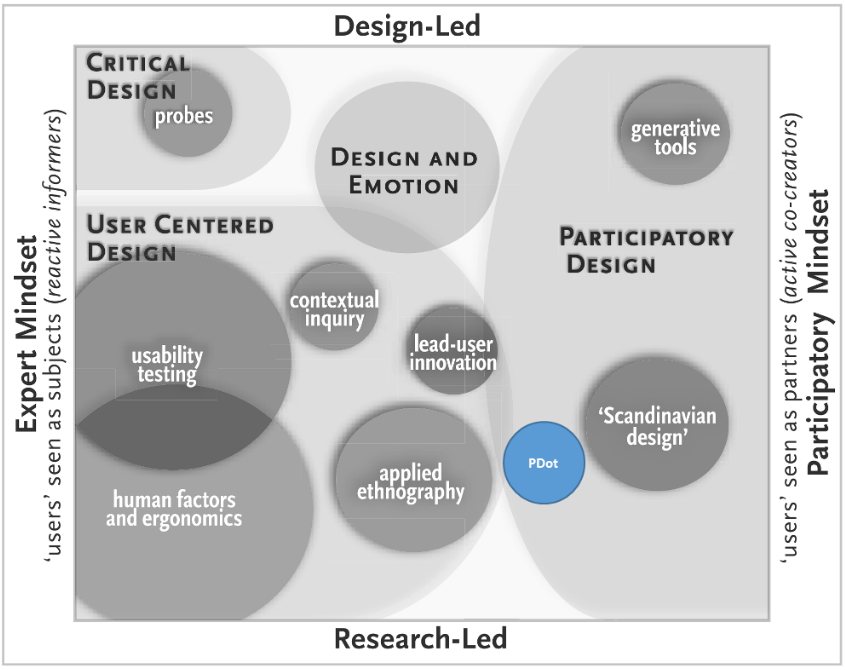


Figure 2.2: PDot located in Sander's Design Research Map (source of figure: Sanders, 2006, desaturated and PDot bubble added).

For the vertical dimension research-led to design-led, PDot is closer to the research-led end, as the tools emerged from the research questions whether and how online tools can support PD activities (Heintz et al., 2014a) and how they are compared to their paper-based counterparts (Heintz et al., 2015b). Instead of defining a new shared design language to be used to allow for and ease communication between end-users and designers, as it is done in approaches falling into the 'generative tools' bubble (close to the design-led dimension end), PDot utilises natural language capabilities for descriptions and freehand drawings for visual expression of ideas.

For the horizontal dimension from Expert to Participatory Mindset, PDot is close to the participatory end, as not the users are the subject/object of study, but the design. PDotCapturer does not focus on the performance of the user (e.g. measuring the time on task as a usability criterion). The emotional responses captured by 'smileys' are primarily used to determine how different parts of the interface are perceived and can be improved, not how the end-user is feeling. An example of the heatmap visualisation created by PDotAnalyser from this information can be seen on the lower left hand side of Figure 5.3. Instead, the PDotCapturer functionality is used to collect the ideas from end-users and their comments on the interface.

While Sanders' map (Sanders, 2008) could be used as a starting point, it has some drawbacks (Stolterman, 2008). For instance, the bubbles in the map are framed in a rather restrictive way as the areas are not necessarily that clear cut. For example the 'usability testing' bubble seems solely seeing the user as the focus of study where applied usability testing should not stop at the list of usability problems but should also come up with improvement suggestions for the interface (Wixon, 2003). Generating useful and usable usability recommendations is recognised as an integral part of usability evaluation to exert real impact on the final product (Hornbæk, 2008; Molich et al., 2007).

Where to draw the lines is not only an issue in the visual representations in Sanders' map. There seems to be a lack of consensus in the HCI community where to draw the boundaries among areas as well, as exemplified by a recent call to reflect on "re-imagining participatory design" (ACMTOCHI, 2016):

> Meanwhile, participatory design often seems to have become synonymous with a more neutral form of 'user-centered' design, concentrating on more local issues of usability and user satisfaction. This is in contrast to earlier work in the field where Participatory Design not only sought to incorporate users in design, but also to intervene upon situations of conflict through developing more democratic processes. (ACMTOCHI, 2016).

PDotCapturer provides functionality to end-users to inform the graphic and interaction design of software products and thus falls into the area of PD focusing on local issues described in the quote above. However, it supports the PD *philosophy* of treating end-users as design *partners* (e.g. Carroll et al., 2000). In the debate in the field regarding the question whether PD should open up or need to go back to its roots (Bergvall-Kåreborn and Ståhlbrost, 2008) it is therefore argued to open up; addressing local issues with a participatory mindset is a valuable part of the PD area.

2.1.4 Further Locating this PD Research Work

The general project setting and time plan, and access to members of the target groups also had an influence on the level of participation for the PD activities in the Go-Lab project. The current state of development of the webapps, when performing the PD activities presented in this thesis, asked for collecting ideas on prototypes rather than starting from scratch. In the IBF (Informant, Balanced, Facilitated design) model by Read et al. (2002), this means the Go-Lab PD activities reported in this thesis were on the Informant Design end of the scale. Designers provided initial designs, interactivity and design were then evaluated by end-users. The evaluation results were then considered by the designers when re-designing the initial prototypes.

On a different, unnamed scale Read et al. (2016) classified PD with school children (one of the main target groups of Go-Lab) based on the exposure of students to the PD activities, ranging from long activities with small groups, over medium length activities with medium size groups to very time-constrained activities with a large number of students. Based on the availability of end-users, in Go-Lab PD workshops with class-sized student groups (ca. 25 of them) or larger groups of teachers (up to 40) were performed in timeslots of only a few hours each. The workshops were therefore on the latter end of the scale, called “the ‘fast and furious’ end of participation in design” by Read et al. (2016, p. 1).

2.2 Advantages and Disadvantages of Paper and Tools in Design

To get an idea where a software tool could support or replace paper, advantages and disadvantages of both media were identified. In addition to the findings from PD research (already presented in Section 1.1) general design research literature was also reviewed to identify strengths and weaknesses of paper and tool for design activities in general and software design in particular (Campos and Nunes, 2007; Guo et al., 2007; Landay and Myers, 1995, 2001; Van de Kant et al., 1998). The results for end-users and designers are summarised in Table 2.1 and described in more detail below.

Strengths of paper-based approaches: Paper-based approaches in software design in general and PD in particular have several advantages (Landay and Myers, 2001; Rettig, 1994; Van de Kant et al., 1998). They can be used by non-computer literate participants and they are cheap and flexible (Van de Kant et al., 1998). With the low-fidelity and ambiguous nature of quickly sketched interfaces (Wong, 1992) they can help to focus on the bigger design and interaction picture rather than getting lost in tiny aesthetic details of single interface elements (Lin et al., 2002; Van de Kant et al., 1998).

	Strengths		Weaknesses	
	General Design	**Software Design**	**General Design**	**Software Design**
End-User				
paper	flexible	no PC experience needed focus on content and functionality rather than details	abstraction from the final design	prototype not interactive media disruption between prototype on paper and tool on PC screen
tool	supports distributed setting	presents prototype in 'actual environment' support features (e.g. undo and redo) pre-defined shapes speed up design process supports interactivity	equipment needed more effort drawing on screen less natural	complicated tools for experts thinking in known patterns, less creatively
Designer				
paper	cheap flexible	no PC experience needed focus on content and functionality rather than details	no versioning support no analysis support	not interactive
tool	supports distributed setting versioning quick retrieval data analysis can be partly automated	supports interactivity pre-defined shapes speed up design process digital data is gathered	equipment needed more effort	thinking in known patterns, less creatively

Table 2.1: Strengths and Weaknesses of paper and tool in (participatory) design from end-user's and designer's perspective.

Strengths of tool-based approaches: Software tools present the software prototype in its 'actual environment' (e.g. on a PC screen rather than a table top, using input devices like mouse and keyboard for the interaction rather than pointing with fingers on paper). Additionally they offer support features (like undo and redo), which are more complicated (undo) or not possible (redo) on paper.

Tools for software design often come with pre-defined shapes. This can speed up the design process as commonly used interface elements are readily available and do not have to be re-created from scratch. When using a software tool to design, the virtual environment offers the option of having interactive designs (Newman et al., 2003; Walker et al., 2002). When applying PD approaches to review existing digital prototypes of a product or service, PD software tools presenting interactive prototypes allow the end-user to interact with the prototype and therefore experience the use more realistically, especially when working with prototypes of higher maturity (e.g. Lin et al., 2000; Rogers et al., 2011; Sundar et al., 2012; Teo et al., 2003; Zhao and Lu, 2012). Such interactivity cannot be supported by paper-based PD approaches representing the artefact to be re-designed only as non-interactive printouts to be annotated. The interactivity of the prototype could create a more engaging situation and could result in more ideas being gathered of possibly better quality, as compared to the paper-based approach.

Most of the PD approaches, techniques, and tools rely on the designer addressing the ideas of end-users while the ideas are expressed. However this is not possible in all PD scenarios, for example when data gathering and data analysis have to be done after one another due to time- or location-constraints. In order to be able to effectively analyse physically represented ideas and therefore data without or with a software tool, it is beneficial or even necessary to digitalise them first. But the manual conversion process can be time-consuming and labour-intensive. Software tools supporting PD activities with end-users can help overcome this shortcoming by allowing ideas to be collected digitally to begin with. Digitalisation of PD data at the time of capturing can therefore enhance the effectiveness and efficiency of data processing overall (e.g. data loss can be mitigated). The digital storage of artefacts created with software tools allows for easier versioning and retrieval than paper (Klemmer et al., 2000).

Depending on the setting and number of ideas gathered the manual analysis and further processing of ideas to reach an end-user inspired (re-)design can be cognitively very demanding. Based on the meta-information gathered together with the PD idea (e.g. location, emotional response of the end-user) a software tool can automatically aggregate and filter the data presented to the designer at any given time, based on criteria specified by the designer. By hiding information currently not relevant to the design task at hand, information overload for the designer can be prevented. On paper this process of identifying currently relevant information would have to be performed by the designer previously or as part of the data analysis task. Software tools can therefore support designers through automation (at least partly) of this process.

Weaknesses of paper-based approaches: One of the main weaknesses of paper-based design approaches is that the design and its interface elements are not interactive (Landay and Myers, 1995; Van de Kant et al., 1998). To mitigate this issue in some paper-based design methods one of the designers 'plays computer' (Rettig, 1994), exchanging interface elements based on the participant's interactions. Although this provides some degree of 'interactivity' it is not on a par with the interactivity a digital system provides (e.g. response times, visualisation of changes). If the prototype is presented interactively on a screen and paper is only used to capture the ideas, there is a media disruption between where the prototype is presented and where the ideas are specified. With regards to the design of information architecture of websites, Designers' Outpost was developed to overcome the identified shortcomings of putting paper sticky notes on walls, e.g. versioning being unfeasible (Klemmer et al., 2000). For the analysis task, paper-based approaches rely on manual piling and sorting of the design ideas and notes by designers (e.g. Rettig, 1994).

Weaknesses of tool-based approaches: The pre-defined shapes in some software design software also come with drawbacks (Landay and Myers, 1995, 2001). The pre-defined design elements can cause participants to think in known patterns, thereby limiting their creativity and hindering their 'out-of-the-box' thinking (Van de Kant et al., 1998). At the same time designing an interface using pre-defined shapes can elicit PD ideas on these shapes, although the goal is to collect insights on the design created by combining shapes. This focus shift towards details can be a limiting factor (Guo et al., 2007; Landay and Myers, 2001; Rettig, 1994; Walker et al., 2002) especially in initial design stages. For example instead of thinking about the question if the prototype needs a printing functionality or not, end-users can end up contemplating if the button needs a label besides the printer icon or not. Commercial tools used by professional designers to design prototypes, like Adobe Photoshop (Newman et al., 2003), are highly specialised with a steep learning curve and are therefore not suitable for end-users to create design alterations and express their ideas.

There are also some constraints from the end-user's as well as designer's perspective for using a software tool as compared to using pen-and-paper, e.g. drawing on a screen is less natural than on paper (Weibel et al., 2011) and computer access is needed to perform the PD activity (to mention one example for each). However, it can be assumed that the benefits (more engaging, digital data gathered and results to work with) could outweigh those constraints. To test this assumption several empirical studies have been conducted, comparing paper-based with tool-based PD activities and their results.

Sketch-based design tools - Hybrid approaches: Sketch-based software tools, like SILK (Landay and Myers, 2001), can combine the strength of paper-based and tool-based approaches (Guo et al., 2007; Landay and Myers, 1995; Newman et al., 2003). However they rely on specific hardware (design tablets), so even with the more and more widespread distribution of tablets and other computing devices with touchscreens (and stylus support), they are not yet commonly available enough for general PD activities. Additionally, even most recent styli have disadvantages not relevant for paper, like latency and offset between real stylus tip and virtual line, negatively affecting the usability (Helps and Helps, 2016). For general PD research it would thus be interesting to find out what could be done without a stylus, using standard mouse and keyboard as input devices.

2.3 Software Tool Support for PD

For gathering and analysing PD ideas on webapps, three basic types of approaches have been identified:

- **Approaches without software tool support:** For PD activities, a number of techniques has been developed that do without the support of a software tool (Sanders et al., 2010; Walsh et al., 2013a), many of which evolved from general, paper-based software design techniques like Storyboarding (Andriole, 1989) and Paper prototyping (Snyder, 2003). One of them is a paper-based method called Layered Elaboration (Walsh et al., 2010), where an initial design created by end-users on an empty slate or a picture of the prototype to be redesigned is covered with acetates on which the end-users can then freely scribble on and write comments. The acetates can be put on top of each other, either to allow end-users to comment on their peers' ideas or for designers to get an aggregated overview of the ideas of several end-users. Further sophisticated analysis can involve quite a few rather tedious and time-consuming steps (e.g. scanning acetates and mock-ups, transcribing comments and describing drawings in a spreadsheet).

- **Approaches with software tool support for the gathering, but *not* for the analysis of ideas:** There are several software tools that can be used to gather ideas in PD activities, which offer no dedicated support for the analysis of the ideas gathered. One of them is DisCo (Walsh et al., 2012), which is a digital representation of the Layered Elaboration approach described above. But instead of providing software-supported analysis when using DisCo, the ideas are transferred to paper sticky notes and arranged on a physical whiteboard for further interactive discussion by end-users and designers. Software could support and (partly) automate this process.

- **Approaches with software tool support for the gathering and analysis of ideas:** Recently some PD software tools have emerged which provide support for the gathering and analysis of ideas. One of them is LaDDI, which is part of the Online Kidsteam Environment (Walsh and Foss, 2015). It has been developed to capture likes, dislikes, and textual design ideas of end-users. To support the analysis task, an additional tool was created, enabling designers to arrange the ideas on a virtual whiteboard (e.g. for clustering and frequency counts). Although there is some software-support for the analysis work, the main task (arranging the comments) is still done manually. As the idea expression functionality of LaDDI is restricted to text, the end-users can express their thoughts less freely (e.g. positioning of ideas in relation to their targets and freehand drawings to provide visual design ideas are not supported).

For all three types of approaches, automated software-support for the analysis task could be beneficial but is not provided adequately. To bridge this gap PDotAnalyser was created.

Instead of software-support other approaches have been developed to support designers in following the PD principles, one of which is making sure all ideas have been considered for the final redesign. To address this issue RAId (Rapid Analysis of design Ideas) was developed (Read et al., 2016). This approach defines lenses for the designers to prevent information overload and fatigue, by allowing them to focus on specific aspects of ideas when considering them. However, applying these lenses can still be mentally demanding, as the designer has to make an effort to actively ignore the aspects of ideas not in the focus of the lenses currently applied. Software support could automatically hide ideas currently not of interest and therefore lead to a lower cognitive workload for the designer.

Besides analysis support existing PD tools (e.g. Walsh and Foss, 2015) lack an essential functionality to gather design ideas: graphical input. Although this is in various degrees a very valuable part of many traditional PD techniques (e.g. sticky notes, paper prototyping, layered elaboration, see e.g. Walsh et al., 2013b)), existing PD solutions still heavily rely only on verbal input expressed through text. The end-users cannot actively participate in the actual design, e.g. by drawing on the proposed design solution.

There have been some attempts in the scientific community to include graphical besides written input functionality in tools for PD. One of them is GABBEH (Naghsh and Andy, 2004), an electronic paper prototyping tool which allows end-users to comment on the current design. Another one is DisCo (Walsh et al., 2012), an online tool to support intergenerational PD sessions. However, DisCo is not yet publicly available for use or even testing and GABBEH only works together with the DENIM tool (Newman et al., 2003). This essentially excludes them from the review.

2.4 Existing Software Tools for Participatory Design

2.4.1 Overview

This chapter is mostly based on a short paper published in an IEEE E-letter (Heintz et al., 2014b). A general key issue when using any online tool is the technical barrier which is caused by the basic requirement to use computers and digital artefacts. E.g. the technical infrastructure needs to be available for every participant and also knowledge how to operate the digital tools is required.

With knowledge and infrastructure given, it can be assumed that the advantages can outweigh the disadvantages. As compared to face-to-face workshops, online tools are available all the time from any computer with Internet connection. Thus end-users can participate whenever and wherever it is convenient for them. Exchange and sharing of physical artefacts (e.g. paper with scribbles) created during PD sessions can be expensive and complicated. With online tools every stakeholder can access the results directly via the Internet. As PD ideas gathered by an online tool are in digital form, this facilitates the retrieval of raw data as well as the results derived.

As none of the existing tools dedicated to PD seems to be sufficient for the PD activities in the Go-Lab project, in this chapter the functionality and usability of existing, publicly available tools are evaluated that are originally not designed for but identified to be useful for PD activities and allow graphical input as well as textual comments.

2.4.2 Functional Requirements

Based on a literature review (especially Muller, 2007; Sanders et al., 2010; Walsh et al., 2013a), a questionnaire about technical conditions at schools (the one presented in Heintz et al., 2015a), empirical observations in previous PD projects and the requirements in Go-Lab, five functional requirements (R) were identified for annotation tools that can support the Go-Lab PD activities for websites and web applications. Each of them is designated with a code R1, R2 and so forth for later reference.

- **R1:** Tool has to work without installation.
- **R2:** Tool has to run in Internet Explorer, Mozilla Firefox and Google Chrome browser, without plugin.
- **R3:** Tool has to provide the option to enable a user to put textual description of the PD idea proximal to a specific interface element of a webpage, not just the whole page.
- **R4:** Tool has to provide at least basic drawing functionality for non-textual expression of ideas.
- **R5:** Tool has to work with interactive prototypes.

To find a tool which is universally applicable both if installation is possible and also if not (with the latter usually being the case in schools), installation of the tool or a browser plugin should not be required. Additionally to maximize the number of potential participants, the technical barrier for the end-user should be minimized. R1 and R2 ensure this low technical barrier. The existing software setup on the end-users computers enables them to participate in the PD activities. Nothing needs to be installed besides one of the major browsers. Atterer and Schmidt (2007) showed the willingness of users to take part in a remote usability test is higher if no change on the clients side is required compared to reconfiguration of the browser and installation of software. As remote usability testing is similar to PD activities this finding can be transferred.

R3 ensures that the expression of the idea will be as specific as possible. If the tool enables users to position an idea directly on interface elements of the prototype, they can show its physical location visually instead of describing the considered element verbally (e.g. the button at the top right hand corner). Nevertheless, written text alone is not sufficient. To fully support PD of the interface, also non-textual expression of ideas has to be possible, resulting in R4.

The tool used for PD activities should be suitable for the early design stages as well as the subsequent phases of interface development. Ideas should not only be gathered on the graphical design but also on interaction, screen-flow, and transitions. Therefore experiencing the prototype and the individual pages in their context needs to be possible. Additionally interactive prototypes have been identified as one possible advantage of using tools. These are the reasons for R5.

Based on the above mentioned five requirements a Google search for tools and web applications has been conducted with search terms 'webapps annotating websites' and 'design feedback on websites'. The five listed requirements only covered the desired functionality of the tool. Although not explicitly listed here, qualitative requirements, like ease of use, have been implicitly considered while checking for the fulfilment of the functional requirements.

2.4.3 Results of Tool Search

Figure 2.3 illustrates the search results by visualizing tools which come close to fulfil all the five requirements (red) and the tools which do (green). R1 and R2 have been applied as filter criterions first when evaluating the search result, thus only two tools are included in Figure 2.3 which require installation as standalone application (⑨ Jing, techsmith.com/jing.html) or as browser plugin (⑦ Floatnotes, floatnotes.org). They are highlighted in Figure 2.3 by an unfilled circle. All the other tools shown fulfil R1 and R2.

Furthermore, the tools are grouped in two ways: Firstly, the grouping is based on the modality for idea expression supported, for instance, textual through notes and/or graphical through basic drawing (cf. the two ovals in Figure 2.3). Secondly, the tools are grouped according to whether they allow commenting on an interactive

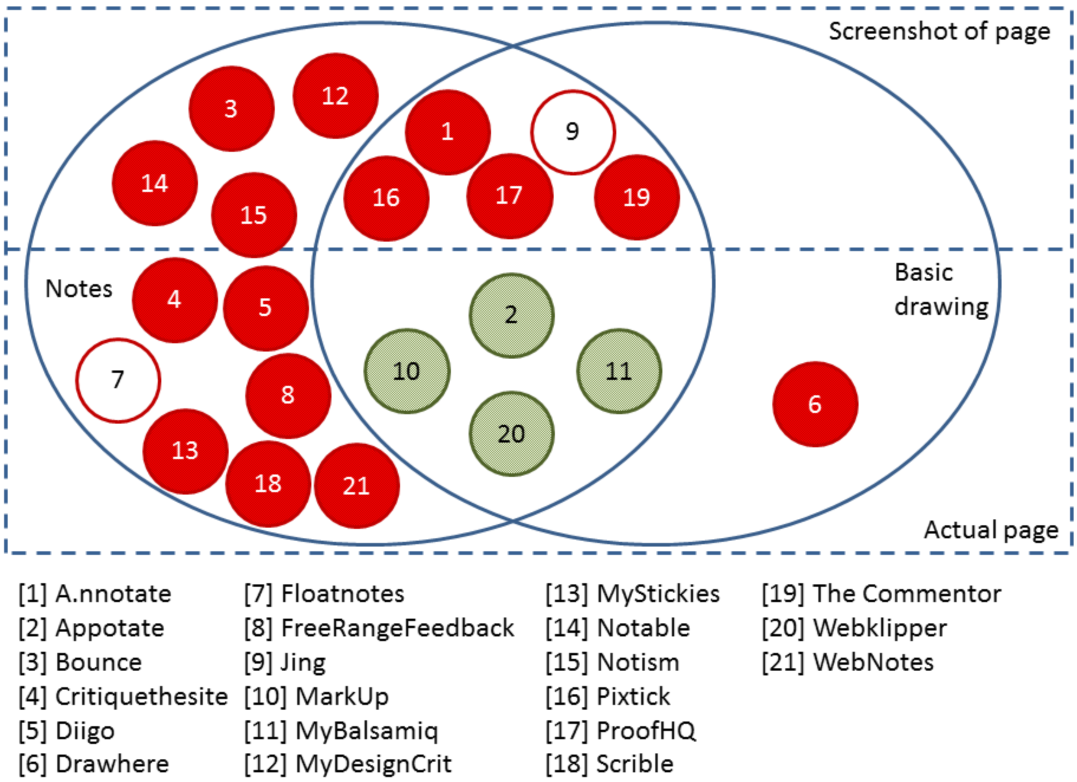


Figure 2.3: Tools grouped by kind of idea expression supported on which form of the page.

prototype or webpage (actual page) or only an image or screenshot (cf. the two dashed rectangles in Figure 2.3).

The first grouping option has been derived from R3 and R4: Notes are needed to give textual descriptions, basic drawing functionality for non-textual expression of ideas. Thus tools falling into the intersection of the two ovals fulfil those two requirements.

The second grouping option has been derived from R5: Working on a screenshot of the webpage is not enough, as this will not allow the user to experience the actual interaction with the prototype. Only tools in the lower part of Figure 2.3 fulfil this requirement. ⑮ Notism crosses the border between screenshot and actual page to a limited extent, because it allows to create interactive areas on pictures, which makes them somewhat interactive.

The top nine tools allow only work with screenshots: ① A.nnotate, a.nnotate.com; ③ Bounce, bounceapp.com; ⑨ Jing; ⑫ MyDesignCrit, mydesigncrit.com; ⑭ Notable, notableapp.com; ⑮ Notism, notism.io; ⑯ Pixtick, pixtick.com; ⑰ ProofHQ, proofhq.com; ⑲ The Commentor, thecommentor.com.

The twelve tools in the lower part of Figure 2.3 work with the actual prototype (as visualized by the dotted line), but seven of those do not support the required basic drawing functionality, including ④ Critiquethesite, critiquethesite.com; ⑤ Diigo, diigo.com; ⑦ Floatnotes; ⑧ FreeRangeFeedback, freerangefeedback.com; ⑬ MyStickies, mystickies.com; ⑱ Scrible, scrible.com; ㉑ WebNotes, webnotes.net.

The opposite is the case for one tool called ⑥ Drawhere, drawhere.com, which supports only graphical input. Therefore textual commenting, if it can be considered so, is only limited to drawing letters with the freehand tool onto the screen.

From all the tools evaluated, the following four have been found to fulfil all five requirements:

- Appotate ②
- MarkUp ⑩
- MyBalsamiq ⑪
- Webklipper ⑳

Appotate (appotate.com) supports the design and implementation process of a website. It brings together the different stakeholders, allowing them to propose ideas on the current prototype by adding comments and suggestions.

MarkUp (markup.io) allows the user to draw and write on a website. It offers a variety of customizing options to choose from (e.g. different colours).

MyBalsamiq (mybalsamiq.com) is the online version of the Balsamiq mock-up tool which can be used to create clickable prototypes enhancing it with online collaboration functionality. There is a wide range of readymade sketched common interface elements to compose the prototype from. With myBalsamiq designers can create interactive mock-ups of the interface of their software or websites looking like sketches. This can be done by either composing existing interface elements provided by the tool or by importing custom sketches or pictures into the tool. Afterwards the different mock-ups can be made (partly) interactive, by linking them and thus responding to clicks by the user. MyBalsamiq has some basic feedback modalities, initially intended for designers to discuss alternatives and propose different solutions. But this can also be used by end-users in PD activities to specify their ideas. A possible issue there might be that expressing PD ideas is not supported ideally, as the interface and interaction options were not designed for this task.

Webklipper (webklipper.com) is an online application that enables the user to not only annotate uploaded documents but also websites and share the results.

2.4.4 Functional Requirement Evaluation

The four tools were tested for fulfilment of the functional requirements described in Section 2.4.2. It was assessed to what extend the requirements were met, by running each tool in each of the three major browsers to test for R1 and R2, by providing example PD ideas to test for R3 and R4 and by annotating an interactive prototype to test for R5. An overview of the results is shown in Table 2.2. It indicates the fulfilment of each requirement as either fulfilled extensively '+' or basically 'o'. The reasoning for these ratings is presented below.

Tool	**Requirements**				
	R1	**R2**	**R3**	**R4**	**R5**
Appotate	+	+	+	o	+
MarkUp	+	o	+	+	+
MyBalsamiq	+	+	+	+	o
Webklipper	+	+	+	o	+

Table 2.2: Results of functional requirements evaluation.

R1 has been fulfilled extensively by all tools. As they all run inside the browser, nothing needs to be installed.

MarkUp has been rated as only basically fulfilling R2, because it needs to be added to the bookmarks toolbar of the browser. But this works fine in all three major browsers tested.

R3 is fulfilled extensively by all tools, because they all offer the functionality to link PD ideas to a specific part of the website, although they all use a slightly different approach.

MarkUp offers freehand drawing functionality, which is the reason why it extensively fulfils R4, compared to the other three tools, which only offer (very) basic drawing functionality of predefined shapes. MyBalsamiq was rated as extensively fulfilling this requirement, as it does not provide freehand drawing but the option to add any shape to the prototype of the shapes that are also used to create the prototype in the first place.

Regarding R5 myBalsamiq works only with prototypes created with the Balsamiq mock-up (or the myBalsamiq) tool. But as the prototypes developed in the Go-Lab project were developed using myBalsamiq, this was no restriction. Consequently myBalsamiq was selected as off-the-shelf tool to compare with a paper-based PD approach (see Section 3.3).

To summarize the State of the Art section: Participatory Design is a very broad and diverse research area. The work presented in this thesis focusses on creating more user-friendly individual systems by providing software support for end-users to inform the re-design in short sessions rather than long collaborations with designers. The related work comprises paper- and tool-based PD approaches, each with their own advantages and disadvantages. To compare the two different approaches, my-Balsamiq has been selected, based on an evaluation of available software to possibly support PD.

Chapter 3

Comparison of Paper-Based Approach with Existing Software Tool

This chapter presents the comparison of the paper-based Layered Elaboration approach with the myBalsamiq tool selected in the previous chapter. To support a qualitative comparison, besides a quantitative comparison of the number of ideas elicited with each approach, the rating scheme CAt+ (Categories plus Attributes) is developed, described in detail, and evaluated. The results of the empirical study performed to compare Layered Elaboration and myBalsamiq show that paper results in a lot more ideas gathered whereas the qualitative differences do not show a consistent trend or meaningful patterns.

3.1 Comparison of paper and tool

To be able to replace paper with a tool (addressing the initial research question RQ1), it first needs to be ensured that paper and tool usage result in the same number and kind of ideas. Otherwise using one or the other would alter the results of the PD activities.

There have been some studies (e.g. Hundhausen et al., 2008; Segura et al., 2012) comparing paper- and tool-based approaches to collect user input empirically in contexts other than but somehow related to PD, including software inspection (MacDonald and Miller, 1998) and multimedia design (Bailey and Konstan, 2003). These comparison studies focused mostly on quantitative results (e.g., the number of defects found during inspection; task completion time) and on subjective opinions of the participants about the use of a tool versus the use of paper for performing specific tasks. However, they hardly compared the quality of the results (except Bailey and Konstan, 2003, where the richness of the user-generated screens was evaluated to some extent).

But to the best of my knowledge, no study has been conducted to compare systematically a paper-based and tool-based approach to determine to what extent software tools can be used to support or even replace paper-based PD activities. This gap was the motivation to conduct such a study.

The first and foremost step of the planned study was to identify an appropriate paper-based approach and software tool to use. As the target group involves children (students) as well as adults (teachers), the decision was made to choose a paper-based approach appropriate for children, and the Layered Elaboration approach (Walsh et al., 2010) was proved to serve this purpose (see Section 3.3.1 for details). It was shown that this approach could somehow be evolved into an online tool (DisCo by Walsh et al. (2012), see Section 2.3). However, the results gathered with DisCo were not yet compared to those gathered with the paper-based Layered Elaboration approach.

Based on the tool evaluation presented in Section 2.4 myBalsamiq was chosen for comparing the effectiveness of the paper-based and tool-based PD approach in terms of ideas elicited.

The drawbacks of exclusively quantitative (or qualitative) approaches are increasingly recognized in the field of HCI (e.g. Law et al., 2014). For instance, the total number of user comments cannot tell whether the comments address the content or the user interface design and interaction concept of the system evaluated. To quantify qualities and thus making them comparable, coding schemes can be used to rate the comments regarding different qualities. One of the challenges of qualitative analysis is the identification or development of a viable coding scheme. As no coding scheme could be found to rate PD ideas on webapps, the following sections describe the development and refinement of a new coding scheme, called CAt+. This is needed to rate the ideas gathered using Layered Elaboration and myBalsamiq and consequently compare the results of the two different approaches.

3.2 CAt+

3.2.1 State of the Art: Coding Schemes

The following sections are taken from a paper published at INTERACT2015 (`http://www.interact2015.org`, last accessed 12/10/2016) (Heintz et al., 2015b).

Various coding schemes have been developed for a variety of topics, ranging from user comments on machine learning (Stumpf et al., 2007), over student comments on the teaching performance of professors (Kindred and Mohammed, 2005) to YouTube comments (Madden et al., 2013). Könings et al. (2010) applied PD in a school setting to plan and improve lessons together with students and developed a coding scheme to code the spoken comments made during discussions. As those schemes are tailored to specific topics and use cases, they are not general enough to be applied directly to coding comments from different domains, e.g. PD ideas on webapp design.

3.2.2 Creation of CAt+

Each user idea gathered in the empirical study comparing Layered Elaboration and myBalsamiq (see Section 3.3) was coded by two HCI researchers (i.e. fully crossed design; Hallgren, 2012) with about two and six years of experience in usability research. Content analysis was applied to generate categories while coding (Krippendorff, 2004). Therefore the categories were generated based on the content of the ideas, while they were coded, initially starting with a blank slate. Both researchers coded the ideas in chunks of about twenty, introducing new category identifiers and definitions where necessary. To make sure that all ideas were coded appropriately, earlier ideas were revisited whenever a new category was introduced. The results were then compared and in case of discrepancy the researchers discussed till a consensus was reached.

Classifying ideas based on their content can help to understand the information contained, but not necessarily enables the comparison of the two approaches used to create the comments. For instance, ideas on design are not necessarily 'better' than ideas on functionality; on this basis no conclusion can be drawn which approach is 'better'. Hence, a broader set of codes with meaningful ratings on the measurable quality of ideas such as specificity (the more specific, the better) was needed. Based on the literature on user defect classification systems (e.g. Vilbergsdottir et al.,

2014) and downstream utility (supporting developers in addressing user feedback, e.g. Hornbaek and Stage, 2006), the following three major attributes were identified:

- **Impact** - the extent to which the mock-up will be changed by addressing the idea expressed in the comment;
- **Specificity (regarding target, reasoning, and solution)** - the detailedness and thoroughness of the idea in terms of explicitly stating the target, reasoning, and possible solution to make an improvement;
- **Uniqueness** - the distinctiveness of the expressed idea.

The initial coding scheme containing categorisation was thus completed by including those three Attributes with values and definitions. Accordingly, the coding scheme was named CAt+: **C**ategories plus **At**tributes, with the **plus** of Attributes compared to other coding schemes that only apply Category (or content) based coding.

3.2.3 Categories

Tables 3.1, 3.2, and 3.3 (the tables have been split to fit on a single page) show the initial categories of CAt+. Each rating is composed of a sub-category together with the main category, e.g. 'Content-Add' (i.e. no idea is just rated as 'Content').

3.2.4 Attributes

Attribute: Impact

The impact rating specifies how much of the user interface would change if this idea is addressed. Its possible values from 0 to 4 are defined as follows:

- **No changes suggested (code: '0').** There is nothing which could have an impact on the mock-up if implemented (e.g. "good idea").
- **Change affecting one element (code: '1').** As implementing the suggested changes would only influence a small part or a single element, the impact of this idea on the whole prototype is small (e.g. "the next [button] needs to be in a different colour to make it clearer.").

Category	**Description**	**Example**
Content	Comments on the learning material.	
Add	Request for more	"Put some text on this page."
Amount of text	Comment on the number of words used	"Shorten the text."
Change	Request for alteration (i.e. what is written), including typos	"However the questions don't seem to link in with the overall subject of the page."
Language	Comment on the wording (i.e. how it is written)	"Use more child friendly language."
Missing description	Request for explanation	"Unclear what to do with these questions."
Positive statement	Supportive comment	"I like the idea of having a video"; "Helpful instructions"
Remove	Request for deletion	"You don't need it to say in three minutes"
Terminology	Comment on only single words and their definition	"What does Buoyancy mean?"

Table 3.1: Name, description, and example for the Content category and its subcategories of the initial CAt+ rating scheme.

Category	Description	Example
Design	Comments on the visual appeal of the mock-up.	
Add	Request for new graphical elements or sound	"add crashing sounds"
Colourful	Comment on the colour or aesthetics	"The tabs could be more colourful."
Friendly layout	Comment on the suitability for children	"Kid friendly layout"
Negative statement	Criticism	"I do not like how the lines meet, it looks messy."
Not specific	General comment without detailed information	"Nothing to grab my attention."
Positive statement	Supportive comment	"Good use of colour to engage the students."
Screen layout	Comment on the positioning, order and size of elements on the page	"Make this bigger to fill the page?"
Terminology	Comment on only single words	e.g. button labels: "Complicated word ['Conceptualization' tab]"
Text layout	Comment on the format (size, style, colour, etc.) of the writing	"Better font, bigger font."
Remove	Request for deletion	"Don't need this [Page 2/2]."
Visual	Comment on the form/shape/sharpness of elements other than text and images	"... a different symbol could be used."

Table 3.2: Name, description, and example for the Design category and its subcategories of the initial CAt+ rating scheme.

Category	Description	Example
Functionality	Comments on interactivity of the mock-up.	
Add	Request for more things to do (e.g. buttons or apps)	"Maybe include zoom in and out buttons ... ."
Missing description	Request for explanation	"What is this for? [Resize element on video]"
Positive statement	Supportive comment	"Good system dragging and getting the answer."
Remove	Request for deletion	"I don't think you will need the calculator."
Picture	Comments on the pictures in the mock-up.	
Layout	Comment on the positioning, order and size of pictures on the page	"... have bigger pictures so you can see better"
Missing description	Request for explanation (including all 'picture unclear' comments)	"Try to describe the photos more so we know what they are."
Unknown	Comments of which the coders could not make sense.	
Not understandable	Comments of which the coder could not make sense	"Isn't the video"
Unreadable	Comments which or important parts of which could not be deciphered.	.
Irrelevant	Comments not related to the mock-up itself	less scribleing (sic!) [comment to scratched out idea of another participant]

Table 3.3: Name, description, and example for the Functionality, Picture, Unknown, and Irrelevant categories and their sub-categories of the initial CAt+ rating scheme.

- **Change affecting several elements (code: '2').** As multiple parts of a mock-up page would change if this idea is addressed, the impact of this comment is medium (e.g. "Do titles for pictures and stuff").
- **Change affecting the page on a level larger than element (code: '3'),** e.g. by adding/removing an element to/from the page, which would cause a change of the layout of the other elements as well. As significant parts or even the whole mock-up page would change if this suggestion is implemented, the idea is rated as having a high impact (e.g. "Add some thing (sic!) here [white space on the right]."). If the idea does not specify an element, it is assumed that the whole webpage is the target and thus affected (e.g. "More colour").
- **Change affecting several pages (code: '4').** As implementing the changes suggested in this idea would change various parts of the whole prototype, its impact is rated as very high (e.g. "log out option").

Attribute: Specificity

The specificity of an idea indicates how detailed it is described. This influences how easily and fast the designer can assess and address the idea. If the target (e.g. an interface object) is specified, the designer is able to identify which part of the mock-up should be changed. If the reasoning for an idea is given, the designer may find a solution, even if none has been specified by the end-user. If a solution is specified by the end-user the designer can decide to implement it or take it as further guidance in finding a feasible solution. If such information is not specified, it might still be possible and reasonable for the designers to make an educated guess. But if the informativeness of an expressed idea is too low, even guessing might not be possible.

Accordingly, a Specificity rating consists of six sub-ratings based on the three aspects of the idea discussed: Target, Reasoning, and Solution, and if they are stated or guessable. The aspects 'Target/Reasoning/Solution stated' can have the value 0, 0.5 or 1. If the respective information is given explicitly in the idea, the rating is 1. If it is somewhat clear what the participant means, the rating is 0.5. If no information is given, the rating is 0. The aspects 'Guessability of Target/Reasoning/Solution' have been introduced to rate if this information can be guessed (1) or not (0). If it is not necessary to guess this information (most of the time because it has been clearly stated), this aspect is rated as 1 (i.e. 'it is guessable').

Attribute: Uniqueness

Each idea is either coded as 1 if it has not been mentioned before or as 0 if it is a duplicate. By adding up the coding over all ideas, the number of distinct ideas can be known.

3.2.5 Further Development of the Initial CAt+ Coding Scheme

After developing the initial CAt+ coding scheme based on PD ideas gathered with a paper-based and a generic-tool-(myBalsamiq)-based approach on mock-ups it was later used to rate PD ideas gathered on more advanced software artefacts, namely prototypes. It is expected that the change of artefacts might introduce new Categories based on the different nature of mock-ups and prototypes. However, changing the digital artefacts evaluated from mock-ups to interactive prototypes aimed at further supporting the generalizability of the CAt+ coding scheme for different purposes in the area of PD and for different maturity of artefacts used for the PD activities.

3.2.6 Adaptation of CAt+

When coding the ideas gathered during the evaluation presented in Section 6.1, CAt+ was adapted accordingly. The differences and changes comparing the initial CAt+ coding scheme with the adapted version are explained in the following sections.

Categories

Results of coding the datasets of the study comparing PDotCapturer with paper support the generalizability of the CAt+ Categories and Sub-Categories compiled and described before, as they were shown to be applicable to datasets other than the ones coded there. As already discussed, some new combinations of main and sub-categories will probably be required when CAt+ is applied in other contexts. Several such combinations for the dataset comparing PDotCapturer and paper were indeed needed.

Combining all existing categories and sub-categories freely to make up for combinations that did not occur in the initial dataset of the previous study used to create the CAt+ coding scheme was mostly sufficient for the rating of the dataset gathered when comparing PDotCapturer and paper.

Only one new subcategory had to be introduced: Fix. This was used in the combined categorization of Functionality-Fix to code ideas that requested to make something work. When thinking about it, it makes sense that this sub-category had to be introduced and was not in the datasets of the previous study performed with mock-ups. As the latter analysed ideas gathered through an evaluation of an only-partly interactive mock-up (navigation was working, but none of the other interactive elements on the page), no bugs or not-working elements could have been identified by the end-users. As the participatory design sessions used to compare PDotCapturer and paper evaluated fully interactive prototypes, issues regarding screen elements not working would be introduced, causing the need for an according new sub-category in the rating scheme.

The following seven combinations of main and sub-category not occurring in the dataset of the previous study have been introduced for the agreed category ratings of the dataset comparing paper-based and PDot-based PD ideas:

- **Content-NegativeStatement** (e.g. "no motivating start"; "representation of the electron movement is bad")
- **Functionality-Change** (e.g. "— Typing the hypothesis not load could be better.")
- **Functionality-NegativeStatement** (e.g. "[Besides Concept Mapper scaffold] Difficult to use ...")
- **Picture-Change** (e.g. "[Arrow from the two balls on the scales] use different colours so students don't assume is the same material")
- **Picture-PositiveStatement** (e.g. "... really nice pictures! :)")
- **Unknown-NegativeStatement** (e.g. "Too difficult! [underlined twice] For our students")
- **Unknown-PositiveStatement** (e.g. when there was a smiley somewhere on the page, without any clear target or explanation)

The introduction of the combination Functionality-Change can again be explained like Functionality-Fix. By having more actual prototypic functionality in the software artefact evaluated the participants were now able to voice their impression on where changes would be necessary. The sub-categories 'Positive statement' and 'Negative statement' make sense for all main categories, it only depends on the dataset if they actually occur or not.

Although six combinations of main and subcategory present in the initial CAt+ coding scheme did not occur while coding the current datasets, they were not removed. Their non-usage can be explained by the nature and fidelity of the artefacts evaluated. Thus to be applicable for coding ideas expressed on prototypes of different or increasing fidelity, CAt+ was augmented by adding new sub-categories without removing the existing ones.

- **Content-Terminology/Design-Terminology:** With Content-Terminology being one of the categories having a high difference between paper and tool in the previous study it is a little bit surprising that it was not needed for this dataset of PD ideas at all. The sub-category 'Terminology' not being used might be explained with the maturity of the prototype (now properly describing terms that might be hard to understand) and using more student-friendly terms in the prototype compared to the mock-ups evaluated in the previous study.

- **Design-FriendlyLayout:** This combination was only used for student comments in the previous study, not for teachers. As the participants in the current study were all teachers, there seems to be a trend, which should be investigated in future studies.

- **Design-NotSpecific:** This again might be explainable with the maturity of the prototype compared to mock-ups. Having a more finalized design to comment on might have enabled the participants to be more specific when expressing their ideas.

- **Picture-MissingDescription:** As there were only two pictures (side-by-side) included in the online lessons as part of the learning content, which were well described, it can be explained, why this combination was not used to code the ideas this time.

Only 4 of the 339 ideas (1.18%) in this dataset were initially coded as 'Picture'. As pictures are either used as design elements or learning content, it is argued that the CAt+ coding scheme can be simplified and thus improved by removing the main category Picture. Those ideas can be put either in the Content or Design category. Accordingly, the 4 Picture ideas were re-coded.

The word 'Unknown' has the implication that something is yet to be discovered, whereas the meaning of this Category is more in the direction of 'Unintelligible' or 'Obscure'. It is therefore proposed to rename this main category to 'Incomprehensible'. This use of a clearer term improves the usability of the CAt+ coding scheme, especially for novice raters.

The set of main categories of the CAt+ coding scheme thus slightly changed by the coding of these evaluation results from the initial set of main categories presented before. The category Picture was removed and the term Unknown changed to Incomprehensible, resulting in the following five main categories: Content - Design - Functionality - Incomprehensible - Irrelevant. Besides Irrelevant, where it would not be meaningful or sometimes even impossible to assign a sub-category, each category rating still consists of a combination of a main and a sub-category, e.g. Content-Add, Design-Layout, Functionality-Remove.

The set of sub-categories of CAt+ used in coding this dataset is the following: Add - Amount of text - Change - Colourful - Fix - Language - Layout - Missing description - Negative statement - Positive statement - Remove - Screen layout - Text layout - Unreadable - Visual.

Attributes

The coding of the attributes Impact (0 = No changes suggested; 1 = Change affecting one element; 2 = Change affecting several elements; 3 = Change affecting the page on a level larger than element; 4 = Change affecting several pages) and Uniqueness (0 = Duplicate; 1 = Unique comment) was applied for this dataset as described for the initial CAt+ coding scheme before.

The initial Specificity rating on the other hand was complicated and some values of 'stated' had an influence on the possible values of 'guessable', e.g. stated 1 would also always be guessability 1. The specificity rating has therefore been simplified and thus improved, by combining the two sub-ratings for each specificity dimension (target, reasoning, and solution). The possible specificity values are now:

- **0**, for 'not even guessable'
- **1**, for 'not stated but guessable'
- **2**, for 'somewhat stated'
- **3**, for 'clearly stated'

This was already done for the analysis of specificity in the previous study, thus it was only reasonable to do it for the coding straight away.

3.2.7 Evaluation of CAt+

To determine which ideas are currently (not) of interest, a coding scheme can be used to categorize and cluster qualitative PD input. Initially the ideas were exported to spreadsheets and coded there, no dedicated software tool supporting the coding was provided. Exporting the ideas makes it harder to do the rating, as the context is lost. PDotAnalyser addresses this issue by integrating CAt+ rating functionality into its GUI. In two of the three evaluation studies conducted to evaluate PDotAnalyser (see Section 5.3), feedback was collected on the integration of CAt+ into PDotAnalyser and also on the coding scheme itself to further improve it (NB: the integration of this functionality was not yet done at the time of the first study).

CAt+ was not compared to existing rating schemes, as this would not have been meaningful. If a comparable rating scheme existed, that could have been used and extended, instead of creating a new one from scratch. Instead, CAt+ was evaluated by applying the rating scheme to more datasets similar and comparable to the one it originated from, to check its applicability and completeness for the rating of PD ideas on webapps.

To evaluate the usefulness of the CAt+ coding scheme for designers when working with and analysing PD ideas the participants in Study 3 evaluating PDotAnalyser (see Section 5.3) were asked questions in the form of "Having the categories/impact/.../uniqueness (CAt+ element) rating would be helpful when addressing the ideas." (see Appendix A.1, question 13). The results are presented in Table 3.4.

When looking at the helpfulness ratings of the CAt+ elements shown in Table 3.4 it can be seen that all elements are rated as being helpful. The categories are considered most helpful. This shows that using CAt+ to code the PD ideas is helpful for the designers and could therefore be applied as support, even when the ideas are captured with tools other than PDotCapturer or methods not using software tools at all.

From the think aloud and questionnaire (Appendix A.1) results two improvement suggestions for the CAt+ coding scheme can be derived. CAt+ categories should be further clustered and the terms streamlined, as they are currently a mixture of

CAt+ element	**Study 3 (n = 5)**
Categories	4.6 (0.89)
Impact	4.0 (0.71)
Specificity - Target	4.0 (1.22)
Specificity - Problem/Reasoning	4.2 (0.45)
Specificity - Solution	3.6 (1.14)
Uniqueness	4.0 (1.73)

Table 3.4: Mean ratings of the helpfulness of CAt+ categories and attributes when addressing ideas on a 5-point Likert scale from 1: Strongly Disagree to 5: Strongly agree, with standard deviations in brackets.

issue-related terms (e.g. terminology) and action-related terms (e.g. add, change, remove). This could be addressed by rating 'issue' and '(proposed) solution' separately (Figure 5.10). Regarding PD idea analysis, the rating on 'solutions' would be of more interest as it classifies re-design ideas proposed by end-users. Nevertheless the issue rating on 'issues' would also be beneficial, for example to identify sets of problems without solution ideas from end-users, to specifically address them in subsequent PD events.

3.3 Paper and General Purpose Tool

To address RQ3 and RQ5 (Section 1.2) the off-the-shelf online tool myBalsamiq (see Section 2.4.3) was compared with a paper-based PD approach regarding number and quality of ideas gathered. Answering these research questions is necessary to justify applying the tool over using paper in order to benefit from the described advantages of software support. But it also goes the other way around: myBalsamiq has shortcomings (e.g. the requirement for an Internet connection) and the question then becomes if paper is an appropriate way to be used as a back-up, e.g. in case of very limited Internet access. For the systematic comparisons of the cross-media results (paper vs. digital) it is crucial to have a coding scheme for analysing and comparing PD ideas. The CAt+ coding scheme described above was used.

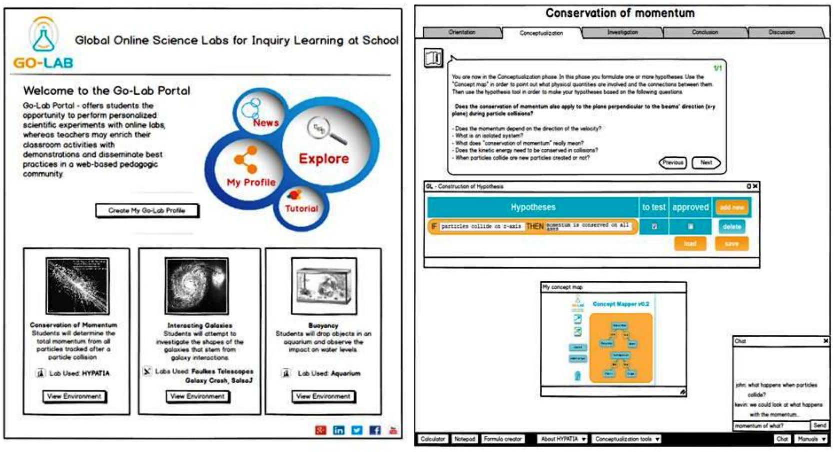


Figure 3.1: Mock-ups of the portal: Homepage (left) and a webpage with learning content and tools provided by the teacher for students (right).

3.3.1 Design of Empirical Study

PD Study with Interactive Mock-ups

With the Balsamiq software, three mock-ups were created that shared the same basic structure but differed in the complexity of the learning content to address students of different academic levels (Figure 3.1). A paper-based and tool-based approached were then applied to collect PD ideas on these mock-ups:

Paper-based: Layered Elaboration is a more recent paper-based prototyping technique (Walsh et al., 2010), which is simple to apply, and has the feature of keeping the initial prototype and comments from an iterative process intact. By overlaying different acetate sheets for the same mock-up printout, researchers can identify, for instance, which features have most frequently been commented on. The process starts by providing each individual or a small group of participants with a usage scenario, a set of ordered numbered printouts of the mock-up, a clipboard, and acetate sheets. Participants are asked to read through the scenario, put one acetate sheet on a printout, one after the other following the given order of printouts, and then provide PD ideas by annotating the acetate sheets with text and sketches, while working through the scenario on their own pace within a 45- or 60-minutes timeslot.

Tool-based: Participants are provided with computer access and work individually or in a small team of two or three people. They are introduced to the mock-up and then shown how to modify the mock-ups using different options provided in the myBalsamiq editor (Figure 3.2). As with the paper-based approach participants are given a usage scenario specific to the mock-up to follow and are asked to provide their ideas while following the scenario. As the tool includes the elements added by the end-user in a transparent layer on top of the prototype, leaving the original prototype intact, it is comparable to the layered elaboration approach. Among the options to express PD ideas there are yellow 'virtual sticky notes', which are added to the prototype by dragging them from the menu, to create a textual comment. Such notes are also commonly used in paper-based PD methods (Druin, 1999). As with the paper-based method the participants progress through the scenario on their own pace within a 45- or 60-minute timeslot.

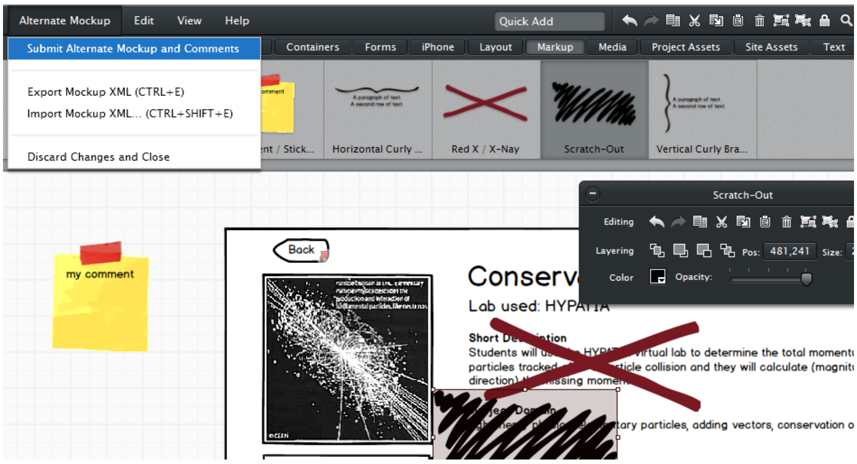


Figure 3.2: The main feedback functions of myBalsamiq (screenshot of myBalsamiq taken and included with the permission of Balsamiq Studios, LLC)

Participants and Procedure

The mock-ups were used in PD workshops in two schools in England (School 1 and School 2) and in an international teachers programme conducted at CERN in Switzerland (Teachers Programme) to gather ideas for the improvement of the prototypes and data for the comparison of the paper-based and tool-based approach. The reason for picking students and teachers as participants was that those were the two main target groups in the Go-Lab project and therefore the two main end-user groups in the project's PD activities. The workshops were started with two presentations (i.e. one briefly on the project and one on the PD process), followed by about a one-hour session of hands-on activities with the mock-ups.

School 1: The first PD workshop took place in a high school. Thirteen science students (mean age: 17 years old) were randomly divided into two groups with six using the Layered Elaboration approach (paper-based) and seven using myBalsamiq (tool-based). Two researchers were present to provide support, and each observed one of the two student groups. The Layered Elaboration technique is typically applied on a group basis and with young children, albeit being applicable to other age groups as well. Because of the low number of participants, they did not work in groups at this event but expressed their ideas individually, to maximize the number of datasets, with no intention to use it as an intervention variable to compare the results from the two settings. Although the Layered Elaboration approach was altered slightly in this case, this did not affect the comparability of the data collected with the two PD approaches, as the participants using the tool to specify their ideas also did not work in groups at this event.

School 2: The second PD workshop took place in an elementary (or primary) school. The procedure was the same as in School 1, except having the students working in groups instead of individually, thus following the Layered Elaboration approach as initially described. 28 students (mean age: 10 years old) were randomly assigned to one of the approaches; 13 (in three groups of three and two groups of two) used the paper-based approach and 15 (in five groups of three) used the tool-based approach to express their ideas. Because of the larger number of participants four researchers (two per approach) were present this time, observing and providing support.

An interesting observation while conducting the PD activities with primary school students was, that the part of the group which was randomly assigned to use PCs and the digital tool to record their ideas, were very excited about using the computer and the other part of the group was first disappointed because they had to work with pen and paper. However that seemed to change when they actually provided their ideas, because some students struggled to use the software tool and envied the students scribbling on the acetates.

Teachers Programme: The third PD workshop was organized as one of the activities in a programme for international high school teachers. It took place at CERN in Switzerland and involved 51 science teachers from 29 countries worldwide. The participants were split in groups of three. PD ideas from 8 groups working with the paper-based approach and 6 groups using the myBalsamiq tool were collected.

3.3.2 Data Analysis

All data were digitalised for further analysis. For the paper-based data, a set of the mock-up printouts and all the annotated acetates were scanned. With the use of the Gimp software the acetate part was removed from the scanned images, making the area where there was no drawing transparent again. Then all ideas from a single session were digitally layered onto the scanned mock-ups (Figure 3.3).

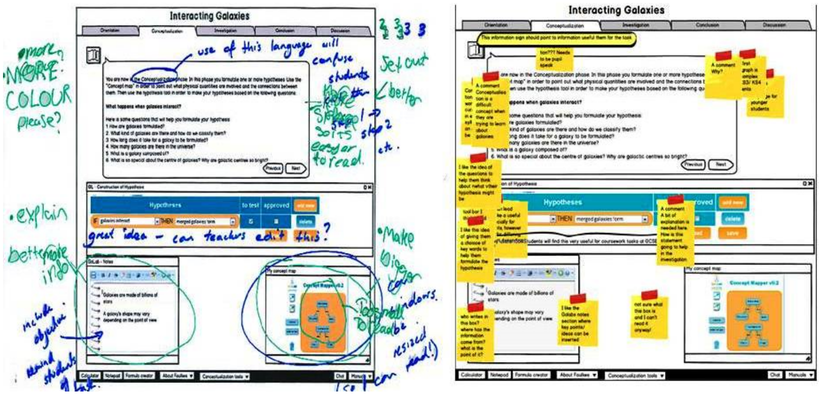


Figure 3.3: Superimposed ideas gathered using the paper- (left) and tool-based (right) approach.

For further data analysis, all comments, including textual and graphical, were recorded in Microsoft Excel sheets with two columns. The first one labelled 'source' contains 'participant and screen ID' allowing the retrieval of the original idea from the digital files. The second column records the textual user comment enhanced with researcher-generated details to make it easier to understand (e.g. description of the position or target) or a description of the drawing. Comments covering several ideas were split into individual rows during this step to prepare for further analysis.

3.3.3 Results

Results regarding RQ3 (Number of Ideas): Through the PD activities, 701 valid ideas (720 in total of which 19 were not related to the mock-up, e.g., a participant wrote a message to the researchers on the acetate "Sorry about my messy handwriting!") were given by participants using the paper-based approach. 190 valid ideas (191 in total of which one was unrelated to the mock-up) were gathered using the tool-based approach.

When comparing the mean number of valid ideas per individual or per group, the paper-based approach resulted in more than twice (School 2: 15.2 compared to 7.0) or even three times (School 1: 51.0 compared to 15.1; Teachers Programme: 27.0 compared to 8.3) as many ideas per individual/group as did the tool-based approach. Table 3.5 shows the results grouped by the main categories. Mean values are chosen in this case because of the relatively low number of participants/groups in each study, meaning that only a rough estimation of number of ideas gathered per participant or group with a bigger sample size can be provided. Table 3.6 shows the percentage (to account for and offset the vast difference in number of ideas gathered with paper and tool) distribution of ideas to categories for the three PD events, for the categories with a difference in the percentage values larger than 5% between the paper-based and tool-based approach in at least one event.

Main Category	**Comments per participant/group**					
	School 1		**School 2**		**Teachers Programme**	
	Paper	Tool	Paper	Tool	Paper	Tool
Content	8.5	3.0	3.7	3.6	13.1	5.8
Design	14.5	8.0	6.3	0.8	4.4	1.0
Functionality	21.0	2.7	2.3	1.4	5.9	1.2
Picture	0	0.3	0.5	0.4	1.9	0
Unknown	6.7	1.1	1.8	0.8	0.8	0.2
Irrelevant	0.3	0	0.6	0	1.1	0.2
Total	**51.0**	**15.1**	**15.2**	**7.0**	**27.0**	**8.3**

Table 3.5: Mean number of ideas per participant (School 1) or group of participants (School 2, Teachers Programme) for each of the main categories and in total.

Summing up, the empirical data of all the three PD workshops with students and teachers indicate that the paper-based approach was much more effective in terms of eliciting ideas.

When looking at the percentage results per category presented in Table 3.6 most of the differences between paper and tool in one event are contradicted by the results of another event. Thus no advantage for paper or tool can be identified for most of the categories. Two exceptions are 'Functionality - Missing description' and Unknown, where on a percentage basis (far) more ideas have been expressed when the paper-based method was used to gather the ideas as when the tool-based method was used.

For the missing description of functionality this might be explained by the paper being non-interactive and presenting the mock-up out of context (e.g. not on a computer screen, not in a browser). Thus users might have a harder time to identify the functionality of screen elements and therefore express in their ideas that a description would be needed.

The differences in 'Unknown' can be partly explained by unreadable ideas but more often by ideas expressed on the paper, where the target was unclear and therefore the problem could not be understood. It can be assumed that this happened less

Category	Percentage of comments					
	School 1		School 2		Teachers Programme	
	Paper	Tool	Paper	Tool	Paper	Tool
Content	16.8	19.8	25.3	51.4	50.7	71.4
Add	5.3	11.3	6.8	2.9	27.5	26.5
Language	1.3	0	5.8	14.3	1.9	4.1
Remove	0.3	0	1.6	0	1.0	6.1
Terminology	3.0	0	4.2	22.9	2.4	8.2
Design	28.6	52.8	43.2	11.4	16.4	12.2
Colourful	5.9	9.4	8.4	2.9	1.4	0
Text layout	1.3	11.3	12.1	0	4.3	2.0
Functionality	41.4	17.9	15.8	20.0	22.7	14.3
Add	23.7	9.4	3.2	17.1	7.7	14.3
Missing description	6.9	3.8	11.1	2.9	7.7	0
Picture	0	1.9	3.7	5.7	6.3	0
Unknown	13.2	7.5	12.1	11.4	2.9	2.0

Table 3.6: Distribution of ideas to categories (with a difference of more than 5% between paper and tool for at least one of the three result sets) in both Schools and the Teachers Programme (in %).

with the tool, because to express ideas verbally there, the end-users had to put a yellow sticky note, which they mostly put onto the screen element causing the issue, thus at least giving a hint regarding the target.

Results regarding RQ5 (Quality of Ideas): All ideas categorized either as 'irrelevant' or 'unknown' (either not understandable or unreadable) were removed from further analysis, as they do not contain useful information for the designers. Therefore 809 ideas (632 paper- and 177 tool-based) were further analysed with regard to the three Attributes of CAt+.

Impact. As presented in Table 3.7, most of the ideas proposed by the participants have an Impact of either 1 or 3 - affecting a single element or the whole page. This might imply that the participants tended to perceive the mock-up from a holistic perspective, although they zoomed in to explore specific elements (e.g. the next button) in detail. Regarding the comparison of paper and tool it can be seen that paper elicited (slightly) more ideas coded as 1 as compared to the tool results with this impact coding. Participants being more willing to specify ideas on small details with paper, might imply that expressing ideas with the tool needs slightly more effort, which was more likely spent on ideas with bigger impact.

Impact	**Percentage of further analysed comments**					
	School 1		**School 2**		**Teachers Programme**	
	Paper	Tool	Paper	Tool	Paper	Tool
0	15.9	10.2	9.6	0	7.5	6.3
1	28.4	22.5	40.7	32.3	32.3	27.1
2	1.5	1.0	7.2	25.8	9.5	14.6
3	35.2	58.2	31.1	32.3	32.8	33.3
4	18.9	8.2	11.4	9.7	17.9	18.8

Table 3.7: Distribution of Impact rating (in %).

Specificity. Table 3.8 presents the percentage of results were the corresponding sub-rating was clearly stated. When looking at the Target it can be seen that this is more specific in paper-based comments than tool-based. This may be explained by the fact that paper allows for a variety of ways to highlight a target (e.g. by drawing a circle around or an arrow pointing towards something, or by underlining text), where the tool used in the evaluation was restricted to a predefined set of feedback elements. Users mainly attached virtual sticky notes, which might explain the lower precision and could be approached by enhancing the functionality of the tool. On the other hand the use of sticky notes might explain the higher specificity of the Reasoning found for the tool-based comments, as the text field on the note might have invited the participant to further elaborate.

	Percentage of further analysed comments					
	School 1		**School 2**		**Teachers Programme**	
	Paper	Tool	Paper	Tool	Paper	Tool
Target	86.1	65.3	78.0	61.3	83.6	68.8
Reasoning	23.7	41.8	26.8	32.3	32.3	45.8
Solution	26.8	20.4	6.6	22.6	40.8	41.7

Table 3.8: Ideas coded as very specific for each sub-rating (in %).

Uniqueness. To determine how many duplicated ideas have been generated with the paper- and tool-based approach, the ideas have been rated based on their uniqueness. The results are shown in Table 3.9. When comparing the percentage of duplicates in the paper- and tool-based results, it can be noticed, that it is mostly higher for paper. If one assumes that there is a limited pool of possible ideas the end-users can come up with, the higher total number of ideas for paper also explains the higher percentage of duplicates as it becomes less likely to come up with a unique idea with an increasing number of ideas. The exception of this for School 1 might partly be explained by having the highest number of comments for the tool-based method throughout the three events, but is still unexpected.

	Percentage of further analysed comments					
	School 1		**School 2**		**Teachers Programme**	
	Paper	Tool	Paper	Tool	Paper	Tool
No. of comments	264	98	167	31	201	48
No. of **unique** comments	225	80	137	30	178	47
% of duplicates	14.8	18.4	17.9	3.3	11.4	2.1

Table 3.9: Results of the Uniqueness rating.

Inter-rater reliability. As two researchers were involved in coding the participants' ideas, weighted Cohen's kappa (Fleiss and Cohen, 1973) was calculated to determine the inter-rater reliability for the different coding criteria. For the Categorization the weight was determined based on the agreement about the main- and sub-category. If only the sub-category differed, a weight of 1 was used (as there was at least agreement about the main category of the idea), if the main category differed, a weight of 2 was applied. For Impact and Specificity, the weight has been determined by the difference between higher and lower value. For Uniqueness, the standard weight was used. For all ratings the value of Weighted Cohen's kappa was above 0.7, ranging from 0.72 (for Uniqueness) to 0.88 (for Specificity - Solution). Although the kappa rating magnitude guidelines in the literature are inconsistent (e.g. Altman, 1990; Fleiss et al., 2003), with all values being above 0.7 it can still confidently be assumed that these results are reasonable or even good.

Pearson's $\chi 2$ analysis of category rating. Some inferential statistics on the results were performed to check whether the observed differences are significant. As the data is categorical, Chi square tests (Maltby and Day, 2002) were used to verify the null hypothesis (H_0):

H_0: the number of ideas in each of the coding dimensions (categories, impact, specificity, uniqueness) is independent of the method used to elicit and capture them (paper or tool).

Table 3.10 shows the results. To get expected values larger than 5 (as required by the Chi square test, e.g. Maltby and Day, 2002) it was necessary to combine some of the results. For categories Picture, Unknown, and Irrelevant were combined to 'Other'. For Impact 0, 1, and 2 were combined to 'less than page level' and 3 and 4 to 'page level and above'. For Specificity the results were combined into three groups, very specific (two or more sub-ratings that are very specific), specific (one sub-rating that is very specific), and unspecific (no sub-rating that is very specific). For Uniqueness no combination was possible, therefore the result for School 2 is included in Table 3.10 although the requirement for Chi square was not met, as the expected value for 'not unique' in 'tool' was less than 5 (about 4.85). A Fisher's Exact test has therefore been performed for the latter, confirming the rejection of H_0 in this case.

As can be seen in Table 3.10 the results are only affirmative for Categories and Specificity ratings. For Categories, H_0 has to be rejected, meaning that the different methods influence the number of comments in different categories. However, the results of the descriptive statistics suggest that no consistent trends or meaningful patterns in terms of the Categorization can be observed. For Specificity, H_0 is not rejected, meaning that the specificity of an idea is independent from the method used for idea gathering. For Impact and Uniqueness the results of the Chi square tests are ambivalent.

To summarise this section on the comparison of paper-based approach with existing software tool: As already highlighted in the motivation section, it is essential to compare paper-based and tool-based approaches in PD. To be able to do this not only on a quantitative basis (i.e. number of ideas gathered), the rating scheme

	School 1	**School 2**	**Teachers Programme**
Categories	$\chi 2$ (3, n=412) =27.19, p<.001**	$\chi 2$ (3, n=233) =15.51, p<.05*	$\chi 2$ (3, n=266) =8.41, p<.05*
Impact	$\chi 2$ (1, n=362) =4.32, p<.05*	$\chi 2$ (1, n=198) =0.004, p>.05 (ns)	$\chi 2$ (1, n=249) =0.03, p>.05 (ns)
Specificity	$\chi 2$ (2, n=362) =3.52, p>.05 (ns)	$\chi 2$ (2, n=198) =1.96, p>.05 (ns)	$\chi 2$ (2, n=249) =4.44, p>.05 (ns)
Uniqueness	$\chi 2$ (1, n=362) =220.24, p>.05 (ns)	$\chi 2$ (1, n=198) =4.30, p<.05*	$\chi 2$ (1, n=248) =20.57, p<.001**

Table 3.10: $\chi 2$ values for independence of number of ideas per rating dimension of CAt+ on method (for all three events).

CAt+ has been created, to quantify different qualities of ideas (impact, specificity, and uniqueness) besides classifying them (using a combination of main and sub-category). The empirical studies comparing myBalsamiq with paper-based counterparts showed that paper resulted in quantitatively more data, the findings on the quality of the ideas gathered are inconclusive.

Chapter 4

PDotCapturer

This chapter presents the iterative development and evaluation of a new PD software tool: PDotCapturer. It is aimed at end-users to support them in their task to explore a prototype and provide re-design ideas. Three versions of the PDotCapturer tool, from the initial prototype to the final version 3, and the changes and improvements from one to the next version are described to show the influence of the evaluations (presented here as well) on the development process.

4.1 Overview

For the development of the PDotCapturer tool an iterative development approach with repeated evaluations of functionality, usability, and user experience was applied. To measure the usability and user experience questionnaires were used. They were filled in by the participants after interacting with PDotCapturer to express their ideas. The questions for this part of the questionnaire were derived from AttrakDiff2 (Hassenzahl and Monk, 2010). Additionally custom questions were added to the questionnaire regarding existing features (e.g. how they are rated and which to keep and which to remove) and missing features.

To define an initial set of requirements several workshops and unstructured interviews were conducted with designers of Go-Lab software, who would later benefit from the ideas gathered using the tool. Additionally the results of the evaluation of existing tools (see Section 2.4) have been taken into consideration regarding features and possible ways of implementation.

4.2 Design and Development

4.2.1 Requirements

This section is based on a CHI2014 poster (Heintz et al., 2014a). Based on the initial set of requirements used for the evaluation of existing tools, as described in Chapter 2.4, deploying myBalsamiq as existing tool in PD activities, as described in Chapter 3.3, and interviews conducted with designers a slightly reworked set of seven key functional requirements for a PD tool used for webapps was identified, which are categorized based on two target groups of the tool: User Requirement (UR) and Designer Requirement (DR). The software tool has to:

- **Interactivity (UR1):** Work with interactive prototypes as opposed to static images.
- **Annotation (UR2):** Enable end-users to describe ideas verbally in a textual comment as well as relate it to a specific interface element.
- **Creativity (UR3):** Support drawing (to provide graphical expression of ideas) and thus more advanced prototype editing (e.g. indicating adding or moving components).
- **Collaboration (UR4):** Allow end-users to provide annotations collaboratively.
- **Access (UR5):** Be easily accessible from anywhere with Internet connection, and work without installation.
- **Instructions (UR6):** Offer instructions to get the end-user started.
- **Aggregation (DR1):** Support aggregation of data (explicit and implicit) from different end-users.

UR6 has been added to support the end-users in their task in case they do not know the prototype targeted by the PD activity or in case designers are not available to support them. In this case the tool should display a set of instructions to the end-users to make them familiar with the functionality of the prototype and the task of providing their design ideas. The Designer Requirement DR1 has been added to not only support the gathering but also the analysis of PD design ideas.

To ensure good usability and high user experience an iterative development approach has been applied when developing PDotCapturer. Therefore an initial prototype was first developed and then further refined.

4.2.2 Initial Prototype

Functionality and Development

Initially a single tool (PDot) should address all requirements, but it was then split into two tools, one targeted at end-users (PDotCapturer) and one at designers (PDotAnalyser), to allow for tailoring of the GUI and features based on the needs of the different target groups.

From the end-user's perspective the general purpose of PDotCapturer (Figure 4.1) is to present the webapp which is the target of the PD activity (Ⓘⓘ in Figure 4.1) and simultaneously provide functionality to express PD ideas (Ⓘ in Figure 4.1). The early prototype of PDotCapturer presented in this section supports these two basic functionalities. The emotional response options only comprise of three options (like, neutral, dislike) rather than more common scales (such as a 5-point Likert scale) to speed up the decision and therefore feedback provision process. The user only has to decide on one of three base emotions regarding the current design and not on one or more nuances in between.

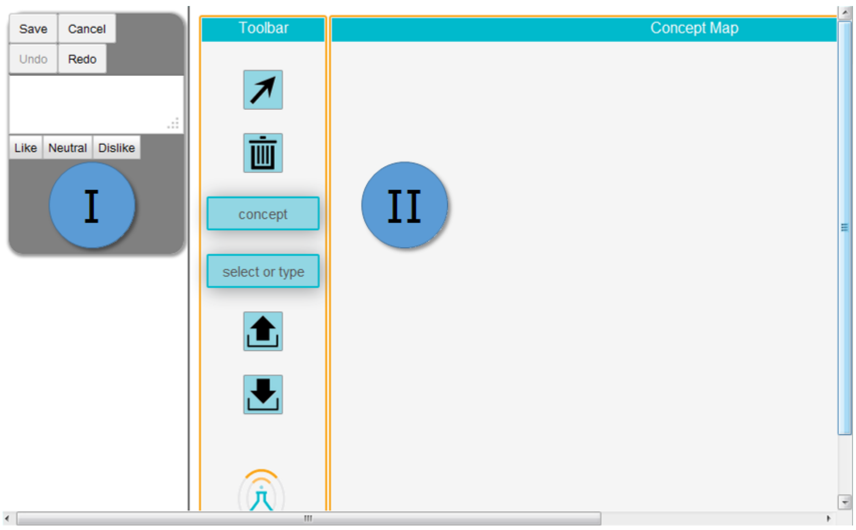


Figure 4.1: Overview of the initial PDotCapturer prototype: Functionality to express PD ideas (I) and webapp targeted by the PD activity (II).

From the designer's point of view, PDotAnalyser aggregates the end-user's ideas and thus allows for easy information retrieval (Figure 4.4). As the latter is not very well supported in the current tools (see DR1 results in Table 4.1), this was a very important functionality to be implemented in PDotAnalyser.

	UR						**DR**
	1	**2**	**3**	**4**	**5**	**6**	**1**
Appotate	+	+	o	+	+	-	-
MarkUp	+	+	+	o	+	-	-
MyBalsamiq	o	+	o	o	+	-	-
Webklipper	+	+	o	-	+	-	-

Table 4.1: Fit between existing tools and refined set of PD requirements. Legend: supported (+), partially supported (o), not supported (-).

Subsequently, it is shown how the initial versions of PDotCapturer and PDotAnalyser address the seven requirements for PD tools. PDotCapturer and PDotAnalyser both open the webapp in an iFrame, thus it is fully functional (UR1) while expressing or analysing ideas. Putting a sticky note on top of the webapp allows the end-user to specify a position to which the idea is proximal (UR2). PDotCapturer offers different means to express ideas and specify additional information (UR3, see Figure 4.2). For instance, by clicking and then moving the cursor, the end-user is able to draw on the webapp. The initial PDotCapturer prototype was kept as simple as possible and used as a starting point to collect feedback and input from participants to identify which additional functionality is needed. Collaborative gathering of ideas (UR4) was not supported in the initial prototype of PDotCapturer but this functionality was added later (see Figure 4.8). To be accessible (UR5) in all major browsers (Google Chrome, Mozilla Firefox and Internet Explorer), PDotCapturer is implemented using the Google Web Toolkit (GWT, `gwtproject.org`). This transforms Java code into HTML and JavaScript, which can be run in the major browsers without the need for a plugin. All idea data created by end-users are stored online in a MySQL database. To support the end-users in the task of providing PD ideas about an unknown webapp (UR6) a guiding scenario is presented in different instruction steps, explaining some interactions with this webapp to get the user started, e.g. in the form of a scenario to follow (Figure 4.3).

Figure 4.4 shows a proof of concept example for data aggregation in PDotAnalyser (DR1). It shows a like/dislike heatmap which highlights all positions where end-users have expressed their likes or dislikes. From this aggregated view, designers can infer which parts of the webapp should stay as they are and which parts need reworking.

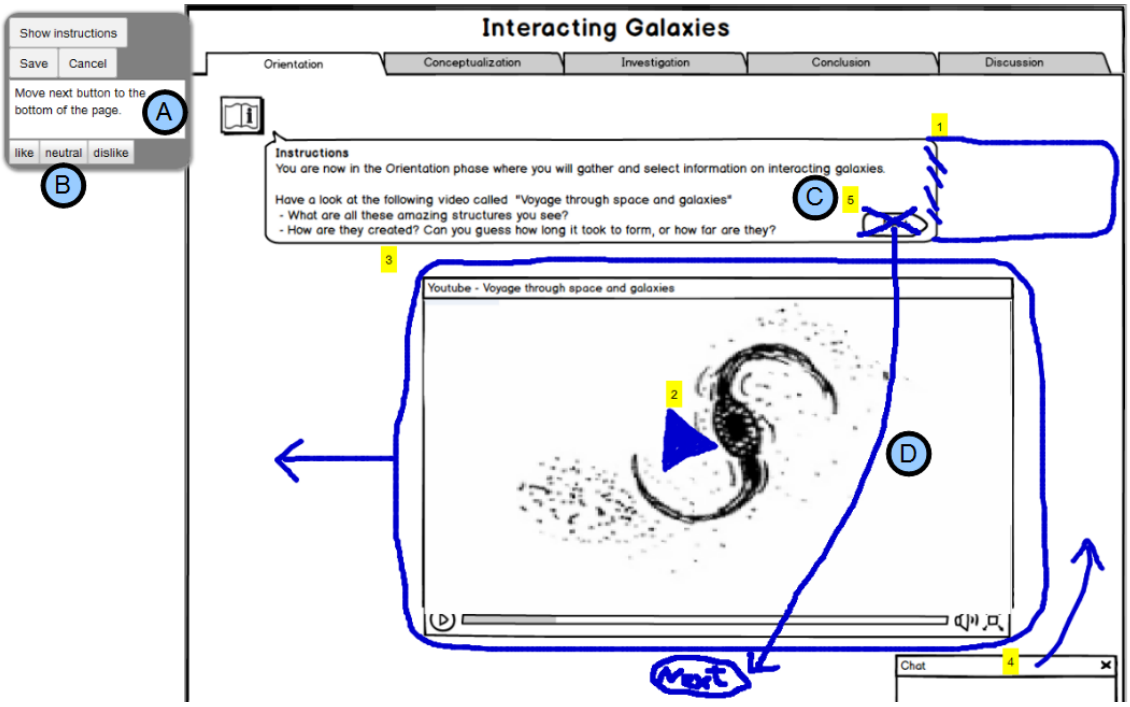


Figure 4.2: Functionality to express PD ideas in the initial PDotCapturer. Sticky notes Ⓒ and textual comments Ⓐ, freehand drawing Ⓓ and mood specification, which can be either like, neutral, or dislike Ⓑ, are supported.

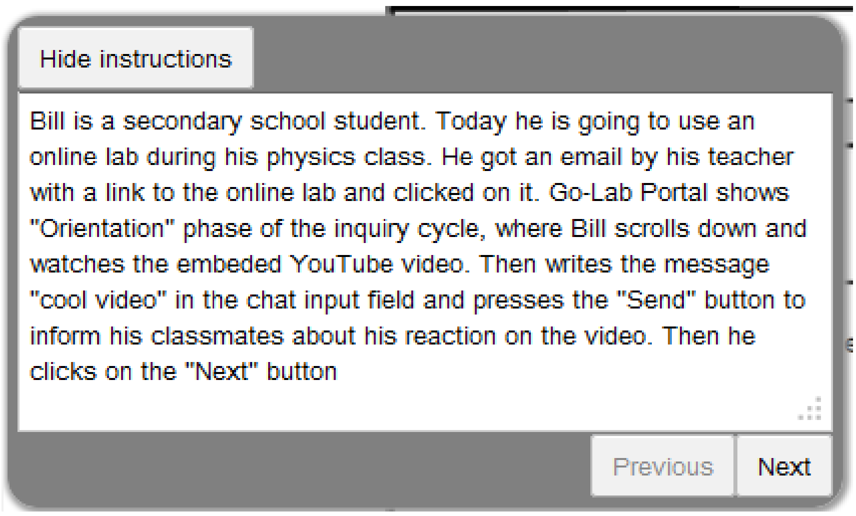


Figure 4.3: Scenario instruction step displayed in the instruction mode of the initial PDotCapturer.

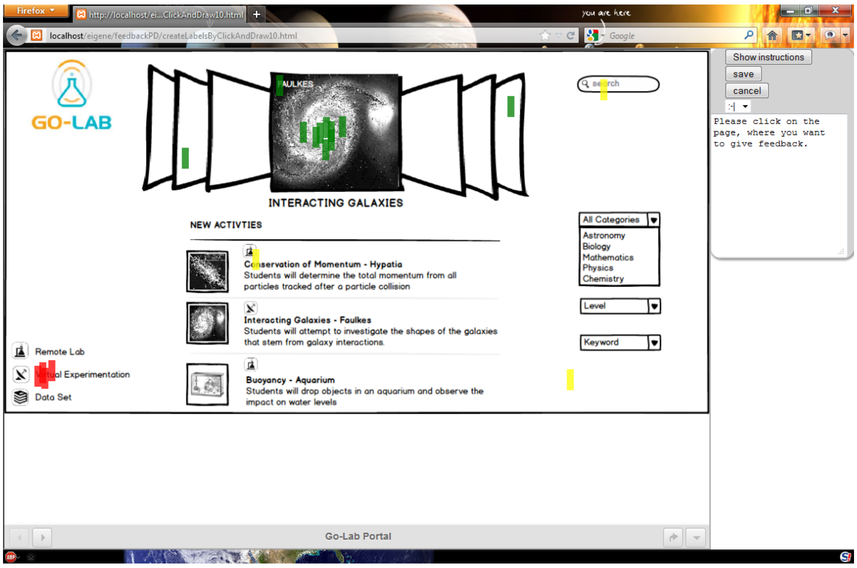


Figure 4.4: 'Like/Dislike Heatmap'. One can tell at a glance that the picture in the upper centre was liked (green highlight) and thus should stay. In contrast the key in the left lower corner was disliked (red highlight) and thus would need to be reworked. To access the details of the ideas expressed the designer could 'dive' into the ideas associated with a dislike emotion. Note that PDotAnalyser for the designer became more elaborate in following versions to address DR1.

To perform a PD activity using the PDotCapturer prototype, the designer needs to create a scenario for the interaction with the targeted webapp. This can be done through creating an evaluation event using the PDotCapturer administration interface (Figure 4.5). This automatically creates accounts for the end-users based on the number of participants specified by the designer and generates a URL to access this event in PDotCapturer.

The end-user then follows the link to PDotCapturer and logs in with the credentials created by the designer. She then reads through the first scenario step (presented in the instruction mode of PDotCapturer, Figure 4.3) and performs the interactions described with the webapp (which is presented in the interaction mode of PDot-Capturer after hiding the scenario step description). Switching between these two modes is currently achieved by pressing the 'Show/Hide instructions' button (see Figure 4.3 for the hide instructions button and Figure 4.2 for the show instructions button). As soon as she has an idea she would like to express, the user switches from the interaction mode to the feedback mode of PDotCapturer (currently by pressing a button labelled with 'Give feedback'). In the feedback mode the webapp is no longer responsive to interactions, but expressing PD ideas on the webapp is enabled. The user then clicks on the position the idea is related to. PDotCapturer displays a numbered rectangle at this position which is numbered in the order it

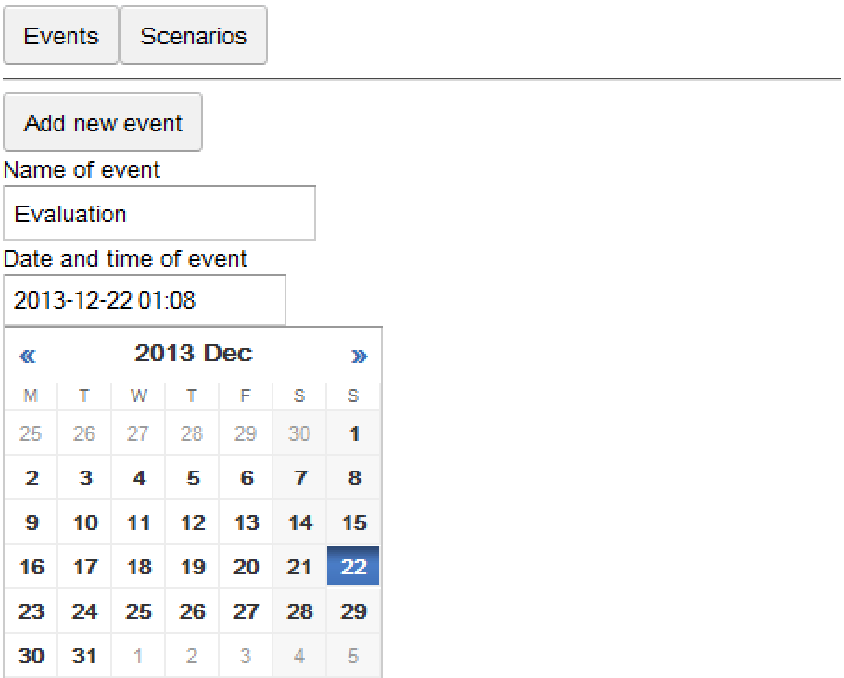


Figure 4.5: Screenshot of PDotCapturer administration interface.

was added. If she wants, the user can then draw on the webapp. To elaborate on the idea the user might also add a textual comment and specify her mood.

In the example displayed in Figure 4.2 the user clicks on the page just to the right of Ⓒ which creates the yellow rectangle numbered with a 5. Then he draws a cross over the next button, indicating it should be deleted there. As he prefers the button being on the bottom of the page he draws it there and adds an arrow to indicate the repositioning Ⓓ. He explains his drawing by adding the textual comment "Move next button to the bottom of the page." Ⓐ. He can then finally specify his mood regarding the initial design by clicking on one of the three toggle buttons Ⓑ.

After finishing with the current scenario step and with expressing ideas on the current page, the user continues the evaluation by switching to the next scenario step (by clicking on the next button in the bottom right hand corner of the scenario step description GUI, as displayed in Figure 4.3). The user then repeats the process for each of the scenario steps pre-specified by the designer to complete the PD activity.

After conducting the PD activity the designer accesses PDotAnalyser to have a look at the aggregated and detailed ideas of all end-users - a functionality not provided in the initial prototype but added later (see Chapter 5).

4.2.3 Version 2 of PDotCapturer

4.2.4 Functionality and Design

Based on the feedback gathered on the initial prototype (e.g. through observer notes and evaluations), the tool was reworked and extended. The details of the evaluation and findings are presented in Section 4.3.1 and summarised here as background information for the changes described:

- Improve usability, as some end-users struggled when using PDotCapturer
- Make it more clear which mode (interaction or idea expression) PDotCapturer is in
- Change mouse cursor to indicate actions that will be performed (sticky note marker or pen)
- Change the design of the emotional response recording toggle buttons (e.g. add smileys to make the design more appealing)

The result of the rework can be seen in the screenshot in Figure 4.6.

When comparing Figure 4.6 with Figure 4.2 the commonalities (e.g. Ⓐ, Ⓒ, Ⓓ in Figure 4.2 compared to ❷, ❶, ❹ and ❺ in Figure 4.6 respectively) and differences (e.g. emotional state is now represented by coloured smileys in addition to the textual expression [❻ in Figure 4.6 compared to Ⓑ in Figure 4.2]; mouse cursor changes to a blue pen when in drawing mode [❸ in Figure 4.6]) can be seen.

In addition to the added images for the emotional expression, the whole tool was redesigned aiming to improve usability and user experience and to have it look less like a prototype and more like a final product. The PDot instructions were displayed constantly on the top of the screen, which removed the need of pressing a button each time the user wanted to see them.

4.2.5 Version 3 of PDotCapturer

Functionality and Development

One major change for the final version of the PDotCapturer tool was a visual re-design, to address the feedback on earlier versions, that this could be further im-

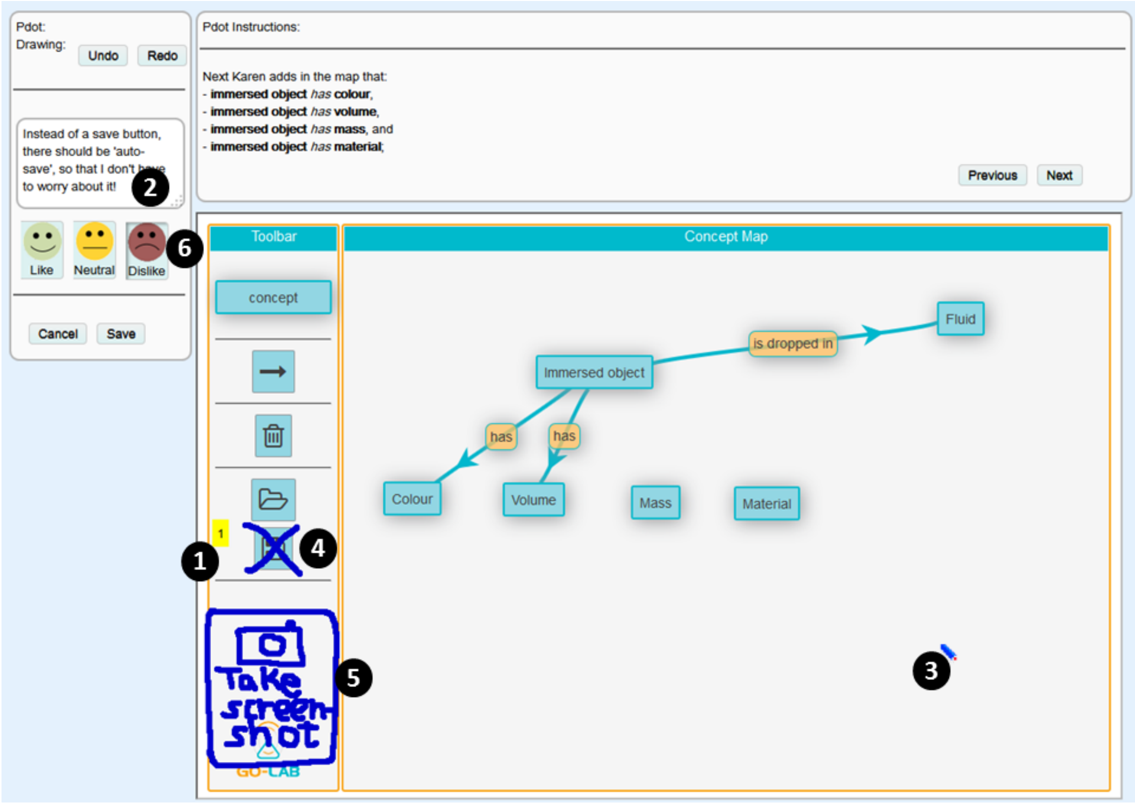


Figure 4.6: Screenshot of version 2 of PDotCapturer.

proved. The changes from the first 'proof-of-concept', over the version 2 to the final version 3 are shown in Figure 4.7.

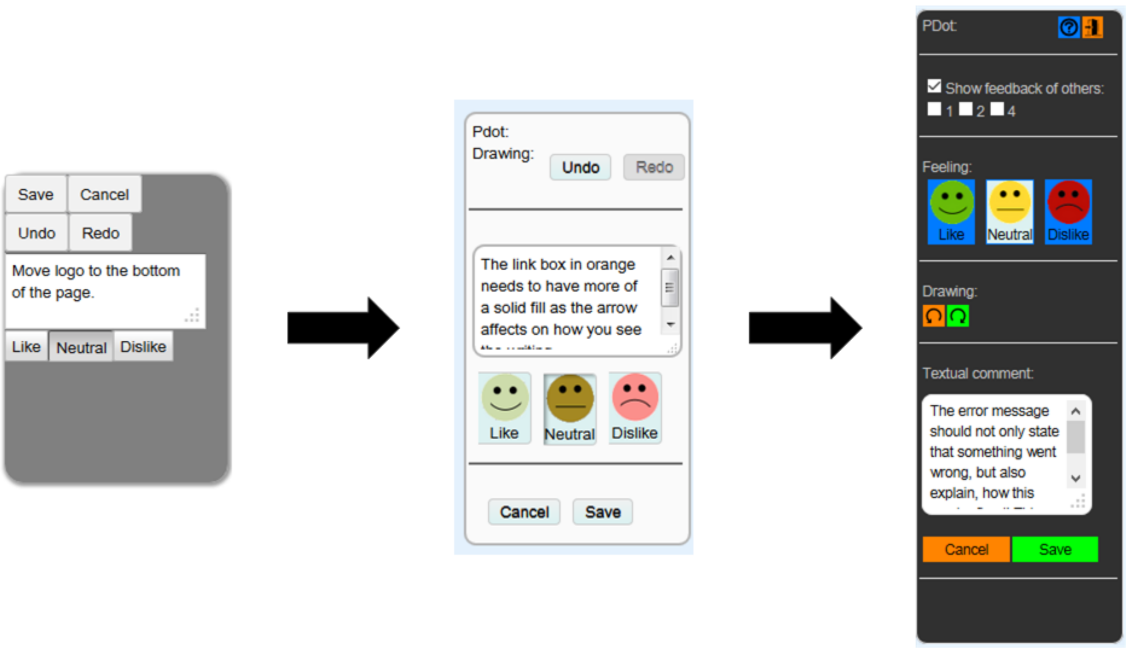


Figure 4.7: Evolution of the visual design of PDotCapturer from 'proof-of-concept' on the left to version 3 on the right.

But besides a visual reworking, many other comments from participants were taken into account when creating the third version of the tool (the final version in the scope of this thesis). The details and findings of the evaluation are presented in Section 4.3.2 and summarised here as background information for the performed changes described in this section:

- Further improve usability of PDotCapturer
- Make PDotCapturer more visually appealing (for students)
- Make specification of the emotional response more prominent
- Make it more clear which mode (interaction or idea expression) PDotCapturer is in

As can be seen one point of criticism was that specifying like or dislike of the current design was mandatory (to be able to create the heatmap and colour-code the markers in PDotAnalyser), but could be overlooked. To prevent this a pop-up with the three smileys is now displayed at the position of the marker when it is created. This way it is much more prominent and much quicker to provide this information.

It has been reported that it can be difficult to identify in which mode PDotCapturer currently is. This was addressed by visually highlighting the feedback mode through a coloured frame around the prototype. An alternative idea was to 'grey out' the prototype while it is not interactive, as it is done in other applications (e.g., the Snipping Tool in Microsoft Windows), but the frame visualisation was picked instead, because it does not alter the presentation of the prototype. These and other slight improvements of the user interface were implemented but are not all described here in detail.

The main feature added in the third version of PDotCapturer is enabling collaborative, PDot-supported PD sessions not using PC-sharing as the mean to collaborate. To do so the tool was extended by adding the functionality to display the ideas of the other end-users participating in the current PD event (Figure 4.8).

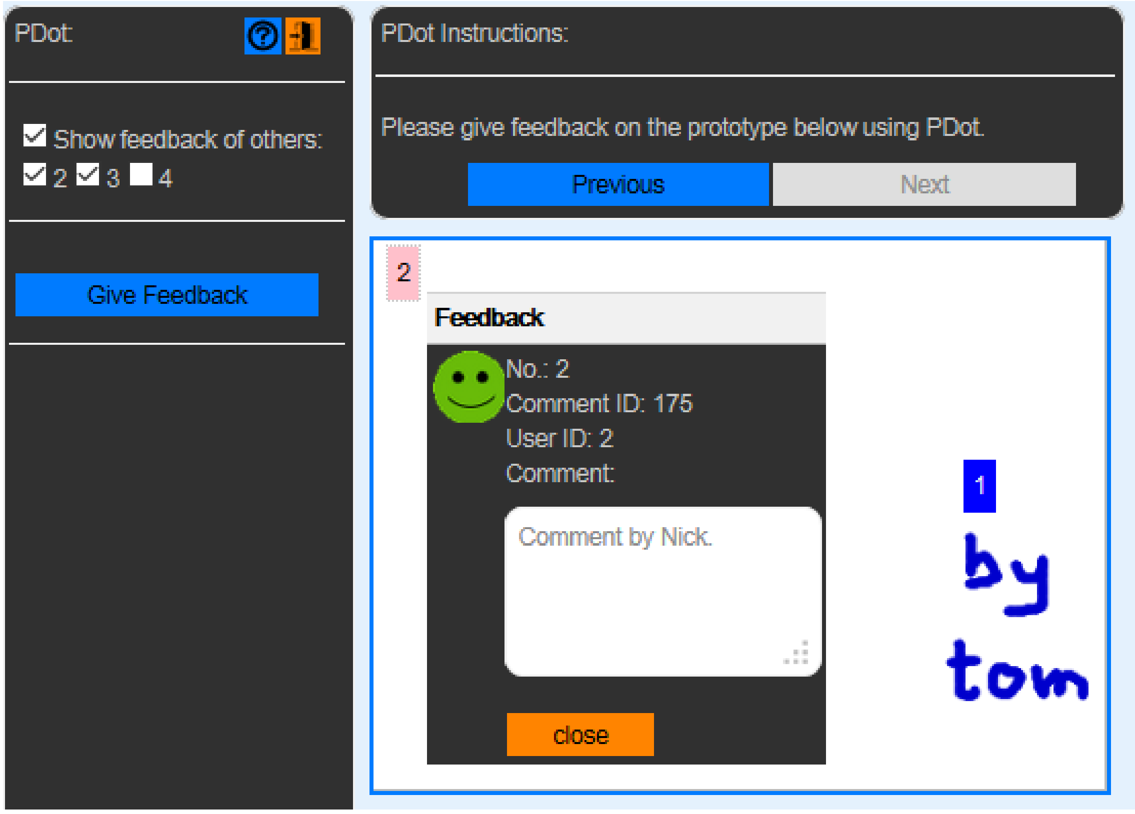


Figure 4.8: Added functionality 'Show feedback of others:' to display ideas of other end-users in PDotCapturer (by username, which is a number).

4.3 Evaluation

4.3.1 Evaluation of Initial Prototype

Goal and Participants

The goal of the preliminary evaluation was to find out how well the initial PDot-Capturer prototype performs with regard to the user requirements (except UR4 which was not implemented in the initial prototype) and usability. For evaluating the early prototype, which was later redesigned based on the feedback gathered and retested iteratively, six participants (2 male and 4 female PhD students in HCI or Computer Science) were involved. Each participant was asked to express PD ideas on a mock-up page guided by a scenario. At the same time they were asked to give feedback on PDotCapturer.

Procedure

Evaluation data were collected by observing the participants' interactions with the initial prototype of PDotCapturer and by conducting a questionnaire as well as a semi-structured interview (sections Questionnaire and Interview in Appendix A.2 respectively).

The questionnaire consisted of general questions about age, gender, field of study / research, adoption of new technology, and experience in requirements engineering. Participants were asked to rate how they liked PDotCapturer overall and also the following qualities: 'natural', 'easy to use', 'easy to learn' and 'flexible' (on a 5-point-scale from 1: 'not at all' to 5: 'very much'). Besides that, they were asked to specify the most critical feature they found currently missing.

The interview was structured by questions on the likelihood of using PDotCapturer to express PD ideas once it is available, and if the participant would recommend it to peers. The interview was concluded by inviting suggestions on how to improve the initial prototype of PDotCapturer.

Results

Observations showed that the participants could easily access the initial prototype of PDotCapturer using a browser (UR5) and successfully explore the interactive prototype of a webapp (UR1) guided by the scenario step (UR6). They were able to annotate (UR2) and draw on the webapp (UR3) but sometimes struggled with using PDotCapturer, implying that its usability needs to be improved.

A major usability problem identified is that some participants found it hard to identify which mode PDotCapturer was currently in (interacting with the prototype or expressing ideas). Insights how this could be visualized more clearly were obtained through the interviews: toggle buttons or tabs could be used to switch between the modes, or they could be displayed more clearly by having a separate box labelled with the name of the current mode or highlighted by using different visual effects.

Other specific suggestions included: Showing a cursor with a sticky note attached to it if next click will create a marker or changing the mouse cursor to a pen when the user can draw on the page. Several people suggested changing the design of the toggle buttons for specifying user mood: having radio instead of toggle buttons or replacing the textual labels with smileys to make the design more appealing.

The trend for usefulness and perceived usability of PDotCapturer, measured using the answers to the questionnaire, looked promising (e.g. a rating of 3.5 and 4.16 out of 5 on average for the qualities 'easy to use' and 'easy to learn', respectively).

4.3.2 Evaluation of Version 2 of PDotCapturer

Procedure and Participants

Version 2 of PDotCapturer has been used and evaluated in several Participatory Design workshops conducted in the scope of the Go-Lab project, two of which are presented here based on a paper describing these evaluations (Heintz and Law, 2015). To give examples from the two main user groups, teachers and students from upper primary school up to universities, the results regarding usefulness, usability and user experience from one student- and one teacher-based event are reported.

The student-based event was conducted in March 2015 with 32 first-year university students in computing who used PDotCapturer to express PD ideas on a complete online lesson on Electricity - An Alternative approach of Ohm's Law (`http://www.golabz.eu/spaces/electricity-alternative-approach-ohms-law` [last accessed: 08/11/2016]), including learning scaffolding apps and an online lab. The students were briefed in class but performed the actual evaluation in their own time, either at home or during another class. 28 (24 male and 4 female; all university students currently studying in the United Kingdom) of the participants filled in the questionnaire about PDotCapturer (Appendix A.3) completely, 4 only partially (not answering the questions reported here).

The teacher-based event was conducted in April 2015 as part of a teacher conference with 20 teachers who used PDotCapturer to express PD ideas on an online chemistry laboratory called BOND lab (`http://www.golabz.eu/lab/bond` [last accessed: 08/11/2016]) on precipitations. 19 (2 male and 17 female; 8 Primary School, 9 Secondary School, 1 Further Education college, 1 university; all from the United Kingdom) of the participants filled in the questionnaire about PDotCapturer (Appendix A.4) completely, one only partially (not answering the questions reported here).

Instruments

The evaluation approach adopted was primarily subjective self-reporting. After using PDotCapturer, the participants were asked to complete a questionnaire on usability and user experience. Section 1 of the questionnaire contained demographic questions (question 1-4 in Appendix A.4). Section 2 is based on two standardized questionnaires, namely AttrakDiff (Hassenzahl and Monk, 2010) and Usability Metric for User Experience (UMUX) (Finstad, 2010) (question 5 and 6 in Appendix A.4 respectively), whereas Section 3 and 4 have been developed to get first impressions on the usability and feature set of PDotCapturer (question 7 and 8-11 in Appendix A.4). Consequently the questions are based on three usability attributes: Learnability (easy to learn), Efficiency of Use (ease of use), and Subjective Satisfaction (useful) (Nielsen, 1994b) enhanced with questions regarding the usefulness of different PDotCapturer features and free form fields to explain in more detail.

- **Background:** Some demographic data have been collected from the participants, including gender, age, school type, and country. These can be covariates for the evaluation results of PDotCapturer, but are not dealt with accordingly in this study, given the small sample size.
- **User Experience (AttrakDiff and UMUX):** These two standardized questionnaires capture quantitative data to be analysed with appropriate statistical methods.
- **Usability:** A set of 11 usability statements collects quantitative data to be analysed with appropriate statistical methods.
- **Features:** The open-ended questions enable participants to give qualitative comments on existing features of PDotCapturer and to propose new ones. However, as the response rate to this question is rather low, the limited findings are not reported here.

Put concisely, AttrakDiff is grounded in the theoretical assumption that the hedonic and pragmatic quality of an interactive product contributes to its attractiveness, which in turn leads to positive user experience and intention to use. UMUX is built upon the traditional notion of usability (ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability) and System Usability Scale (SUS, Brooke, 1996) with the aim of producing an even more parsimonious scale for industrial use.

To avoid the questionnaire becoming too long and tedious to fill in for the participants, a focus was put on usability and User Experience. However, it would also have been interesting to measure other factors such as cognitive load while using the tool.

Results

As students and teachers are the two main target groups of the Go-Lab PD activities and thus the main user groups for PDotCapturer in the context of this work, it is not only of interest how they see the tool, but also, how the views compare and where teachers and students differ. This is important to make sure that the right adaptations are made to PDotCapturer to appeal to both user groups and that positive changes for one group do not have a negative effect on the other.

To compare the responses of students and teachers regarding the 'Usability of PDotCapturer' statements in the questionnaire a non-parametric Mann-Whitney-U test has been performed, given that the results of a Shapiro-Wilk test of normality indicates a non-normal distribution of the data. Only two question showed a significant statistical difference between teachers and students: "Sticky notes marking the spot of my feedback were useful." with U=177.50, $p<0.05$, student mean rank=20.84, teacher mean rank=28.66 and "It was easy to switch between different modes (instructions, interact with app, give feedback)." with U=158.50, $p<0.05$, student mean rank=20.16, teacher mean rank=29.66. A possible explanation why teachers (mean=3.53) found it easier, as compared with their student (mean=2.79) counterparts, to switch between PDotCapturer modes might lie in the different software artefacts evaluated by the students and teachers. The teachers were evaluating an online lab, which requires less navigation and fewer switching between interacting with the lab and expressing ideas using PDotCapturer as compared to the students' evaluation of a whole online lesson, which not only includes an online lab but also several scaffolding apps and other online resources (e.g., Electrical circuit lab, Conclusion Tool and YouTube videos). While on average both user groups found the virtual sticky notes useful, it is unclear why teachers (mean=4.32) did this significantly more than students (mean=3.71).

The other questions did not show any significant difference in the ratings between the students and teachers, suggesting that their perceptions of the functionalities of PDotCapturer are similar. For the teachers' responses, the mean values were all higher than 3.0, indicating their overall neutral attitude towards the statements or their tendency to agree with the statements. For the students' responses, the mean values were less than 3.0 for two of the statements, suggesting that they tended to disagree on them while being neutral or inclined to agree on the other statements. The first statement on which the students disagreed is "Design of PDot was visually appealing." (mean = 2.79, SD = 1.1) and the second one is "It was easy to switch between different modes (instructions, interact with app, give feedback)." (already discussed above). The rating of students for the former statement, albeit not significantly different from the teachers (mean = 3.32, SD = 1.06), indicates that the design seems sufficient for the teachers whereas the students seem to see the need for improvement.

To compare the responses of students and teachers regarding the 'User Experience of PDotCapturer' statements (which were rated by 26 students and 19 teachers) in the questionnaire, an independent samples t-test has been performed for pragmatic and hedonic quality measured through AttrakDiff, given that the result of a Shapiro-Wilk normality test indicates the normal distribution of the data (significance levels greater than 0.05, see Table 4.2).

		Shapiro-Wilk		
		Statistic	**df**	***p* value**
Pragmatic Quality	**students**	.975	26	.748
	teachers	.938	19	.245
Hedonic Quality	**students**	.975	26	.755
	teachers	.907	19	.064

Table 4.2: Normality test results for pragmatic and hedonic quality (from AttrakDiff word pairs).

For the User Experience question based on AttrakDiff, no significant differences in the perception of the pragmatic quality between students and teachers have been found. But teachers rated the hedonic quality significantly higher than students did with t[43] = -2.18, p<0.05 (student: mean = -0.27, SD = 0.78 and teacher: mean = 0.37, SD = 1.19). Both mean ratings are in the average region, thus there is still room for improvement, but the teachers already perceived a higher potential for getting pleasure from the tool usage. When analysing the statements individually using a non-parametric Mann-Whitney-U test, as the results of a Shapiro-Wilk test of normality indicates a non-normal distribution of the data, two of the word pairs show significant differences. One is "dull - captivating" with U=142.00, p<0.05, student mean rank=18.96, teacher mean rank=28.53 and the other one is "cheap - premium" with U=127.00, p<0.01, student mean rank=18.38, teacher mean rank=29.32. The mean student rating of these statements goes towards the negative and the mean teacher rating towards the positive side of the scale. This is calculated from the

7 point scale by using values from -3 for the word on the left (with a negative connotation) to +3 for the word on the right (with a positive connotation), see Table 4.3 for results.

	Students		**Teachers**	
	Mean	**SD**	**Mean**	**SD**
Dull - Captivating	-0.77	1.47	0.26	1.45
Cheap - Premium	-0.65	1.06	0.37	1.50

Table 4.3: Ratings of word pairs from AttrakDiff with significant differences between students and teachers.

These findings correspond to the results of the analysis of the Usability questions, where the students stated that they did not find PDotCapturer visually appealing where the teachers did. For the 4 statements of the UMUX questionnaire, no significant differences in rating between the students and teachers have been found.

Analysing more and possibly bigger dataset in the future, including teachers and students evaluating different or the same applications using PDotCapturer, will help to overcome one of the limitations of the current data analysis. As the samples reported on here not only differed regarding their characteristics (teachers or students), but also regarding the artefact evaluated using PDotCapturer, the latter might also influence the results, thus limiting the certainty of the reasons for the differences. Future analysis and results can show if the responses of each target group are consistent, when using PDotCapturer to express ideas on the same and different digital artefacts, overcoming this limitation of the preliminary data analysis presented here.

The preliminary results of the usability and User Experience evaluation of the improved prototype show that usability and usefulness of PDotCapturer are perceived as sufficient, but there is still room for improvement, at least from the students' perspective especially regarding the visual representation of the tool. These results have been taken into consideration for the re-design and further development of PDotCapturer (i.e. visual re-design for version 3). Besides asking usability-related questions, the questionnaire also gathered functionality-related information, which were incorporated in the new version of PDotCapturer presented in Section 4.2.5.

As PDotCapturer allows participants to freely express their ideas and provide feedback (through freehand drawings and in a single textbox), not only PD ideas are gathered. The responses include a mixture of usability issues, feature requests (requirements), and other matters (e.g. feedback regarding the learning content), but all of these ideas are beneficial to re-shape and further improve the Go-Lab resources.

4.3.3 Further Iterative Evaluation of PDotCapturer

When using PDotCapturer in Go-Lab PD events besides gathering PD ideas observer notes on the tool usage were also collected. This unstructured data was also used to come up with improvement suggestions for the tool. Additionally other evaluations, like the study comparing paper-based and PDot-based PD approach (described in Section 6.1), were used to further evaluate PDotCapturer, e.g. through questionnaires filled in by the participants after they performed the task of expressing ideas. These results were not substantial enough to report them here but also helped to shape the development and improve PDotCapturer.

4.3.4 Evaluation of Version 3 of PDotCapturer

The evaluation of version 3 of PDotCapturer was performed in conjunction with a comparison of PDotCapturer with paper. It is therefore presented in Section 6.2.1.

To summarise the section on PDotCapturer: Based on a set of initial requirements, an initial prototype and two subsequent versions of the tool have been created. The changes between the different versions were informed by different evaluations, collecting feedback on the current state of the tool.

Chapter 5

PDotAnalyser

This chapter presents the iterative development and evaluation of a new PD software tool: PDotAnalyser. It is aimed at designers to support them in their task to analyse the gathered PD ideas. The roots of the tool in PDotCapturer are explained, together with the reasoning for creating a dedicated tool. The two versions of PDotAnalyser created and the changes and improvements from one to the other are described to show the influence of the evaluation (presented as well) on the development process.

5.1 Methodologies

As for PDotCapturer for PDotAnalyser an iterative development approach was applied. An initial design and requirements were derived from PDotCapturer and the way it was used to analyse end-user ideas before the dedicated PDotAnalyser tool was developed.

To measure usability and user experience AttrakDiff2 was used as part of a questionnaire. The remainder of the questionnaire were questions regarding demographics and more detailed questions on functionality and its usefulness to determine future development of the tool. In addition to the questionnaire the think-aloud technique was used to gain insight into how designers used the tool and what issues they encountered.

5.2 Design and Development

5.2.1 Initial Prototype: PDotCapturer to Analyse Ideas

PDotCapturer focused on supporting end-users in expressing ideas in textual and graphical modes. Idea analysis was possible, but not optimally supported: The ideas were retrieved and analysed by logging in with the individual end-users' credentials to view each end-user's recorded ideas individually in the tool. This was tedious and allowed the analysis of the ideas of only one end-user at a time. Alternatively, all ideas could be exported into a spreadsheet where, however, the contextual information provided by PDotCapturer for understanding the idea would be removed, leading to the risk of misinterpretation. These limitations called for more sophisticated means for idea analysis which were implemented in PDotAnalyser.

Based on a set of initial functional and usability requirements informally elicited from the participants who attempted to use PDotCapturer for idea analysis and thus identified its enhancement potential, a first prototype of PDotAnalyser was developed. As designers are the end-users of PDotAnalyser all requirements are classified as designer requirements (DR). PDotAnalyser has to:

- **Interactivity (DR1):** Work with interactive prototypes, to allow designers to recreate the interactions reported by end-users, which lead to issues.
- **Idea Presentation (DR2):** Visualise the relation between interface elements and ideas and present the verbal description given by the end-user when specifying the idea.
- **Drawing Visualisation (DR3):** Present drawings in a way that makes it clear to which idea they belong.
- **Collaboration (DR4):** Allow designers to analyse annotations collaboratively.
- **Access (DR5):** Be easily accessible from anywhere with Internet connection, and work without installation.
- **Context (DR6):** Provide instructions given to the end-user as context of ideas.
- **Aggregation (DR7):** Support aggregation of data (explicit and implicit) from different end-users.

5.2.2 Version 1 of PDotAnalyser

Features

Where PDotCapturer only aggregates all the ideas specified by one participant per screen, PDotAnalyser aggregates all ideas expressed by all participants per screen (or more precisely: per PDotCapturer scenario instruction step).

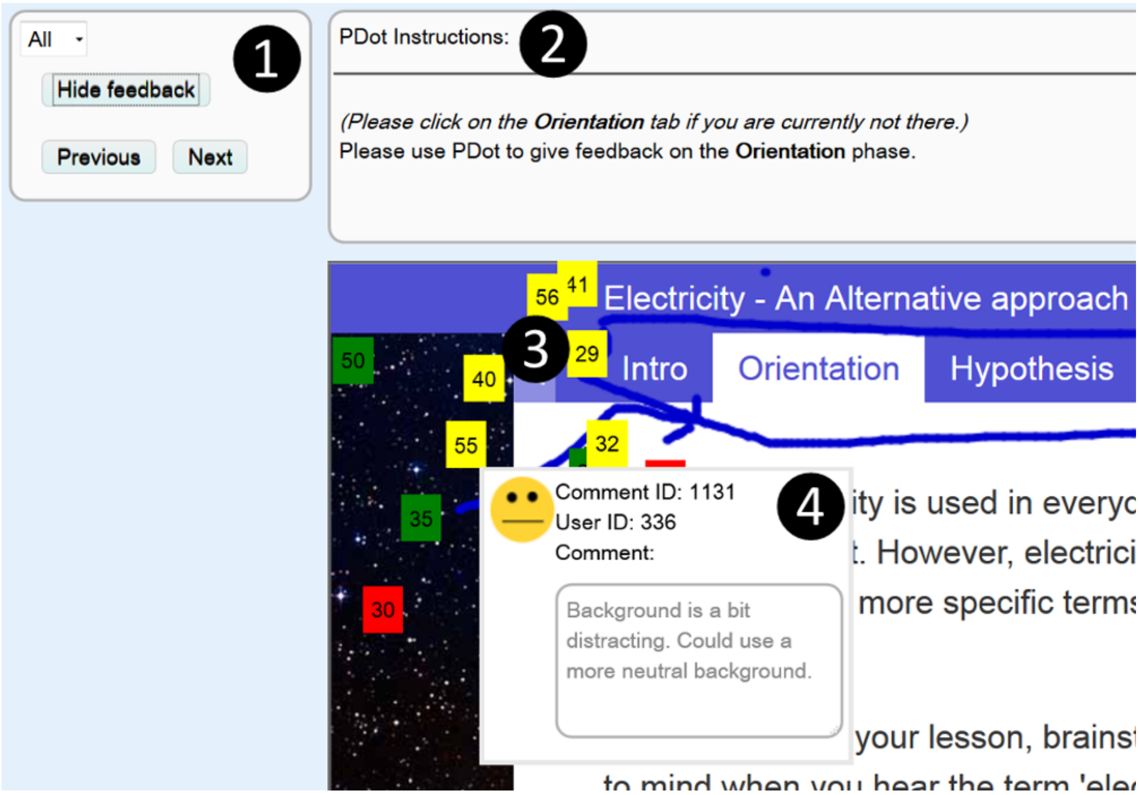


Figure 5.1: Screenshot of PDotAnalyser showing the 'aggregated mode', displaying all ideas that have been expressed on this page of the prototype (numbers in circles added for referencing).

PDotAnalyser consists of three major parts, namely: navigation, instruction, and visualisation (Figure 5.1). The tool can be run in two modes: single (Figure 5.2) and aggregated, and has several distinct features (Figure 5.3 and Figure 5.4). They are elaborated in the following.

Three Major Parts

Navigation box: In the upper left hand corner the functionality to navigate through different ideas generated by all end-users participating in a PD workshop is provided, together with filter options where applicable (❶ in Figure 5.1).

PDot instructions: The instructions for end-users are shown on top of the page (❷ in Figure 5.1). When analysing the ideas, the designer can use these instructions as context information on what the end-user was doing when expressing the ideas.

End-user idea visualisation: Markers (numbered rectangles that are colour-coded based on the emotional response the end-users specified regarding the current design when expressing their idea [green = like, yellow = neutral, red = dislike]) and drawings created by the end-users are shown overlaying the prototype (❸ in Figure 5.1). Detailed information (textual description of the idea, IDs, and emotional response rating) is shown in a pop-up (❹ in Figure 5.1) when moving the mouse cursor over a marker.

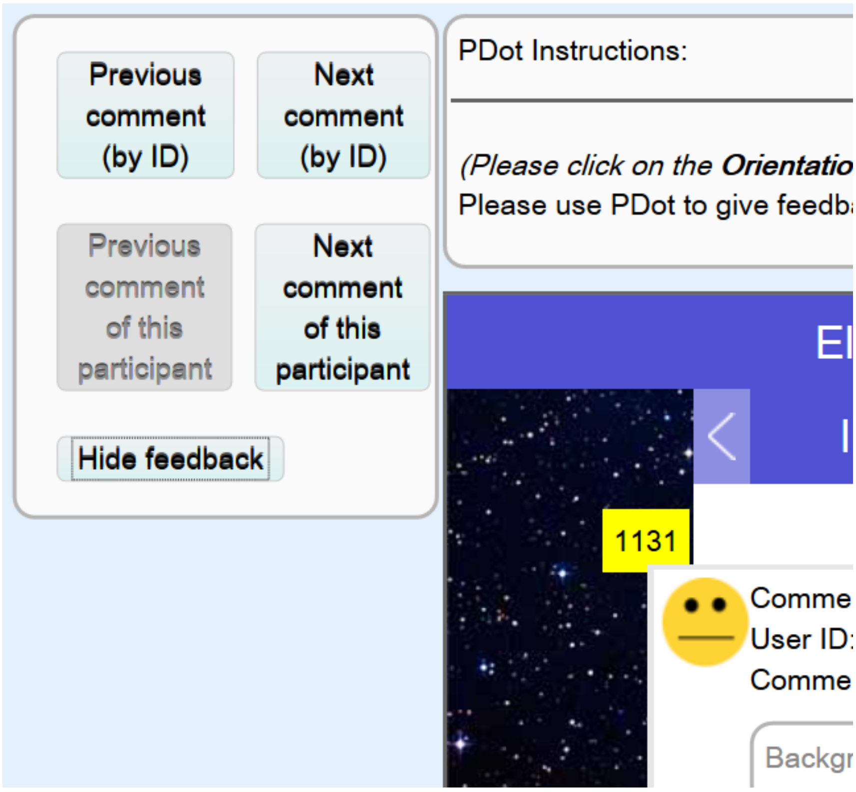


Figure 5.2: PDotAnalyser in 'single mode'.

Two Modalities

PDotAnalyser has two modes: single mode (Figure 5.2) and aggregated mode (Figure 5.1).

In the **single mode** only one idea (together with its meta information) is shown at a time. The designer analysing the ideas can navigate either through the ideas of one end-user or through the ideas of all participating end-users using the 'previous' and 'next' buttons provided. To show the details for one particular idea, its ID can also be specified as part of the URL when accessing the tool. Additionally, the interface can be switched between displaying the ideas gathered and interacting with the prototype. This helps the designer to get a better understanding of the context and possible reasoning of the ideas. It enables the reproduction of interactions that end-users performed with the prototype and subsequently described when expressing ideas.

In the **aggregated mode** all ideas expressed on the current screen of the prototype are displayed at once. This gives a quick overview about all ideas and the areas of interest, where many or no ideas have been given. Analysis is further supported by colour-coding of the markers based on the reported emotional response of the end-user regarding the existing design. At the same time this mode supports quick navigation through several ideas one after the other. This can be done either by hovering over the markers with the mouse cursor or by using the 'previous' and 'next' buttons provided in the navigation box. The meta information for the idea is then displayed in a pop up and the associated freehand drawing is highlighted by altering its colour. An additional filter option allows designers to view the ideas of a single end-user who participated in the PD workshop instead of all.

Distinct Features

Figure 5.3 and Figure 5.4 highlight some of the distinct features of PDotAnalyser to support idea analysis, which are visible in the GUI, described in the following sections.

The ideas of end-users are presented on top of the (interactive) prototype (Figure 5.3, top) where it was given (Figure 5.3, bottom). This supports the process of making sense of the ideas by providing the visual and interactive context of each idea as background information as well as providing the possibility to recreate the interactions the end-user performed before expressing the idea.

PDotAnalyser automatically generates a 'heatmap' visualisation of the emotional response (dislike, neutral, like) the end-user specified regarding the current design when expressing their idea. This is done by creating a semi-transparent circle coloured red, yellow, or green around the position of each marker (left hand side in bottom screenshot of Figure 5.3). This heatmap can either be displayed on its own or as a layer between the idea markers and the prototype. It allows for a quick overview of 'problem areas' (red/orange/yellow), where negative opinions on the de-

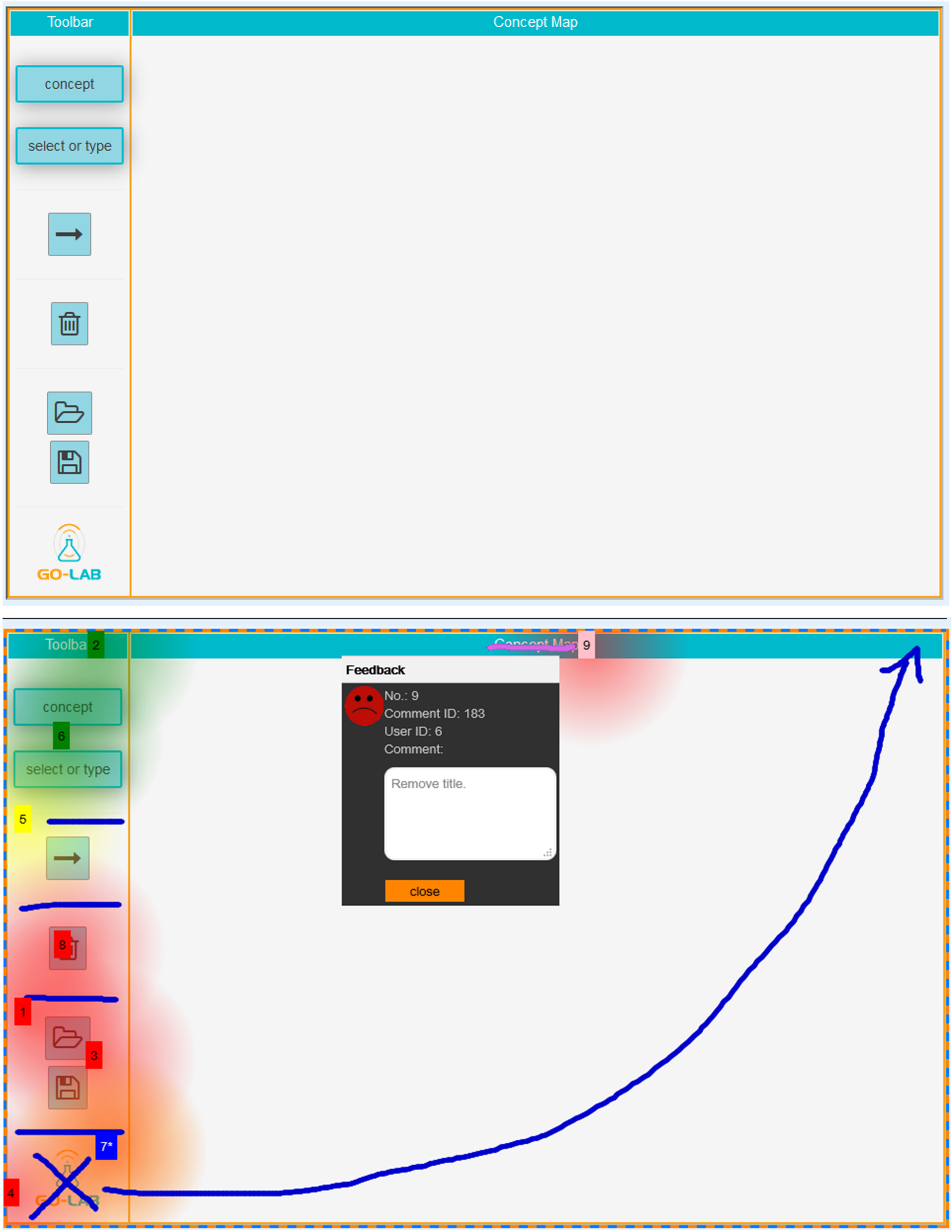


Figure 5.3: Top: A screenshot of the prototype to be reviewed by end-users (i.e. the target of the PD activity). Bottom: The idea visualisation features of PDotAnalyser overlay the prototype.

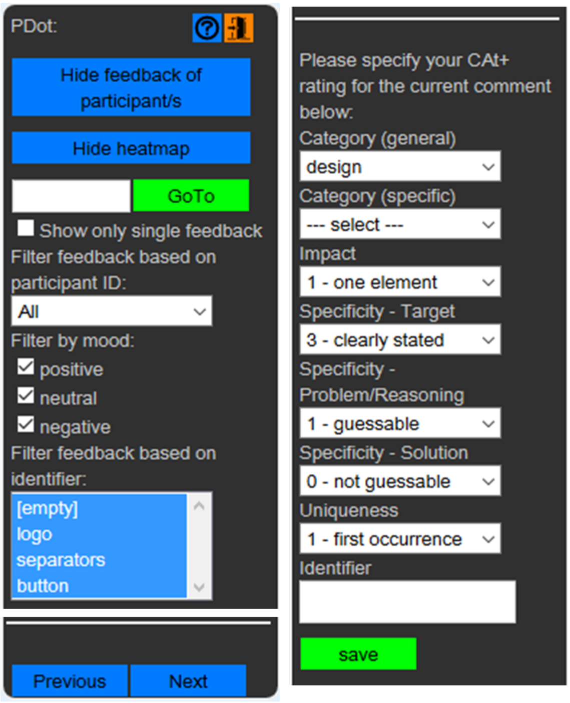


Figure 5.4: Filter options and CAt+ coding GUI in PDotAnalyser.

sign have been expressed. At the same time, the heatmap also displays an overview of 'encouraging areas' of the prototype, which have been perceived positively (green overlay). This information can be used by end-users and designers together to facilitate a discussion about the ideas. Alternatively, it can be utilised by the designers on their own, to analyse the ideas further and come up with a redesign for the prototype.

Automated aggregation of ideas (based on location), options to filter the ideas (Figure 5.4, left), and 'single mode' (Figure 5.2) support the analysis task and sense making process by preventing information overload. The integrated support for further analysis through rating of ideas using the CAt+ coding scheme (Figure 5.4, right) as an integral part of the tool makes it possible to have all the context information and to use the other features while doing the rating. The CAt+ rating can then support further filtering, clustering, and aggregation when working with the ideas expressed by the end-users participating in the PD activities.

5.2.3 Version 2 of PDotAnalyser

The evaluation and findings of version 1 of PDotAnalyser (Study 1 and 2), leading to the changes described in this section, are presented in Section 5.3 and summarised here:

- Combine aggregated and single mode
- Improve navigation between ideas
- Add element identifier

To address these improvement suggestions in a later iteration of PDotAnalyser aggregated and single mode were integrated into one, with the option to switch by ticking a check-box in the GUI. Regarding the improvement suggestion for the navigation an input field was added to specify the ID of an idea to navigate to. In the area of the CAt+ coding scheme GUI the option to specify element identifiers was added.

Another major change from version 1 of the PDotAnalyser tool to the final (in the scope of this thesis) version 2 was a visual re-design (according to the re-design of PDotCapturer).

Figure 5.5: Aggregated marker that is displayed in case two or more ideas have been given at the same spot.

One issue reported with version 1 of PDotAnalyser was that idea markers could be hard to access when two or more ideas were expressed at the same location. In this case the visual representations of the markers were created on top of each other and thus the user could only interact with the topmost. To address this issue 'aggregated idea markers' have been introduced to PDotAnalyser. If two markers are too close to each other they are now combined to one aggregated marker (which indicates the existence of 'sub-markers' through an asterisk next to the marker number). On mouseover this aggregated marker shows a vertical list of the markers that have been aggregated (see Figure 5.5 for details). From there the user can interact with the markers as usual (e.g. on mouseover they show a pop-up with the detailed information regarding this idea).

Besides overlapping markers, another issue reported with markers was, that accidentally moving the mouse over a different one, while working with the idea functionality (e.g. performing a CAt+ rating), would open the information and pop-up related with this marker, so that the initial focus was lost. To address this issue, 'locking' of idea information was implemented. Besides showing the pop-up with the information related to this idea on mouse-over over the idea marker, clicking on the idea marker kept it open and deactivated reacting to mouse-over for all other idea markers. To close the pop-up the user could either click on the marker again or use a newly added close button in the pop-up.

With the rework of the CAt+ coding scheme (see Section 3.2.6 for details), resulting in a smaller set of CAt+ items, also the PDotAnalyser interface to specify a CAt+ rating for each idea was slimmed down (Figure 5.6). Additionally autosave was implemented for the CAt+ rating information entered.

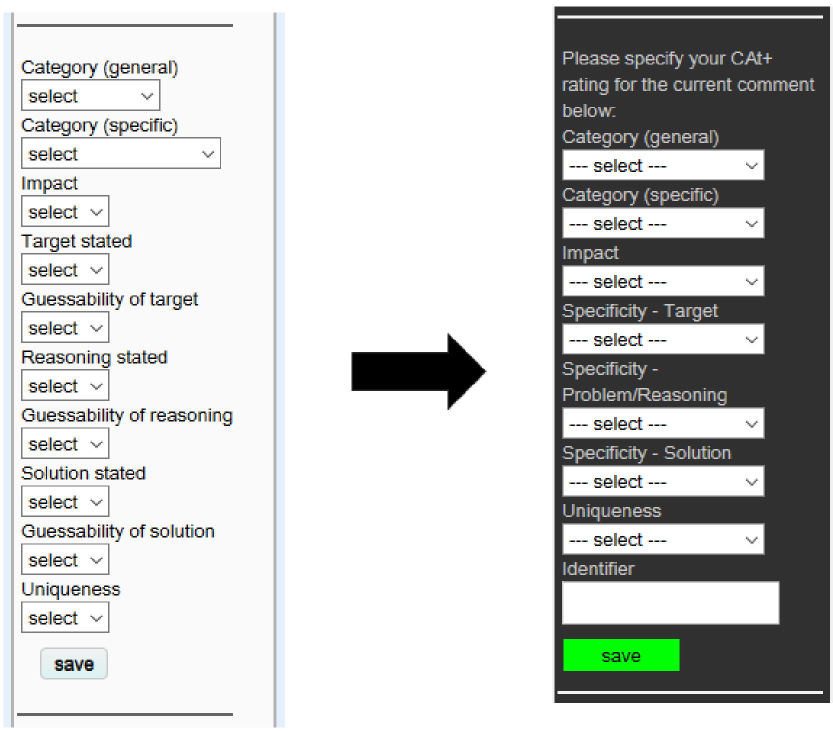


Figure 5.6: Reduced set of CAt+ items in the interface.

5.3 Evaluation

To evaluate PDotAnalyser, three studies were conducted (Figure 5.7) evaluating the usefulness, usability (Nielsen, 1994b) and user experience (Garrett, 2010) of this tool. In earlier studies (Workshop A and Workshop B) two iterative versions of PDotCapturer were used for PD activities on webapps to generate a database for the PDotAnalyser evaluations in Study 1, Study 2 and Study 3. The reason for not using the same set of gathered ideas for all three studies was that, like PDotAnalyser, PDotCapturer was also improved over time; Study 2 and Study 3 reflected this by using ideas gathered with the improved version of PDotCapturer.

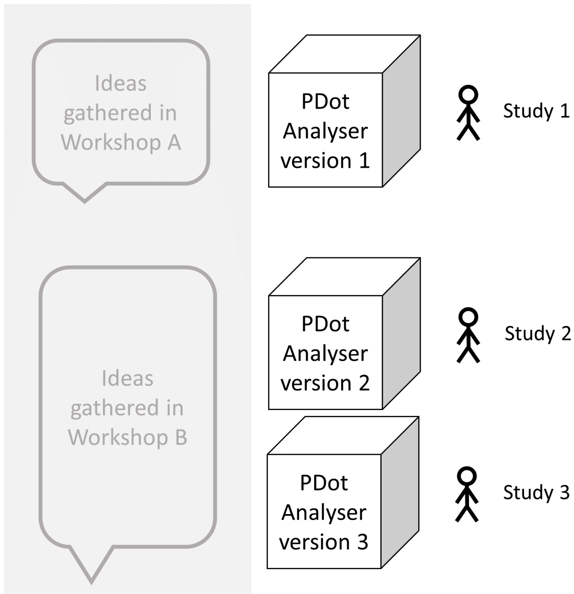


Figure 5.7: Formal Evaluations. Ideas as database were gathered in Workshop A and Workshop B using PDotCapturer.

5.3.1 Procedures and Participants

Participants with a range of experience as designers and familiarity with the webapps of interest were recruited, from computer science undergraduates to developers of those webapps. As PDotAnalyser, like most software tools, should support both novice and expert users, it was evaluated with both user groups. It was started with novices based on the conjecture that the core features of the tool proven usable for less experienced users should also be usable for their more experienced counterparts, but this logic is not necessarily applicable the other way round.

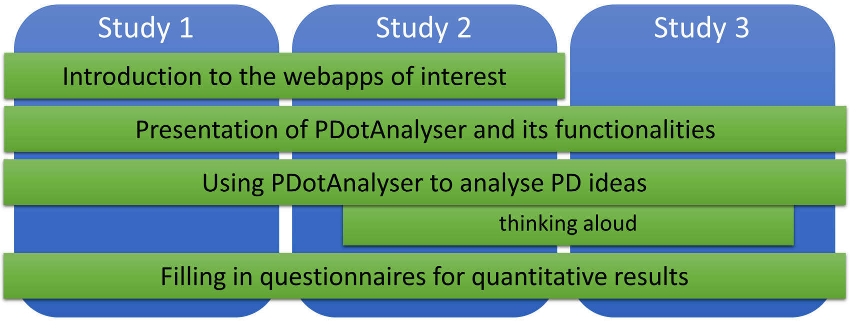


Figure 5.8: Study design for PDotAnalyser evaluation.

Study 1 was conducted with 19 Computer Science students (16 male, 3 female) as part of their programming and HCI education, including lectures and practical sessions on Participatory Design (Figure 5.8). As preparation for the evaluation they had lectures covering Participatory Design principles and scenarios where the webapps of interest could be used. In the beginning of the two-hour PDotAnalyser evaluation session, the participants were given a short presentation of the tool and its functionalities. They were then asked to analyse ideas gathered earlier with PDotCapturer (Workshop A in Figure 5.7). Afterwards they filled in a questionnaire (Appendix A.5) to collect data regarding usability and User Experience (UX) of PDotAnalyser (measured by AttrakDiff2, Hassenzahl and Monk, 2010) as well as feedback regarding existing and additionally requested features (Figure 5.8). The AttrakDiff2 questionnaire was used for this quantitative study, because it is a standardized tool and fitted the evaluation goal. Usability and UX of PDotAnalyser were measured, because it should not only be usable and useful but also pleasant and desirable to use.

Study 2 was conducted with five early stage researchers and practitioners in the field of information technology (3 male, 2 female). They had various backgrounds but generally a good knowledge of computer programming. After a presentation of PDotAnalyser, its purpose and functionalities, the participants were asked to analyse ideas in the form of textual comments and drawings gathered in a workshop about PDotCapturer earlier (Figure 5.7). Compared to Study 1 the participants were using a slightly enhanced version of PDotAnalyser which included functionality to rate the gathered ideas using the CAt+ coding scheme (described in Section 3.2.6). In addition to the questionnaire used in Study 1, to gather more quantitative feedback, which was extended by a set of questions regarding the CAt+ rating (question 11 in Appendix A.6), further qualitative feedback was collected using the think-aloud approach while the participants interacted with the tool (for about an hour, Figure 5.8). This approach was applied to get a deeper understanding of how the participants actually used PDotAnalyser and which issues they encountered in doing so.

For Study 3 a further improved version of PDotAnalyser and the CAt+ coding scheme were evaluated by five developers who were involved in implementing the webapps of interest. The same approach, methods, and PD idea data as in Study 2 were used (Figure 5.8). As in Study 2 the think-aloud approach was used during this phase to gather in-depth feedback on the tool and the CAt+ coding scheme. When the participants felt confident that they had a clear understanding of the tool as well as the coding scheme and had provided the qualitative comments they could think of, they were then asked to complete a slightly updated version of the questionnaire used in Study 2 (e.g. SUS questions were added, Appendix A.7).

5.3.2 Evaluation Results

Usability

To measure the usefulness and perceived usability of PDotAnalyser, the three qualities 'usefulness', 'ease of use', and 'ease of learning' were rated in the questionnaire by the participants. The results are presented in Table 5.1.

	Study 1 (n = 18)	**Study 2 (n = 5)**	**Study 3 (n = 5)**
	Aggregated mode		
usefulness	3.44 (.86)	4.60 (.55)	4.80 (.45)
ease of use	3.11 (.96)	3.80 (1.01)	4.20 (.45)
ease of learning	3.17 (1.01)	3.40 (.89)	4.40 (.55)
	Single mode		
usefulness	3.17 (.86)	4.20 (.84)	4.80 (.45)
ease of use	2.94 (.80)	3.20 (1.01)	4.20 (.45)
ease of learning	3.11 (.90)	3.60 (.89)	4.40 (.55)

Table 5.1: Mean ratings of the statements regarding the qualities of PDotAnalyser on a 5-point Likert scale from 1: Strongly Disagree to 5: Strongly agree, with standard deviations in brackets.

When comparing the usability results of the different studies over time (Table 5.1) it can be seen that the prototype improved in all three qualities (usefulness, ease of use, ease of learning) and for both visualisation modes (aggregated and single mode).

User Experience

To measure the perceived UX of PDotAnalyser the responses to AttrakDiff2 questions were analysed. The results are presented in Table 5.2.

AttrakDiff2	Study 1 (n = 18)	Study 2 (n = 5)	Study 3 (n = 5)
PQ	4.42 (.67)	4.80 (.69)	5.45 (.45)
HQ	4.11 (.77)	4.20 (.76)	4.80 (.74)
Goodness	4.28 (1.41)	5.20 (1.30)	5.80 (.84)
Beauty	3.94 (1.35)	4.40 (.55)	4.60 (1.34)

Table 5.2: Mean ratings of Pragmatic Quality (PQ), Hedonic Quality (HQ), Goodness, and Beauty from AttrakDiff2 items (7-point semantic differential scale) with standard deviation in brackets.

The AttrakDiff2 results also show a trend of improvement (Table 5.2). This shows that the changes performed for the different iterations resulted in a higher perceived hedonic and pragmatic quality, goodness, and beauty.

Feedback on Heatmap Visualization

The heatmap showing the emotional responses (⑥ in Figure 5.9) was found especially useful by the participants. One of the participants in Study 3 even wanted more of such heatmap visualisations, in combination with the CAt+ ratings. For example, a colour-coded 'Impact' heatmap displaying the impact distribution of ideas generated by the end-users.

Improvement Suggestions

Suggestions on PDotAnalyser features resulting from Study 1 are based on four free text questions in the questionnaire. The following list shows an abbreviation for each question together with the feature mentioned most frequently:

- Liked features - Being able to see all the ideas in one screen at once in the aggregated mode.
- Disliked features - Markers were sometimes overlapping.
- Features to be added - An option to filter the ideas by mood.

- Features to be removed - ‘nothing’ or ‘none’.

To address these improvement suggestions ‘aggregated markers’, which combined overlapping markers into one, and an option to filter ideas based on emotional response information were added in later iterations of PDotAnalyser (see above).

In Study 2 a variety of suggestions was gathered through think-aloud. Three of the main improvement suggestions are presented in the following:

- **Connect the aggregated and single mode:** The two modes initially had two different URLs to access them and were thus separated from each other. The participants asked to add functionality to seamlessly switch between the two modes.
- **Improve navigation:** Instead of only having buttons to navigate to the previous and next idea, a more flexible navigation should be provided, e.g. a ‘pagination’ type of navigation, allowing to jump to the first or last idea on this screen and to specify a idea number in the list to go there directly.
- **Add element identification:** Functionality to specify the design element addressed in a comment should be added, allowing to group or filter ideas based on elements (e.g. ‘show me all ideas regarding the ‘login button”).

From the think-aloud responses in Study 3 several improvement suggestions can be derived. The following list contains the points most commonly raised (Figure 5.9):

- Instructions should indicate how many more steps there are, e.g. by displaying this information as 2/5 ①.
- All buttons, but the logout button, should have the same colour, not some being blue and others green ②.
- The ‘filter by user id’ functionality should not filter out the ideas by this end-user, but the ideas by all others ③.
- While rating the idea using CAt+ you should be able to write down and save your guesses for ideas for which you can only guess a target, problem/reasoning, or solution ④.
- It should be indicated more clearly (e.g. by changing the mouse cursor) that clicking on a label in the PDotAnalyser GUI opens a help pop-up with more detailed information ⑤.

- The terms used in the CAt+ coding scheme should be adapted so that they are either all verbs or all adjectives, not a mixture of both as it is right now.

- CAt+ categories could be further clustered and the terms streamlined, as they are currently a mixture of issue-related terms (i.e. terminology) and action-related terms (i.e. change). This could be addressed by rating 'issue' and '(proposed) solution' separately (Figure 5.10).

Figure 5.9: Final version of PDotAnalyser evaluated in Study 3. Purple circles with numbers added to indicate participant feedback described in the text.

The different think-aloud sessions created a vast number of re-design suggestions that cannot all be reported here. Additional issues that have been raised by several of the participants are:

- Reduce the number of items in the CAt+ interface

- Autosave the data entered

- Instead of dropdowns to specify the CAt+ rating other input modalities could be provided
- If several idea markers are (exactly) on top of each other, only the top one can be accessed with the mouse

Not all of these suggestions could be addressed in the scope of this research, because of time restrictions, but provided useful input for further refinements of the PDotAnalyser tool.

Another interesting result from the think-aloud sessions is that some participants did not find the added-value of the 'single mode'. Instead of hiding all other ideas they found that just dimming them (making them semi-transparent) would be sufficient.

To summarise the section on PDotAnalyser: Initially PDotCapturer was also used to access the ideas gathered. However, as this was cumbersome and did not realise the full potential of software support for designers, a dedicated tool was developed based on the experience of using PDotCapturer. Input from end-users gathered on PDotAnalyser and its existing feature set influenced the changes and extensions for the following versions.

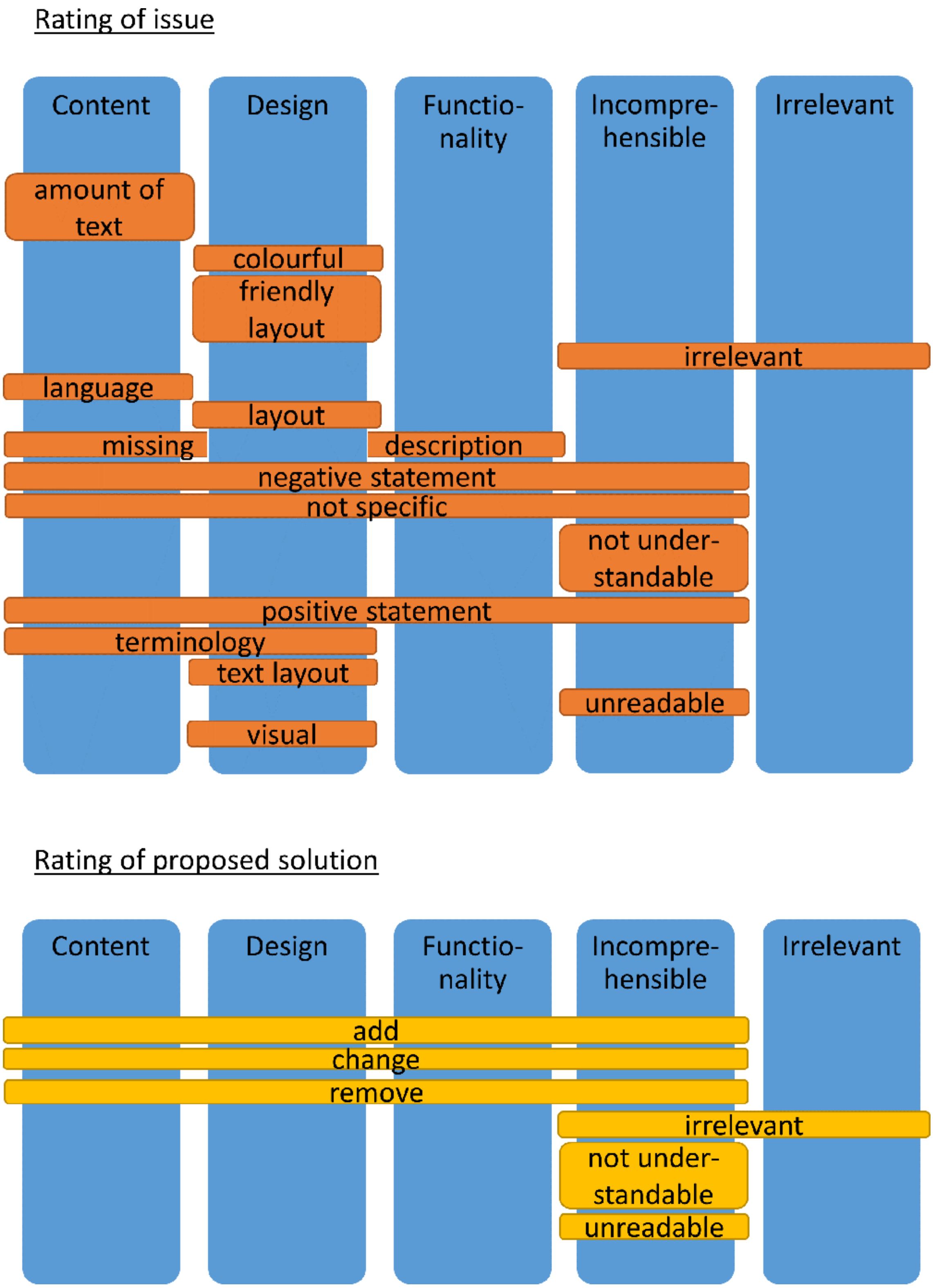


Figure 5.10: Proposed changes for CAt+ categories: Rate issue and proposed solution separately.

Chapter 6

Comparison of Paper-Based Approach and PDotCapturer

As explained in the introduction it is very important to compare tool-based approaches with their paper-based counterparts to identify similarities and differences. Therefore this chapter presents empirical studies evaluating the different versions of PDotCapturer comparing them with similar paper-based PD approaches. Besides quantitative differences, qualitative differences are identified by applying the CAt+ rating scheme presented earlier.

6.1 Paper and Version 2 of PDotCapturer

It is assumed that using a dedicated tool (PDotCapturer) rather than a general tool (myBalsamiq) could improve the outcome from the tool-perspective. This motivated the repetition of the comparison presented in Section 3.3, this time with a dedicated tool. In addition to addressing RQ4 and RQ6 (Section 1.2) the study presented in this section aims to add to the database for the comparison of paper and tool by conducting another study leading to a clearer picture and more thorough findings in which context to use which option.

With the higher maturity of the software artefacts evaluated (prototypes rather than mock-ups) the paper-based approach had to be altered slightly from the one described in Section 3.3 to take into account their higher interactivity. As this could no longer be reflected on paper accordingly instead of just using printouts for the evaluation, the end-users worked through the interactive prototype on the computer screen and used the paper only to note down their ideas, not to assess and evaluate the prototype from there.

6.1.1 Design of Empirical Study

PD Study with Interactive Prototypes

Based on earlier PD activities, including evaluation and re-design of interactive mock-ups, interactive prototypes of online lessons have been created. Those are composed of learning material (for example as text and embedded YouTube videos), scaffolding apps to support the learning tasks, and an online lab for conducting experiments. From the list of available prototypes on the project portal two of equal complexity but with different topics (to cover a broader range of teacher expertise) were selected for the evaluation. Figure 6.1 shows a screenshot of the 'Electricity' online lesson on the left (teaching the students about Ohm's law) and the 'Splash' online lesson on the right (teaching the students about Archimedes' principle).

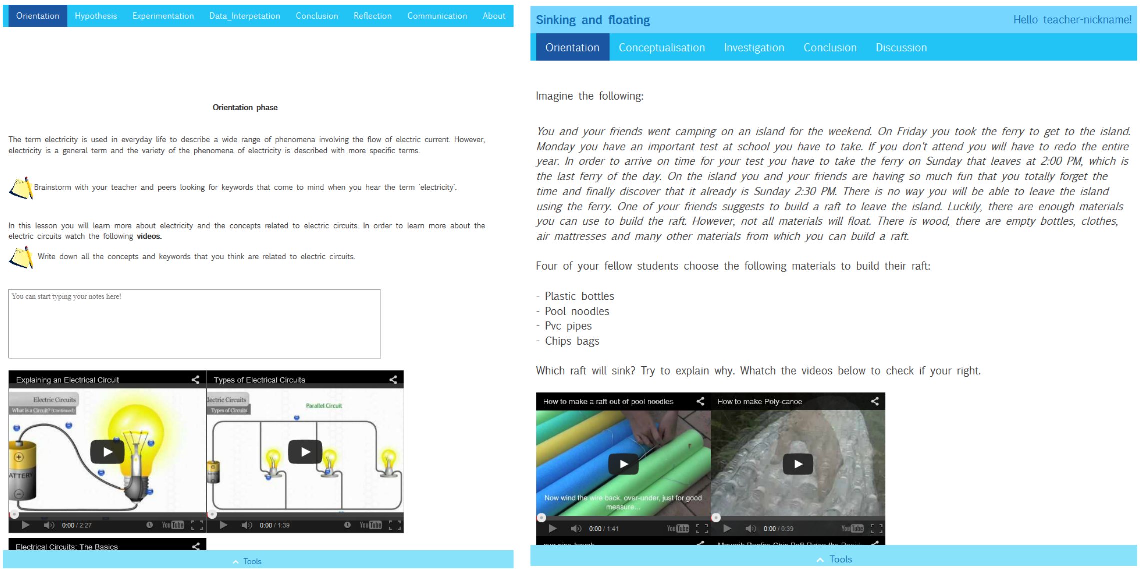


Figure 6.1: Prototypes of online lessons used in the evaluations ('Electricity' on the left and 'Splash' on the right).

The two approaches used to evaluate the interactive prototypes, paper-based and tool-based, are explained in more detail in the following two sections.

Paper-Based Evaluation of Interactive Prototypes

For the paper-based evaluation of the interactive prototypes ‘evaluation booklets’ have been created. Those were printouts of each phase in the online lesson, sometimes spread over several pages, if they were larger than one screen or if a scaffolding app had different tabs of which only one could be visible on each printed page. The participating teachers then got the task to try out the features that their students would use while learning going through this online lesson. Because of time restrictions they were not asked to explore the entire online lesson in full, but to try out each different feature at least once. For example if there were several YouTube videos embedded, the teachers were not asked to watch all, but to look at least at one of them to get an idea of the interaction mechanisms and to therefore be able to express ideas on the inclusion of and interaction with YouTube videos in general. Besides going through the prototypes on their computers, the participants had the task to note down their PD ideas in their evaluation booklet. During the evaluation two researchers took observer notes, helped out where needed, and if necessary encouraged the participants to note down their findings also in the booklet and not only voice them together with their questions to the observers.

Tool-Based Evaluation of Interactive Prototypes

The tool-based evaluation of the interactive prototypes was very similar to the paper-based evaluation. To familiarize the participants with the PDotCapturer tool used for the evaluation, it was first presented using an example unrelated to online learning (to not bias the input gathered later on). The participating teachers were then given the task to work through the online lesson (as described for the paper-based evaluation) and use PDotCapturer to express their ideas. As the paper- and tool-based evaluations took place in the same room at the same time the same two researchers supporting the participants using the paper-based approach again took observer notes, helped out where necessary, and encouraged the participants to report the ideas they had.

6.1.2 Participants and Procedure

The PD study was conducted at a summer school event with 39 teachers from 9 European countries in July 2014.

To fit in the overall time schedule of the one week summer school, the PD study was split up into two workshops on two different days. On the first day about a third of the participants used the paper-based approach (6 groups, one of which switched to the tool after the first page of the evaluation booklet) and the rest of the participants (12 groups) used the tool-based approach to express ideas on the online lesson on Electricity in Workshop-1. On the second day they swapped the approach and expressed ideas on the Splash online lesson on Archimedes' principle (with 13 groups using the paper-based and 4 the tool-based approach) in Workshop-2.

On both days providing PD ideas was embedded in a PD session with introductory material, questionnaires on usability and User Experience, and group feedback. The session on the first day was 120 minutes whereas the one on the second day was only 80 minutes, the participants therefore had a little bit less time with the Splash online lesson (40 min.) than with the Electricity one (55 min.). This time difference should not have a huge influence on the results as it effected both approaches equally, the participants were already a little experienced in interacting with the online lessons and scaffolding apps, and because the Splash online lesson has less phases and can thus be finished a little bit faster than the Electricity one. Because of limited hardware resources the teachers worked mostly in pairs (some alone and some in groups of three), sharing a PC (and evaluation booklet in case they were using the paper-based approach).

The above described setup enabled two within-subject comparisons: Comparison 1 involved the groups who used paper in Workshop-1 and then used the tool in Workshop-2. Comparison 2 involved the groups who used the tool in Workshop-1 and then paper in Workshop-2.

6.1.3 Data Analysis

To ease the data handling during analysis, all results have been transcribed (from paper-based approach) or exported (from tool-based approach) to Excel tables. In doing so one (already anticipated) advantage of the tool became visible: While it took about 8.5 hours to transcribe the paper-based ideas, it took only a couple of seconds to a few minutes to export the tool-based ideas from the database. Regarding Research Question 3 it can thus be concluded that using an appropriate PD tool can decrease the data processing time tremendously, compared to paper.

To check back with the actual ideas, in case the transcriptions were not clear without the context, the evaluation booklets and PDotAnalyser were available throughout the analysis for the raters to look at the idea in its context. The data has then been coded in a fully crossed design (Hallgren, 2012) by two HCI researchers with about one and seven years of experience in usability research.

6.1.4 Results

Results regarding RQ4 (Number of Ideas)

By conducting the PD activities 343 idea items were collected, 202 of which were gathered through the paper-based approach. 141 of the ideas were expressed using the tool-based approach, thus using the paper-based approach resulted in about 1.43 times the number of ideas compared with the tool-based approach.

Although that is still a noticeable margin it is a considerable change over the differences between paper and general tool reported as result of the previous studies, where overall paper resulted in more than twice or even more than three times the number of ideas collected with the tool. Regarding RQ3 this shows the influence of using a more appropriate tool for the task on the number of ideas gathered. Because the participants and artefacts evaluated in the current study differ too much from the previous study, the numbers cannot be directly compared and all the changes necessarily attributed to the tool used, but the results can nevertheless be used as an indicator how using a prototype of a dedicated tool influences the number of ideas gathered with the tool-based approach.

There are several explanations for the differences still existing between the number of ideas gathered with the two different approaches. Some of them are of general nature and it might therefore per se not be possible to overcome this shortcomings of a tool compared with paper. But some of them are based on the current state of the tool prototype and the evaluation environment and can thus be overcome in upcoming versions of the tool and future evaluations:

- **Paper being more natural than tool.** With PDotCapturer the participants needed to learn how to use it and get familiar with expressing their PD ideas using it. Paper and how to write and draw on it on the other hand is already familiar to the participants so they can start right away with providing ideas and they might also be quicker providing written comments on paper, e.g. depending on their handwriting compared to their typing speed.

This is a general issue of tool over paper usage, which cannot be resolved. But it can be tried to mitigate it by making the tool even easier to learn and use. Therefore the feedback on the tool and possible issues participants had while using it to express their ideas was taken into account to improve version 3 of the PDotCapturer tool.

- **Tool still being a prototype** Although version 2 of PDotCapturer and thus an improved version of the tool was evaluated, it still being a prototype could have an influence on the User Experience as well as on the ability, speed of, and willingness to use the tool to express PD ideas.

  This issue will automatically be resolved in the future, with the tool becoming more and more mature which should further improve the task and evaluation experience for the participants.

- **Slow Internet connection at the workshop locations.** With the study presented being conducted as part of a summer school for teachers, the technical conditions were not as freely configurable as in more controlled evaluation settings. This led to some issues with slow Internet connection when using the tool to express ideas. It also slightly influenced the paper-based approach as participants looked at the online lessons on their PC and only used the paper to note down their ideas, but might have a bigger impact on the number of ideas gathered with the tool-based approach.

Besides an overall evaluation of the number of ideas, the ideas have also been further analysed based on CAt+ categories and attributes assigned to them, to see if one of the approaches resulted in significantly more ideas in one of the coding dimensions. The null hypothesis (H_0) to be verified by using Chi square tests (Maltby and Day, 2002) was accordingly formulated as follows:

H_0: the number of ideas coded as one of the rating options specified by CAt+ is independent of the approach used to gather them (paper- or tool-based) for each of the possible values and dimensions (categories, impact, specificity, uniqueness).

The results of the Pearson's $\chi 2$ analysis can be seen in Table 6.1. The attributes Impact and Uniqueness could be analysed right away. Specificity was aggregated based on the number of specific sub-ratings for Target, Reasoning, and Solution. All ratings with two or more values of 3 for the sub-ratings were aggregated to 'very specific', all ratings with at least one sub-rating of 3 were aggregated to 'specific', and the remaining comments were aggregated to 'unspecific'. For the Categories the analysis had to be done on the aggregation of sub-categories based on their main category (see first line in categories column of Table 6.1). As this still resulted in two cells (16.7%) with an expected frequency of less than 5, 'Picture' and 'Irrelevant' were combined to 'Other' and the test was repeated with these aggregated categories (see second line of categories column in Table 6.1) to reach expected values above 5, which are advised for the Chi square test (e.g. in Maltby and Day, 2002). The result remained the same, H_0 was rejected.

	Result of Pearson's $\chi 2$ analysis
Categories	$\chi 2$ (5, n=343)=16.254, p<.05* $\chi 2$ (4, n=343)=14.167, p<.05*
Impact	$\chi 2$ (4, n=277)=27.075, p<.001**
Specificity	$\chi 2$ (2, n=277)=1.574, p>.05 (ns)
Uniqueness	$\chi 2$ (1, n=277)=0.926, p>.05 (ns)

Table 6.1: $\chi 2$ values showing if the number of ideas in each CAt+ dimension is independent from the approach used to gather the ideas or not.

As can be seen in Table 6.1, the null hypothesis that the number of ideas is independent of the approach used to record them, is rejected for Categories (significant) and Impact (highly significant). Regarding Specificity and Uniqueness H_0 is not rejected, meaning that paper and tool result in a comparable number of unique and repeated ideas of equal specificity.

Results regarding RQ6 (Quality of Ideas)

Data processing to compare ideas gathered with paper- and tool-based approach

To account for the differences in number of ideas gathered using the two different approaches the data was normalised for the qualitative analysis. This was done by dividing the number of ideas assigned to the individual categories and attribute levels by the number of ideas gathered in the respective workshop and with the respective approach. The following tables show the percentages of the respective sub-totals.

Categories

As shown in Table 6.2, both comparisons show consistent trends for some categories and are inconsistent for others. The numbers of Content and Design ideas gathered are higher when the tool-based approach is used, while the number of Functionality ideas is higher when the paper-based approach is used. The results for the two categories Incomprehensible and Irrelevant are inconclusive, because the trends go in opposite directions.

	Comparison 1		**Comparison 2**	
	Paper	**Tool**	**Paper**	**Tool**
Content	23.53	40.00	27.20	35.54
Design	4.41	24.00	8.00	13.22
Functionality	54.41	20.00	44.00	36.36
Incomprehensible	10.29	16.00	18.40	9.92
Irrelevant	7.35	0.00	2.40	4.96

Table 6.2: Percentage of ideas per main category.

When looking at the distribution of ideas in the three main categories that showed consistent trends (Content, Design, and Functionality) it can be seen that the tool created a somewhat more equal distribution of ideas to categories than paper (see Table 6.2). The difference between the highest and lowest percentages of ideas is less than 3 times for the tool, as compared to more than 12 times for paper.

A possible explanation for the differences in the distribution of Content, Design, and Functionality ideas might lie in the prototype presentation. The non-interactivity of the printouts on paper might have triggered the participants to question how interaction elements work, why they work this way, and how they expect them to work. With the interactive presentation of the prototype in PDotCapturer and the option to express ideas without a switch of the medium, the participants might have been enabled to follow the online lesson more smoothly, therefore focusing more on its content and noticing design issues. Additionally the separation of the online lesson onto different screenshot pages in the evaluation booklet might have caused the participants to focus more on details and separate interaction elements rather than the design and content of the complete page.

As it is not meaningful to do a CAt+ Attributes rating for Incomprehensible and Irrelevant ideas, the subsequent results only include ideas that were coded as Content, Design, or Functionality.

Impact

The fact that Impact level 2 was not present in three of the four datasets coded this time implies that it might be feasible to merge it either with level 1 or 3. Thus the ideas coded as having an impact of 2 were re-visited. As they all affected the prototype on a local level of individual elements without affecting the whole screen, the decision was made that for this datasets Impact 1 and 2 can be merged. When comparing this finding with the results presented earlier, a similar trend can be seen. But the trend in the datasets presented there is not as strong as in the current datasets. Hence, more evidence is needed to decide on retaining or removing this Impact level from the CAt+ coding scheme. Nevertheless, considering the possible merge of Impact level 1 and level 2 when coding PD results with CAt+ in the future can be suggested.

When checking the CAt+ Impact rating results (see Table 6.3) for trends, it can again be seen that the two comparisons show some consistent trends and some inconsistent ones. The number of ideas with an Impact rating of 0 or 3 is higher with the tool whereas the number of ideas with an Impact rating of 1 & 2 is higher when using the paper-based approach. The results for Impact rating of 4 are inconclusive.

	Comparison 1		**Comparison 2**	
	Paper	**Tool**	**Paper**	**Tool**
0	41.07	57.14	24.24	30.10
1 & 2	44.64	19.05	49.49	22.33
3	10.71	23.81	20.20	39.81
4	3.57	0.00	6.06	7.77

Table 6.3: Percentage of ideas per Impact level.

The overall rather high percentage of impact ratings of 0 could be explained by the fact that by actively asking participants about their mood (see ⑥ in Figure 4.6), the tool might elicit more ideas stating generic positive or negative feelings towards screen elements. As such ideas are helpful to get a general impression what participants like or dislike, but do not have any immediate influence on the prototype re-design, they would be rated as having an impact of 0.

The especially high number of 0 Impact ideas for the tool in Comparison 1 might be explained by the motivational issue described earlier. When the participants who initially chose the paper-based approach in Workshop-1 had to use the tool in Workshop-2 they might have been less motivated to express ideas and therefore have produced less well-thought and impactful ideas.

A possible explanation why the paper-based approach gathered more localised (1 & 2) ideas addressing single elements, where the tool-based approach gathered more global (3) ideas addressing the whole page could again be the presentation of the online lesson. Splitting it up as different screenshot pages in the evaluation booklet might have led to the participants focusing more on details rather than the complete page. Although in cases covering multiple printout pages they would also have to scroll in the prototype, the latter might have been perceived as one continues page (with scrolling) rather than separated sections of a page (navigated in between by scrolling). The triggering of ideas targeting several pages might be independent of the approach used as neither of the approaches is giving an overview over several pages at the same time.

Specificity

As there are no significant differences for the qualitative attribute Specificity (see Table 6.1) it can be said that paper- and tool-based approach led to ideas of equal quality regarding the Specificity of ideas. Although not significantly relevant the aggregated results are shown in Table 6.4 to provide a complete picture of all the results.

	Comparison 1		**Comparison 2**	
	Paper	**Tool**	**Paper**	**Tool**
unspecific	1.79	13.64	15.15	4.85
specific	71.43	50.00	45.45	57.28
very specific	26.79	36.36	39.39	37.86

Table 6.4: Percentage of ideas per Specificity level.

The results for the Specificity rating are based on the three sub-ratings. For instance, it is less helpful to have a clear target but unspecific problem description and solution compared to a clear target together with a clear problem description and a well described proposed solution. Therefore three possible levels of Specificity for a comment are differentiated as:

- **unspecific:** All sub-ratings have a value of less than 3.
- **specific:** Only one sub-rating has a value of 3.
- **very specific:** Two or more sub-ratings have a value of 3.

The CAt+ rating results regarding Specificity (Table 6.4) are inconclusive when comparing the paper-based and the tool-based approach.

Uniqueness

The Pearson's $\chi 2$ analysis showed that the uniqueness or repetition of ideas is independent from the method (paper or tool) used to gather them (see Table 6.1). It can thus be concluded that paper and tool usage result in the same ratio of duplicates to unique ideas.

	Comparison 1		**Comparison 2**	
	Paper	**Tool**	**Paper**	**Tool**
0	7.14	0.00	11.11	6.80
1	92.86	100.00	88.89	93.20

Table 6.5: Percentage of ideas per Uniqueness value.

When comparing the CAt+ Uniqueness rating results (Table 6.5) it can be derived that the tool-based approach resulted in fewer duplicates than the paper-based approach. This can partly be explained by the higher number of ideas gathered with the paper-based approach over the tool-based approach. With more ideas also the chance of getting duplicates increases (assuming the pool of possible ideas is limited and even more so for ideas that are obvious).

Another possible explanation for the paper-based approach gathering more duplicates might be that the repetition of similar looking pages on the printouts might have elicited the same ideas several times.

Inter-rater reliability

To evaluate the inter-rater reliability of the two raters who coded the data, weighted Cohen's kappa (Fleiss and Cohen, 1973) was calculated. For the Categories a weight of 2 was applied when main and subcategory differed and a weight of 1 if there was at least agreement on the main category but differing sub-categories assigned. For Impact and the three Specificity sub-ratings the difference between higher and lower value assigned was used as the weight. For Uniqueness the standard weight was applied. The results are shown in Table 6.6.

	Weighted Cohen's kappa
Categories	0.76
Impact	0.85
Specificity-Target	0.60
Specificity-Reasoning	0.70
Specificity-Solution	0.79
Uniqueness	0.79

Table 6.6: Weighted Cohen's kappa to determine inter-rater reliability.

As said before the guidelines on kappa rating magnitudes are not consistent in the literature. But with a result of 0.6 or higher (mostly 0.7 or higher) for each rating dimension, it is reasonable to assume that the inter-rater reliability is good.

Comparison of Tools

Although there is still a noticeable difference in the number of comments gathered with the different approaches, it is a vast improvement over the difference reported in the previous study. This implies that using a more appropriate tool for the task results in a noticeable improvement in terms of the number of comments gathered. As no consistent trend between myBalsamiq and PDotCapturer regarding the CAt+ Category coding can be seen, it can be assumed that switching the tool did not have an influence on the type of comments gathered.

A caveat is that the participants of the study and artefacts evaluated were different from their counterparts in the previous study. Thus there might be factors other than the change of the tool (from myBalsamiq to PDotCapturer) influencing the number of comments gathered. Nevertheless, it is reasonable to argue that deploying a dedicated PD tool can contribute at least partially to the improvements observed.

Although the numbers can also not be compared directly when comparing the qualitative results for myBalsamiq reported in the previous study with those for PDotCapturer reported here, some consistent trends can be seen for the different CAt+ attributes.

Using the dedicated PDotCapturer tool resulted in more comments with an Impact rating of 0. As myBalsamiq, like paper, does not encourage specifying emotions, this difference can be explained in the same way as above. The trends for Impact ratings of 1(& 2) and 3 are much stronger in the datasets presented here. This can be explained by the dedication of the tool emphasizing the differences between paper and tool. As the findings are inconclusive regarding Impact of 4 for both tools, this supports the assumption that the number of ideas with this level might be independent of the approach used to gather them and therefore the reasons for differences here lie beyond the tool used to gather these ideas.

Opposite to the findings reported for myBalsamiq, where the paper consistently resulted in a higher number of very specific Specificity-Target comments, the results for PDotCapturer are inconclusive. The increase in target specification when using PDotCapturer can be explained by the improved way to indicate the position of a comment in PDotCapturer over myBalsamiq.

Regarding the Uniqueness the difference in duplicates between paper and PDot-Capturer is mostly lower than reported for paper and myBalsamiq. This can be explained by the equalization of total number of comments gathered between paper and tool when using PDotCapturer instead of myBalsamiq.

6.2 Paper and Version 3 of PDotCapturer

6.2.1 Participants and Procedures

For the final evaluation of PDotCapturer comparing it with paper at the same time, 38 HCI students were randomly (based on their Student ID) split into four groups to evaluate either a low-fidelity or a high-fidelity prototype of a Quiz Tool using either a paper-based or PDotCapturer-based approach.

To compare the use of PDotCapturer with the use of paper to elicit PD ideas in different stages of the design process the previously described four conditions were created. Two of them use a low-fidelity prototype, representing an early design state and the other two use a high-fidelity prototype, representing a later design state. To keep the results comparable by having comparable prototypes just of different fidelity a hand-drawn sketch drawing was created, based on the high-fidelity prototype used in the second set of conditions. For both prototypes a paper-based approach and PDotCapturer are used to gather ideas. This resulted in the following four conditions:

Condition 1 (n = 9) is a paper-based approach on a low-fidelity prototype. Here the participants got a booklet of photocopied hand-drawn screens of the Quiz Tool, together with a scenario to follow and to pretend to perform using the provided screens. This experience was followed by a questionnaire on the paper-based method they just used to express their ideas (Appendix A.8).

Condition 2 (n = 10) is PDotCapturer on a low-fidelity prototype. Here a scan of the low-fidelity prototype is presented in PDotCapturer together with the same scenario as in the paper-based condition. Afterwards the participants filled in a questionnaire about the features, Usability, and User Experience of PDotCapturer (Appendix A.9).

Condition 3 (n = 9) is a paper-based approach on a high-fidelity prototype. Here the participants interact with the interactive prototype on a PC screen, following the same scenario as the participants in the other conditions. They use a booklet with print-outs of screenshots of the prototype to specify their ideas. Then they answer the same questionnaire as the participants in Condition 1 (Appendix A.8).

Condition 4 (n = 10) is PDotCapturer on a high-fidelity prototype. Here the interactive prototype is presented in PDotCapturer together with the scenario to follow. The participants then interact with the prototype there and use the functionality of the tool to express their ideas. Afterwards they fill in the same questionnaire as the participants in Condition 2 (Appendix A.9).

As can be seen the conditions reflect the conditions from earlier evaluations comparing paper-based and myBalsamiq/PDotCapturer supported PD activities.

The prototype used in this evaluation was a Quiz tool, which allows teachers to create quizzes with a variety of question types (e.g. multiple choice, open text). Figure 6.2 and Figure 6.3 show the two versions of different fidelity used with the two different approaches to gather ideas. The Quiz tool was selected for this evaluation, as it is quite intuitive what you are supposed to do with it and as it has a limited set of functionalities so that it could be covered completely in the time available for the evaluation.

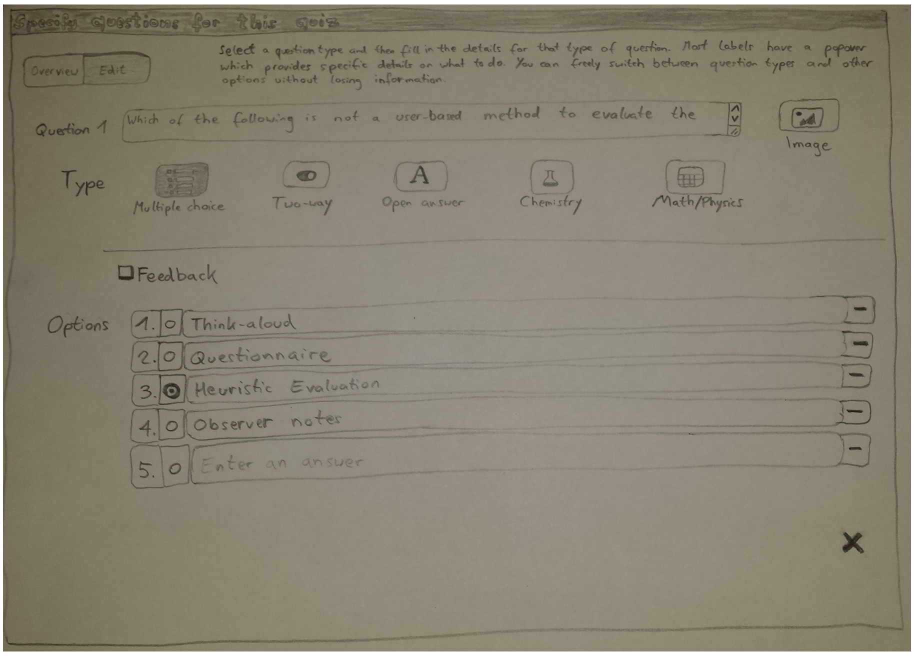


Figure 6.2: Sketch of Quiz tool used as low-fidelity prototype.

To let the participants know what they are expected (to pretend) to do with the Quiz tool, a scenario describes the different steps and interactions associated with creating a multiple choice question using the Quiz tool. The task for the participants

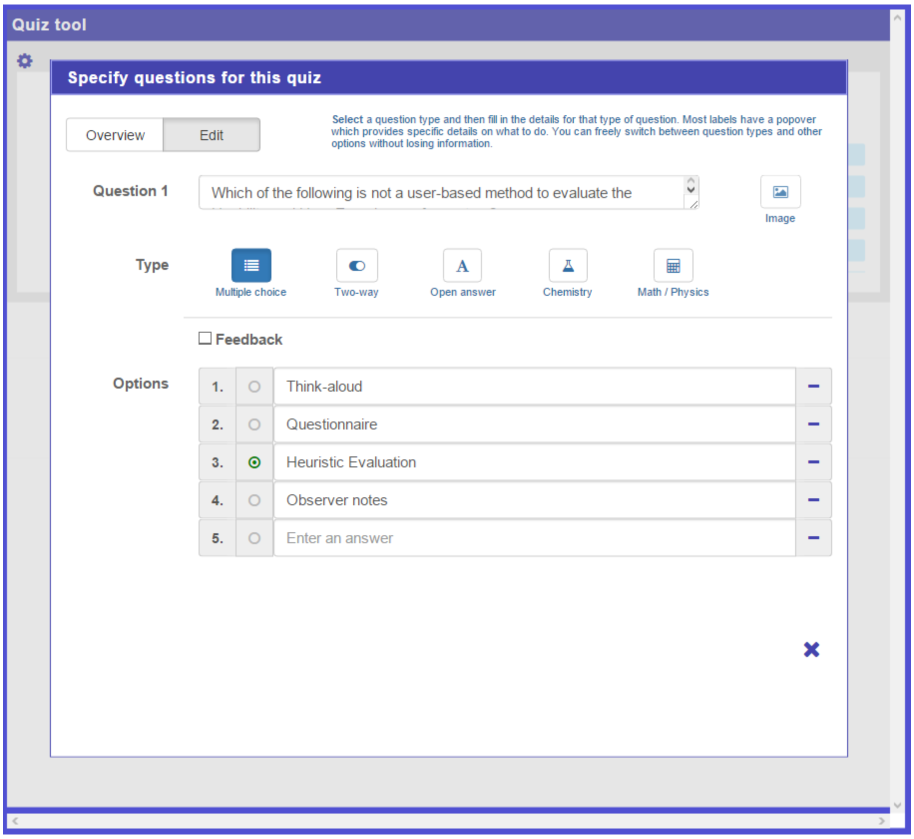


Figure 6.3: Screenshot of Quiz tool used as high-fidelity prototype.

was (to pretend) to perform the steps presented in the scenario and specify their PD ideas while doing so.

The evaluation session started with a presentation of background information. The students then had about an hour to follow the scenario and provide their ideas. Afterwards they were asked to fill in a questionnaire. To compare the two approaches these questionnaires included questions derived from System Usability Scale (SUS, Brooke, 1996). Additionally they contained questions regarding general information.

The PD ideas gathered were afterwards rated using the latest version of the CAt+ coding scheme. Due to time constraints the decision was made to not perform a fully-crossed design. Instead only 20% of the ideas were coded by a second rater to calculate the inter-rater reliability.

Pearson's $\chi 2$ analysis on the results was performed to check the significance of the findings (compare with Section 3.3.3). As CAt+ ratings are categorical data, Chi square test (Maltby and Day, 2002) was applied to verify the null hypothesis (H_0):

H_0: the number of ideas in each of the coding dimensions (categories, impact, specificity, uniqueness) is independent of the method used to elicit and capture them (paper or PDotCapturer).

The results can be seen in Table 6.8.

	Pearson's $\chi 2$ analysis for prototype (low or high)
Categories - Issue	$\chi 2$ (4, n=632)=3.514, p>.05 (ns)
Categories - Solution	$\chi 2$ (4, n=632)=4.423, p>.05 (ns)
Impact	$\chi 2$ (4, n=622)=11.870, p<.05*
Specificity - Target	$\chi 2$ (2, n=622)=7.788, p<.05*
Specificity - Problem/Reasoning	$\chi 2$ (3, n=622)=8.330, p<.05*
Specificity - Solution	$\chi 2$ (3, n=622)=.996, p>.05 (ns)
Uniqueness	$\chi 2$ (1, n=622)=2.997, p>.05 (ns)

Table 6.7: $\chi 2$ values showing if the number of ideas in each CAt+ dimension is independent from the prototype on which the ideas were gathered or not.

To ensure the perceived differences were not caused by the different prototypes, a second null hypothesis (H_0) was tested:

H_0: the number of ideas in each of the coding dimensions (categories, impact, specificity, uniqueness) is independent of the prototype used to elicit and capture them (low- or high-fidelity prototype).

The results can be seen in Table 6.7. The only case were a significant result is retrieved for both null hypotheses is Specificity - Problem / Reasoning. As it is not clear if the significant correlation in this case is caused by the prototype or the approach, the significant result is not going to be taken into account for the approach.

6.2.2 Results

Number of Ideas

Table 6.9 shows the number of ideas gathered with each approach. As the number of participants was slightly different in the four conditions, the number of ideas cannot be compared directly. To account for this the number of ideas is divided by the number of participants for the respective condition. For the low-fidelity

	Pearson's $\chi 2$ analysis for approach (paper or PDot)
Categories - Issue	$\chi 2$ (4, n=632)=3.869, p>.05 (ns)
Categories - Solution	$\chi 2$ (4, n=632)=12.712, p<.05*
Impact	$\chi 2$ (4, n=622)=8.591, p>.05 (ns)
Specificity - Target	$\chi 2$ (2, n=622)=2.134, p>.05 (ns)
Specificity - Problem/Reasoning	$\chi 2$ (3, n=622)=15.906, p<.05*
Specificity - Solution	$\chi 2$ (3, n=622)=27.731, p<.001**
Uniqueness	$\chi 2$ (1, n=622)=5.489, p<.05*

Table 6.8: $\chi 2$ values showing if the number of ideas in each CAt+ dimension is independent from the approach used to gather the ideas or not.

	Paper	**PDotCapturer**
low-fidelity	180 (n = 9) 20.00	187 (n = 10) 18.70
high-fidelity	151 (n = 9) 16.78	139 (n = 10) 13.90

Table 6.9: Number of ideas in total, number of participants in brackets, and mean number of ideas per participant for the four conditions.

prototype paper resulted in about 1.07 times the number of ideas compared with PDotCapturer. For the high-fidelity prototype paper resulted in about 1.21 times the number of ideas compared with PDotCapturer. Both values are lower than the 1.43 times reported for the comparison of version 2 of PDotCapturer with paper.

Although the numbers are not directly comparable because of the different evaluation scenarios (e.g. target of the PD activity), a trend is visible, hinting at the fact that improving the tool also improves the comparability (number-wise) of paper-based and tool-based PD results. With the distance between paper and tool being even less than in the last paper and PDotCapturer comparison, this furthermore supports the assumption that a dedicated tool results in higher comparability with paper than a general tool (over 2 to over 3 times the number gathered on paper compared with myBalsamiq).

For the further analysis 25 of the ideas gathered with PDotCapturer on the high-fidelity prototype had to be removed from the dataset, as these ideas were given on interface element only visible in the live Quiz environment, which were not displayed on the printouts and in the prototype drawings (e.g. header of the page containing the quiz tool).

Categories

Table 6.10 shows the percentage of ideas per main category of the CAt+ coding scheme. Percentages are presented instead of counts to address for the different number of ideas gathered in the different conditions. The inferential statistics performed show that the differences for the categories for solution are significant between paper and PDotCapturer. This means while paper results in significantly more Design and Functionality solutions, PDotCapturer results in significantly more Content and Not needed solutions. Solutions are usually classified as not needed when the issue has been rated as positive. A possible explanation for this is, that PDotCapturer has the mandatory specification of feelings when expressing an idea. This might trigger the expression of positive (which is one of the feeling options) ideas. A possible explanation for paper causing more Functionality ideas being expressed might be the non-interactivity of the paper, triggering the end-user to think about (possible) functionality of interface elements and consequently expressing ideas accordingly.

As in previous data analysis using the CAt+ rating scheme, Incomprehensible and Irrelevant comments have been excluded from further analysis, as it is not meaningful to rate such comments regarding the CAt+ attributes.

	low-fidelity		high-fidelity	
	Paper (n = 180)	**PDot (n = 187)**	**Paper (n = 151)**	**PDot (n = 139)**
Issue				
Content	17.22	13.90	13.25	13.16
Design	46.11	44.92	47.68	42.98
Functionality	35.00	40.11	37.75	41.23
Incomprehensible	1.67	1.07	1.32	0.88
Irrelevant	0.00	0.00	0.00	1.75
Solution				
Content	27.78	28.88	18.54	27.19
Design	30.00	29.95	35.76	29.82
Functionality	19.44	11.76	18.54	7.89
Incomprehensible	2.22	1.6	2.65	3.51
Not needed	20.56	27.81	24.50	31.58

Table 6.10: Percentage of ideas per main category. Due to space restrictions PDot-Capturer is abbreviated to PDot in the table headers.

	low-fidelity		high-fidelity	
	Paper (n = 177)	**PDot (n = 185)**	**Paper (n = 149)**	**PDot (n = 111)**
0	25.99	29.19	26.17	38.74
1	49.72	55.14	56.38	51.35
2	12.99	6.49	8.72	4.5
3	5.08	4.86	7.38	5.41
4	6.21	4.32	1.34	0.00

Table 6.11: Percentage of ideas per Impact value.

Table 6.11 shows the Impact coding results. It looks like paper results in ideas with bigger impact, but this observation is not significant.

	low-fidelity		**high-fidelity**	
	Paper (n = 177)	**PDot (n = 185)**	**Paper (n = 149)**	**PDot (n = 111)**
Target				
3	97.74	97.30	96.64	92.79
2	1.13	2.16	3.36	7.21
1	1.13	0.54	0.00	0.00
0	0.00	0.00	0.00	0.00
Problem / Reasoning				
3	42.94	43.78	26.17	45.05
2	12.99	24.86	17.45	16.22
1	42.37	30.81	55.03	36.04
0	1.69	0.54	1.34	2.70
Solution				
3	42.94	24.32	34.9	30.63
2	12.99	29.73	36.24	13.51
1	42.37	45.41	27.52	54.95
0	1.69	0.54	1.34	0.90

Table 6.12: Percentage of ideas per Specificity value.

Table 6.12 shows the CAt+ coding results for the specificity of ideas for the four conditions. As can be seen from the inferential statistics the differences are highly significant for the solution. As before it is assumed that this result is influenced by the higher number of positive ideas specified using the tool. As a positive idea does not have a clear solution (besides an implicit 'keep it'), those were mostly rated as 1.

	low-fidelity		**high-fidelity**	
	Paper (n = 177)	**PDot (n = 185)**	**Paper (n = 149)**	**PDot (n = 111)**
0	22.60	14.59	26.85	20.72
1	77.40	85.41	73.15	79.28

Table 6.13: Percentage of ideas per Uniqueness value.

Table 6.13 shows the results of the uniqueness coding for each of the conditions. The inferential statistics performed show that PDotCapturer results in significantly more unique ideas than paper. With the similar number of ideas gathered with paper and tool, this can this time not necessarily be explained with a higher probability of repeated ideas when having more ideas. A possible explanation of this might be that several similar looking images on paper caused repeated expression of the same idea. The presentation of the sketches and prototype in PDotCapturer on the other hand might have been perceived as interacting with a tool where content is swapped within the same interface element, not inducing the necessity to provide ideas on the same issue on a different screen again.

Pearson's Chi Squares

The Pearson's Chi squares results regarding independence of the number of ideas in each CAt+ dimension from prototype and approach are presented in Table 6.7 and 6.8 respectively.

6.2.3 Results of Questionnaire

When comparing the results from the four conditions comparing paper and PDotCapturer it can be seen that PDotCapturer is rated better than paper for the low-fidelity prototype conditions in all but one of the statements (Table 6.14). For "I needed to learn a lot of things before I could get going with this method" paper does better than PDotCapturer. This can be explained by the fact that the PDotCapturer tool, opposite to drawing and writing on paper, was not known to the participants before the evaluation.

For the high-fidelity prototype condition the outcome is opposite to the low-fidelity conditions: paper is rated better than PDotCapturer in all statements (Table 6.14). The conjecture that integrating a high-fidelity prototype into PDotCapturer instead of a low-fidelity one somehow makes it more complicated and less user-friendly to use is supported when comparing the same approaches on the different maturities of the prototype. Besides "I needed to learn a lot of things before I could get going with this method" the paper approach on the high-fidelity prototype is rated better (or equal for "I think that I would like to use this method again") than the one on the low-fidelity prototype. PDotCapturer is rated better for the low- than the high-fidelity prototype in all statements.

Statement	**Paper**	**PDot**	**Prototype**
I think that I would like to use this method again	3.33 (1.33)	3.50 (1.50)	**low**
	3.33 (1.15)	2.40 (1.20)	**high**
I found this method/PDotCapturer unnecessarily complex	1.67 (1.25)	1.50 (0.92)	**low**
	1.33 (0.47)	2.20 (1.08)	**high**
I thought this method was easy to use	3.56 (1.17)	4.30 (1.00)	**low**
	4.33 (0.67)	3.90 (0.83)	**high**
I would imagine that most people would learn to use this method very quickly	4.00 (1.33)	4.50 (0.67)	**low**
	4.11 (0.74)	3.50 (1.02)	**high**
I found this method very cumbersome to use	2.56 (0.83)	2.40 (0.92)	**low**
	2.11 (1.20)	3.50 (0.92)	**high**
I felt very confident using this method	3.67 (1.25)	4.30 (0.64)	**low**
	4.22 (0.79)	3.60 (1.28)	**high**
I needed to learn a lot of things before I could get going with this method	1.56 (1.26)	1.90 (0.94)	**low**
	1.67 (0.94)	2.20 (1.40)	**high**
Being able to create freehand drawings was useful to express my opinion	2.89 (1.37)	3.50 (1.43)	**low**
	3.78 (1.13)	3.40 (1.36)	**high**
Being able to provide textual feedback was useful to express my opinion	4.00 (0.94)	4.30 (1.00)	**low**
	4.44 (0.68)	3.90 (0.83)	**high**

Table 6.14: Replies to questions comparing paper and PDotCapturer on a 5-point Likert scale from 1: Strongly Disagree to 5: Strongly agree, with standard deviations in brackets.

6.2.4 Inter-Rater Reliability

To evaluate the comparability of the codings of the two raters, weighted Cohen's kappa (Fleiss and Cohen, 1973) was calculated. The resulting inter-rater reliability for the different criteria is shown in Table 6.15. As in previous studies the weight for the Categories was based on agreement in both, main- and sub-category (weight of 0), agreement in at least the main category (weight of 1), or disagreement in the main category (weight of 2). This weighting was applied as agreement in the main category is more critical (e.g. is it a Design idea or an idea about the Functionality) than the subcategory (e.g. should existing Content be Changed or new Content be Added). For Impact and Specificity it was assumed that the options on the 'scales' are somewhat equally distributed and therefore the distance/difference between the higher and lower value was applied as weight. For Uniqueness the standard weight was used, as there are only two options to choose from.

	Weighted Cohen's kappa
Categories - Issues	0.98
Categories - Solutions	0.78
Impact	0.97
Specificity-Target	0.83
Specificity-Reasoning	0.86
Specificity-Solution	0.89
Uniqueness	0.46

Table 6.15: Weighted Cohen's kappa to determine inter-rater reliability.

As can be seen in Table 6.15 for all ratings but Uniqueness the value of weighted Cohen's kappa is above 0.7 (mostly even above 0.8). Even with the inconsistent kappa rating magnitude guidelines in the literature (Altman, 1990; Fleiss et al., 2003; Landis and Koch, 1977), it can again (cf. earlier studies) be confidently be assumed that these results are reasonable or even good.

The low result for Uniqueness can be explained by the fact that the second rater only coded 20% of the ideas, thus in that dataset some other ideas that would make the coded ones not unique were not present. Consequently the 'unique' rating of the second rater mismatched the 'not unique' rating of the first. For the combined rating the first rater pointed out the previous idea that made the current one not unique and the ideas were coded accordingly.

The second lowest inter-rater reliability was achieved in the Categories rating for solutions. This can be explained by the observation that some of the end-user ideas describe the solution indirectly (e.g. ‘text is too small’ hinting at the solution that it needs to be bigger) or arbitrarily (e.g. the issue of the small text could be addressed by Removing Content [making the remaining text bigger if it still occupies the same space], Changing the Design [making the text layout larger], or by Adding Functionality [e.g. settings or buttons that allow the user to change the font size interactively]). When coding these ideas raters can come to different conclusions. However the still substantial agreement shows that the CAt+ rating scheme is robust enough to account for this.

6.3 Paper and PDotCapturer: Creating a design from Scratch

This section and its sub-sections is based on a full paper presented at BritishHCI 2017. A common PD scenario is creating the design of an interface (or product) from scratch. The end-users are asked to freely come up with ideas and create an initial design on a blank page. To evaluate PDotCapturer in this scenario and compare it with its paper-based counterpart (Layered Elaboration, Walsh et al., 2010) the study described in the following sections was conducted.

6.3.1 Procedures and Participants

28 Informatics students used Layered Elaboration and PDotCapturer in a cross-comparison study to design a sleeping personal informatics tool home screen and personalized dashboard for their university homepage. These applications were selected because they fit the curriculum, the students just had learned about personal informatics applications, and because of the familiarity of the target group with the university homepage. Additionally for both applications it was reasonable to design a single screen, a limitation necessary for the study due to time constraints.

The two-hour session started with a presentation explaining Participatory Design and where it would typically be applied in the design life-cycle when designing an application from scratch. The Layered Elaboration method and PDotCapturer were described as two options for PD activities. Then the first task, to design a sleeping personal informatics tool home screen (showing e.g. sleep last night/this week, alarm time) was presented. The participants were randomly divided into two groups of 14 each (Figure 6.4), starting with the Layered Elaboration approach (green box on top in Figure 6.4) or PDotCapturer tool respectively (yellow box on the bottom in Figure 6.4). The two subgroups were further divided into six teams: four teams of two and two teams of three participants (grey team boxes in Figure 6.4) to let the participants perform the design in small groups as proposed in the Layered Elaboration description. Six teams per approach were created to let three teams each discuss and exchange their designs. The participants then had 15 minutes to come up with an initial design, and specify it, either on paper or on a blank white slate in PDotCapturer. Then they explained their design to the other two groups and exchanged designs. They had then 10 minutes time to annotate the initial design, either on acetates or on a virtual layer in PDotCapturer. The process is displayed in Figure 6.5 and Figure 6.6. Example results can be seen in Figure 6.7 and 6.8.

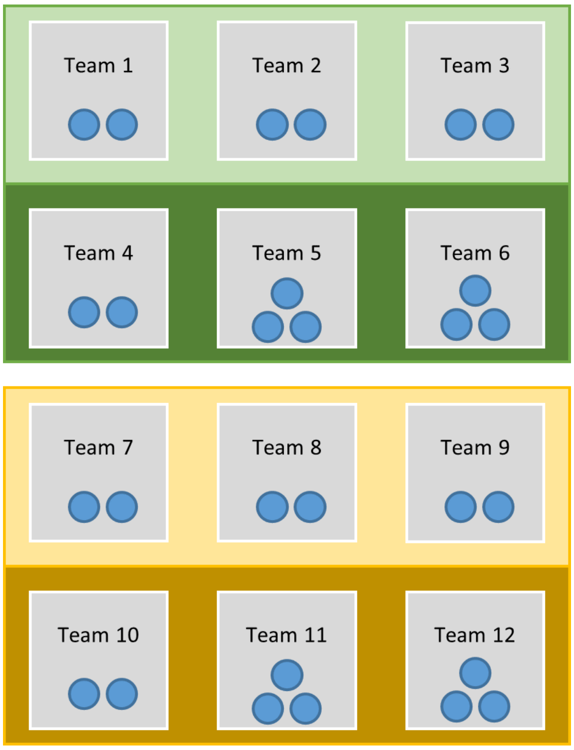


Figure 6.4: Division of participants.

To avoid order effects the participants were then shown an animation on the Anti-Counterfeiting Trade Agreement (ACTA, `https://www.youtube.com/watch?v=glR5Kf6GvpY&list=PL5th9f1P3KP-5pVlCSIqYeE3F9iRkklsT&index=5`, last accessed 24/03/2017). This video was unrelated to either one of the two design tasks, and a following short discussion on it was used to distract the participants. After this the participants swapped the PD method, the ones who used paper before now used PDotCapturer and vice versa. The task this time was to design a university

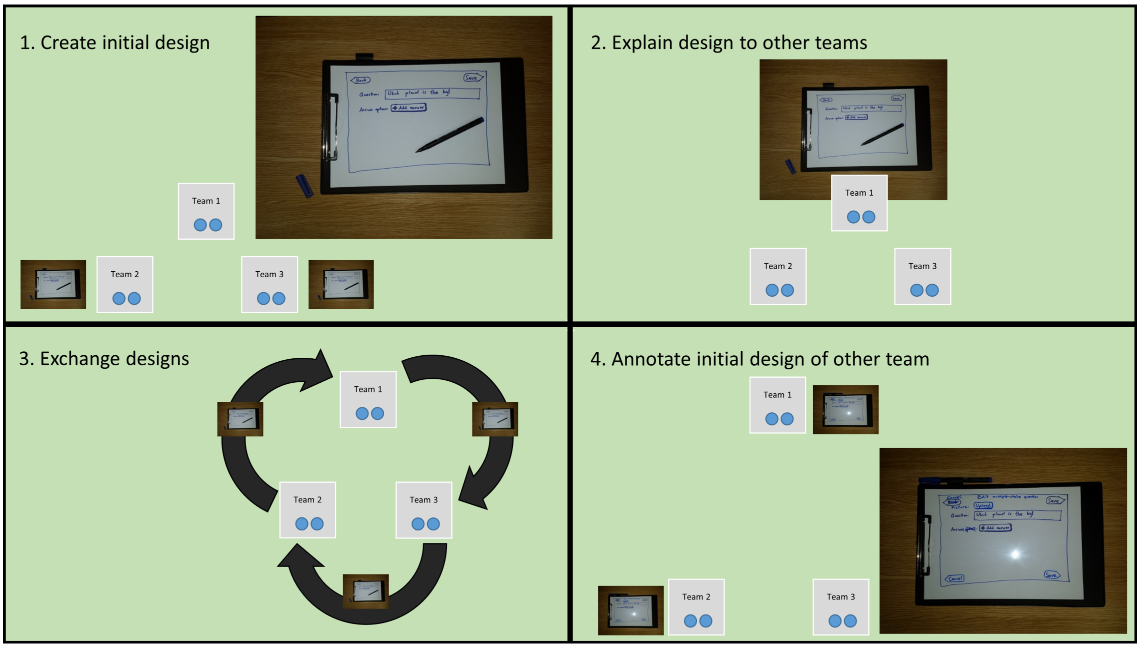


Figure 6.5: Layered Elaboration approach as performed with paper and acetates (the same process was followed by Team 4, Team 5 and Team 6).

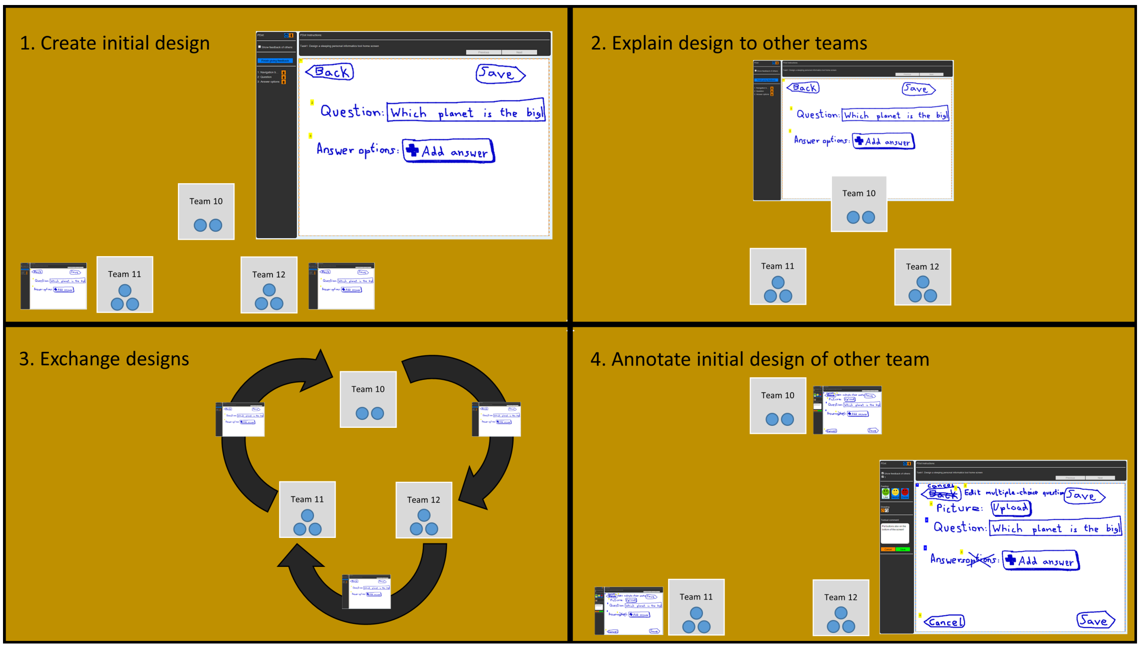


Figure 6.6: Approach as performed with PDotCapturer (the same process was followed by Team 7, Team 8 and Team 9).

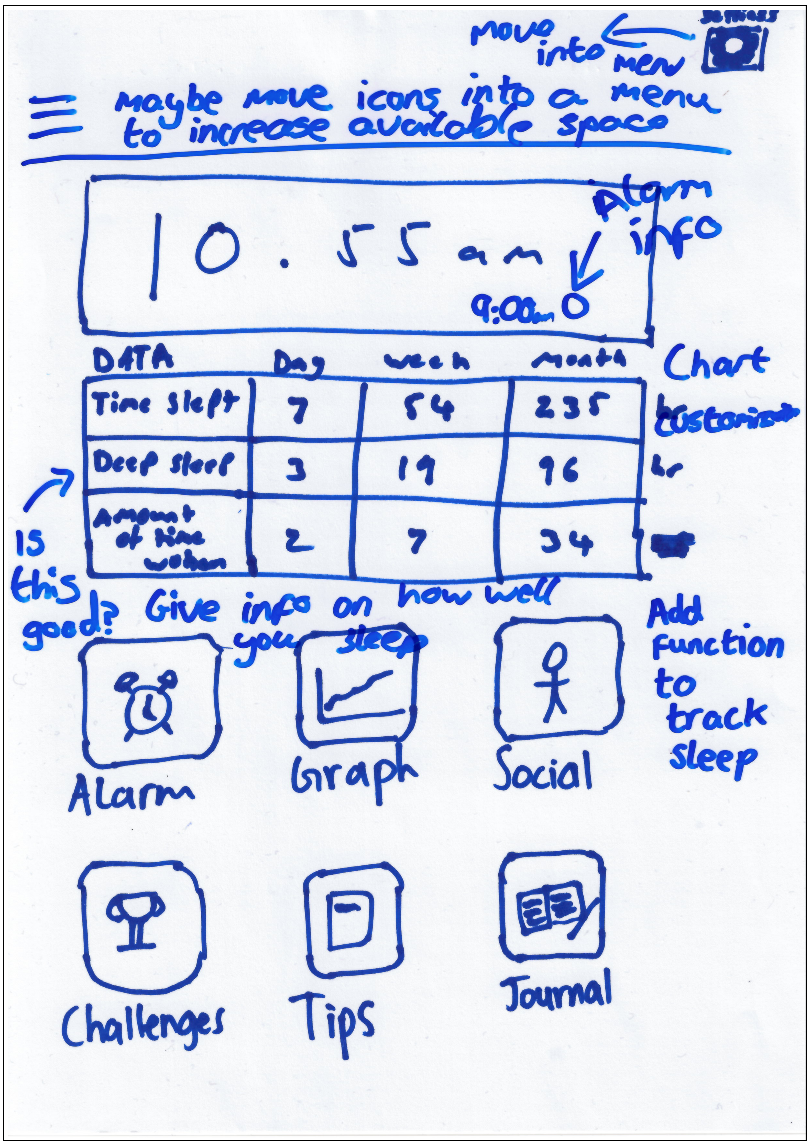


Figure 6.7: Example of sleeping informatics tool home screen designed with the Layered Elaboration approach.

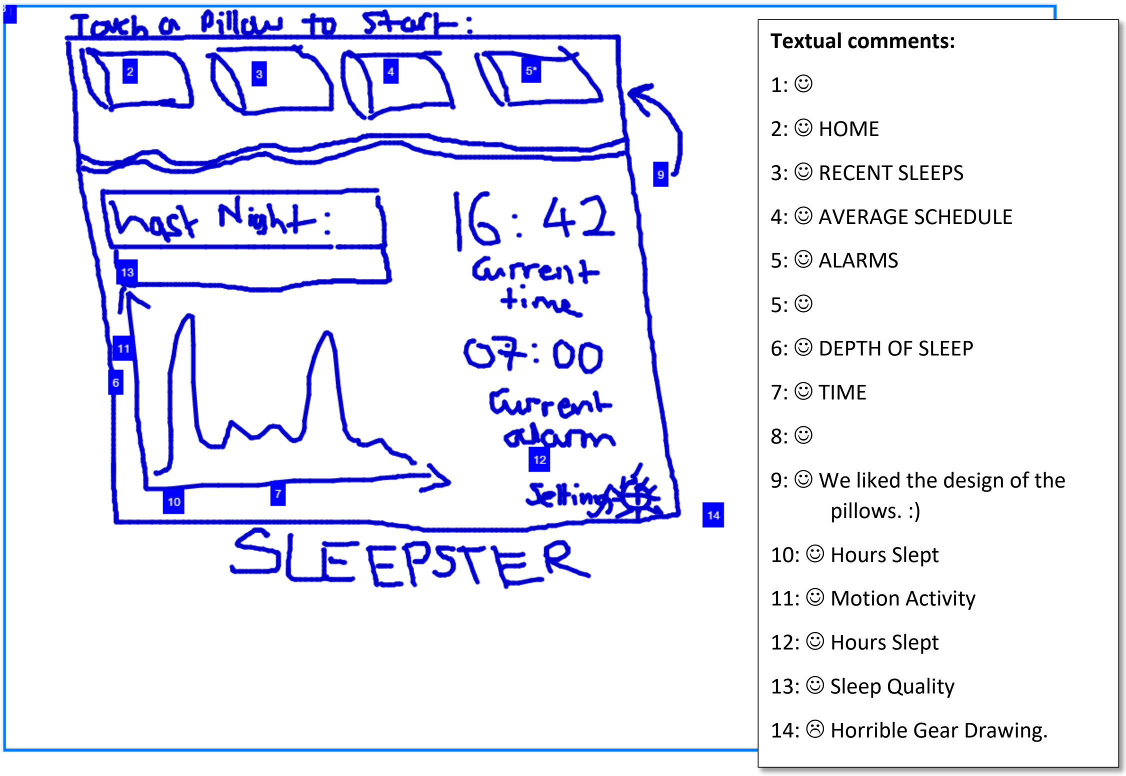


Figure 6.8: Example of sleeping informatics tool home screen designed with PDot-Capturer.

customizable dashboard (showing e.g. announcements, appointments, timetable, overview of emails). The following procedure of creating an initial design, explaining, exchanging, and annotating it were the same as for the sleeping personal informatics tool. Example results can be seen in Figure 6.9 and Figure 6.10. In total eleven designs created with the Layered Elaboration approach and twelve designs created with PDotCapturer were handed in by the participants.

After using both methods the participants filled in a questionnaire (Appendix A.10). It consisted of a section capturing demographic data (e.g. age and gender) and a section comparing the paper-based approach and PDotCapturer (inspired by SUS, Brooke, 1996) and asking for preference of one or the other regarding different PD idea expression tasks (see Figure 6.11 for an example and Table 6.23 for a list of the statements) followed by a text field to elaborate on why and for what the participants preferred which method.

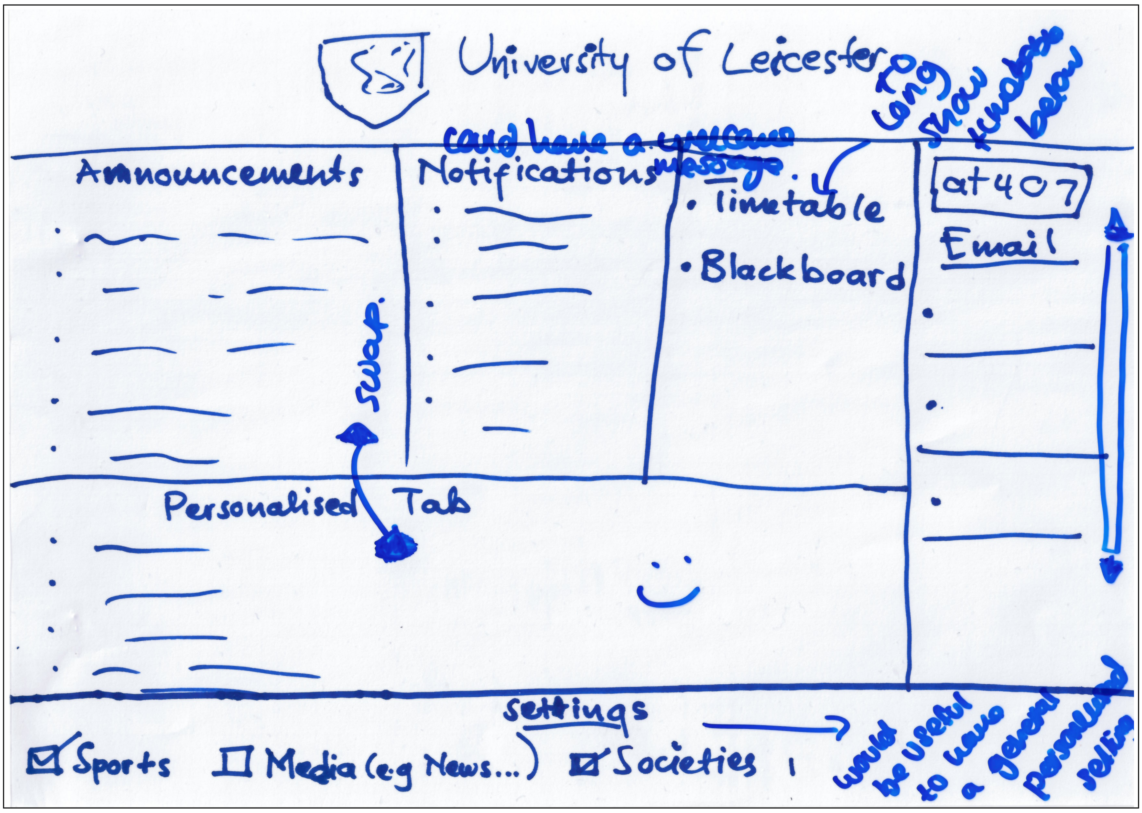


Figure 6.9: Example of the university customizable dashboard designed with the Layered Elaboration approach.

6.3.2 Data Analysis

To compare the designs created on paper and with PDotCapturer, they were rated by two HCI specialists regarding three qualities: aesthetics, usability, and relevancy. They first did the rating individually and then met to discuss the discrepancies.

Aesthetics

To rate the aesthetics of the designs, rating scales for classical and expressive aesthetics were used, developed by Porat and Tractinsky (2012). For classical aesthetics the qualities ‘clean’, ‘pleasant’, ‘symmetrical’ and ‘aesthetic’ are rated on a 7-point scale ranging from strongly disagree (1) to strongly agree (7). For expressive aesthetics the qualities ‘original’, ‘sophisticated’, ‘spectacular’, and ‘creative’ are rated using the same 7-point scale. An example Aesthetics rating for one of the designs can be seen in Figure 6.12.

To make sure the results were not influenced by the two different topics of the design task the ratings were tested for independence of the design task. A Shapiro-Wilk test

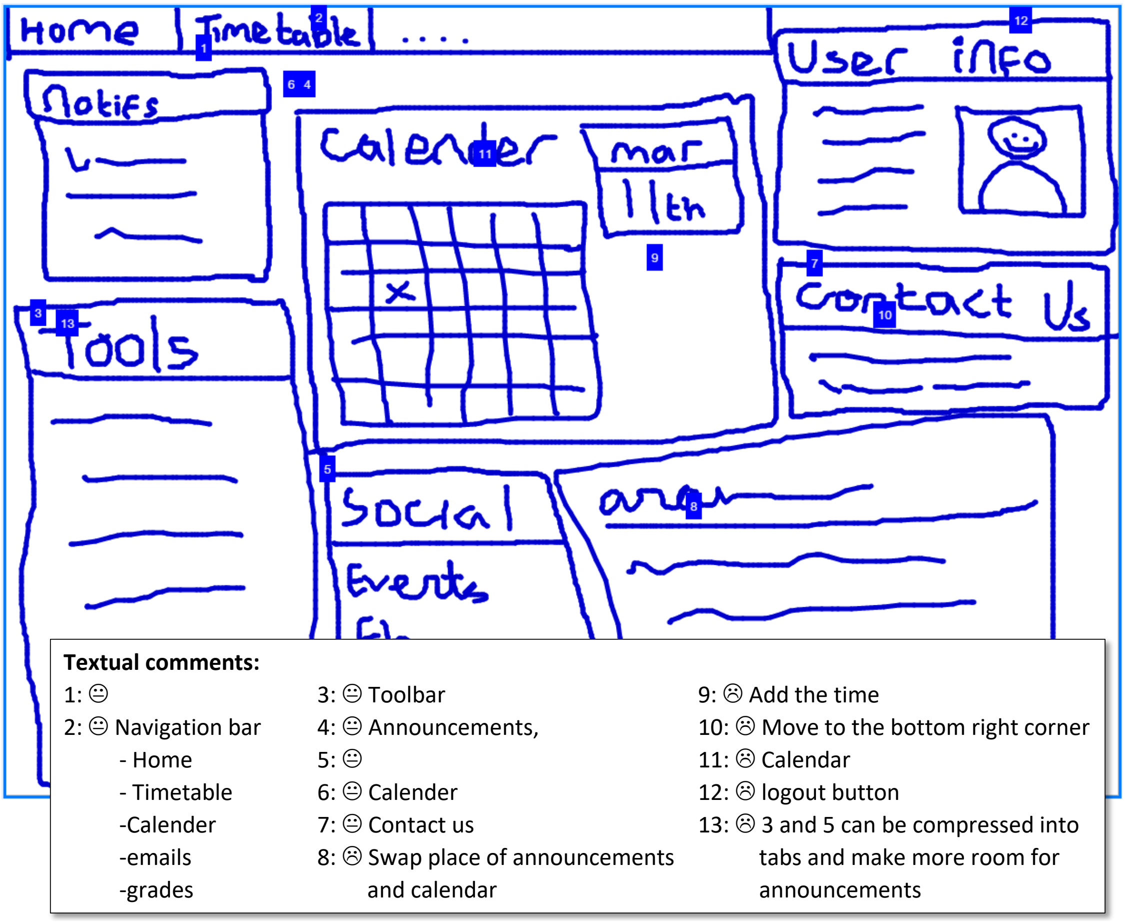
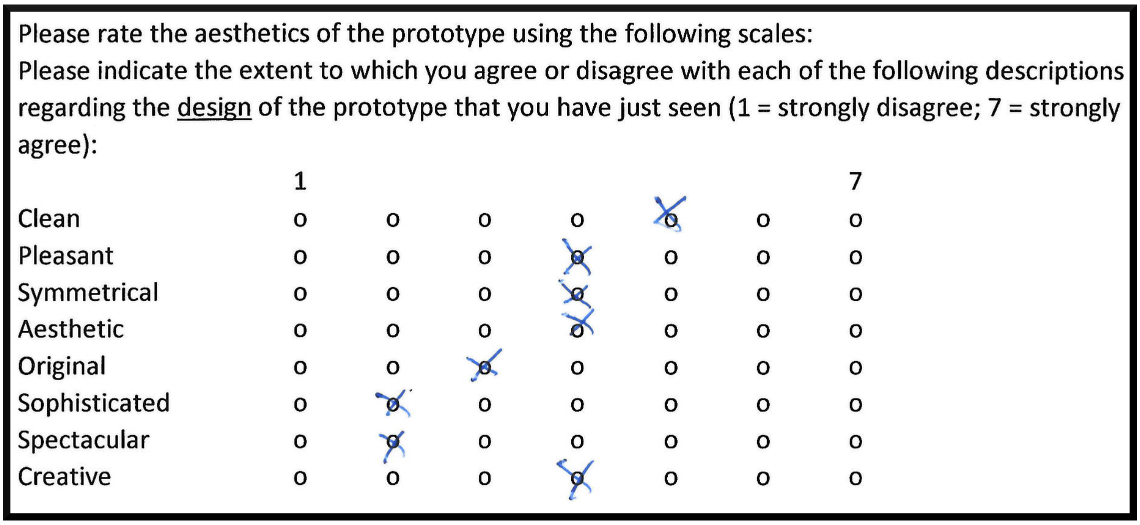


Figure 6.10: Example of the university customizable dashboard designed with PDot-Capturer.

Figure 6.11: Example statement from the questionnaire comparing paper with PDot-Capturer.


Please rate the aesthetics of the prototype using the following scales:
Please indicate the extent to which you agree or disagree with each of the following descriptions regarding the design of the prototype that you have just seen (1 = strongly disagree; 7 = strongly agree):

	1						7
Clean	o	o	o	o	X	o	o
Pleasant	o	o	o	X	o	o	o
Symmetrical	o	o	o	X	o	o	o
Aesthetic	o	o	o	X	o	o	o
Original	o	o	X	o	o	o	o
Sophisticated	o	X	o	o	o	o	o
Spectacular	o	X	o	o	o	o	o
Creative	o	o	o	X	o	o	o

Figure 6.12: Example Aesthetics rating for one of the designs.

Quality	Mean rank University dashboard	Mean rank Sleeping tool	
Clean	12.23	11.79	U = 63.5, p = .869
Pleasant	10.95	12.96	U = 54.5, p = .449
Symmetrical	14.00	10.17	U = 44.0, p = .145
Aesthetic	9.36	14.42	U = 37.0, p = .054
Original	10.00	13.83	U = 44.0, p = .151
Sophisticated	11.73	12.25	U = 63.0, p = .831
Spectacular	10.18	13.67	U = 46.0, p = .144
Creative	9.77	14.04	U = 41.5, p = .108

Table 6.16: Mann-Whitney U test results for differences in aesthetics qualities between the two design tasks.

for normality showed that the data were not normally distributed for the different rating categories and tasks. Therefore a Mann-Whitney U test was performed. Results showed that no significant differences in aesthetic ratings between the two design tasks were detected (Table 6.16). They suggested that the aesthetics ratings were independent of the type of prototype designed.

Usability

For the usability rating Nielsen's 10 Usability Heuristics were applied to the designs (Nielsen, 1994a). For each heuristic the raters specified if it was violated or not. In case the heuristic could not be checked based on the prototype the raters could specify 'not applicable' for this heuristic regarding the prototype. An example Usability rating for one of the designs can be seen in Figure 6.13.

To test for independence of the usability ratings from the design task, a Chi square test was performed. This test was selected as there are two categorical variables to be compared: heuristic violation (yes, no, not applicable) and task (university dashboard or sleeping tool home screen). Three of the heuristic ratings could not be analysed, as they were rated as 'not applicable' for all designs: "Error prevention", "Flexibility and efficiency of use", and "Help users recognize, diagnose, and recover from errors" (see Section 6.3.3 below for details). For the other heuristic no significant relation between rating and task performed was found (Table 6.17).

Heuristic	**Result of Pearson's $\chi 2$ analysis**
Visibility of system status	$\chi 2$ (1, n=23)=2.561, p>.05 (ns)
Match between system and the real world	$\chi 2$ (1, n=23)=0.290, p>.05 (ns)
User control and freedom	$\chi 2$ (1, n=23)=0.048, p>.05 (ns)
Consistency and standards	$\chi 2$ (1, n=23)=0.434, p>.05 (ns)
Error prevention	N/A
Recognition rather than recall	$\chi 2$ (2, n=23)=1.345, p>.05 (ns)
Flexibility and efficiency of use	N/A
Aesthetic and minimalist design	$\chi 2$ (1, n=23)=1.495, p>.05 (ns)
Help users recognize, diagnose, and recover from errors	N/A
Help and documentation	$\chi 2$ (1, n=23)=1.155, p>.05 (ns)

Table 6.17: $\chi 2$ values showing if the usability rating is independent from the design task or not.

Please rate the usability of the prototype by indicating which of the following 10 Usability Heuristics (if any) has been violated:

Heuristic	Violated	Not violated	N/A
Visibility of system status The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.		✓	
Match between system and the real world The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.		✓	
User control and freedom Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.		✓	
Consistency and standards Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.		✓	
Error prevention Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.			✓
Recognition rather than recall Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.		✓	
Flexibility and efficiency of use Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.			✓
Aesthetic and minimalist design Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	✓		
Help users recognize, diagnose, and recover from errors Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.			✓
Help and documentation Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.	✓		

Figure 6.13: Example Usability rating for one of the designs.

Relevancy

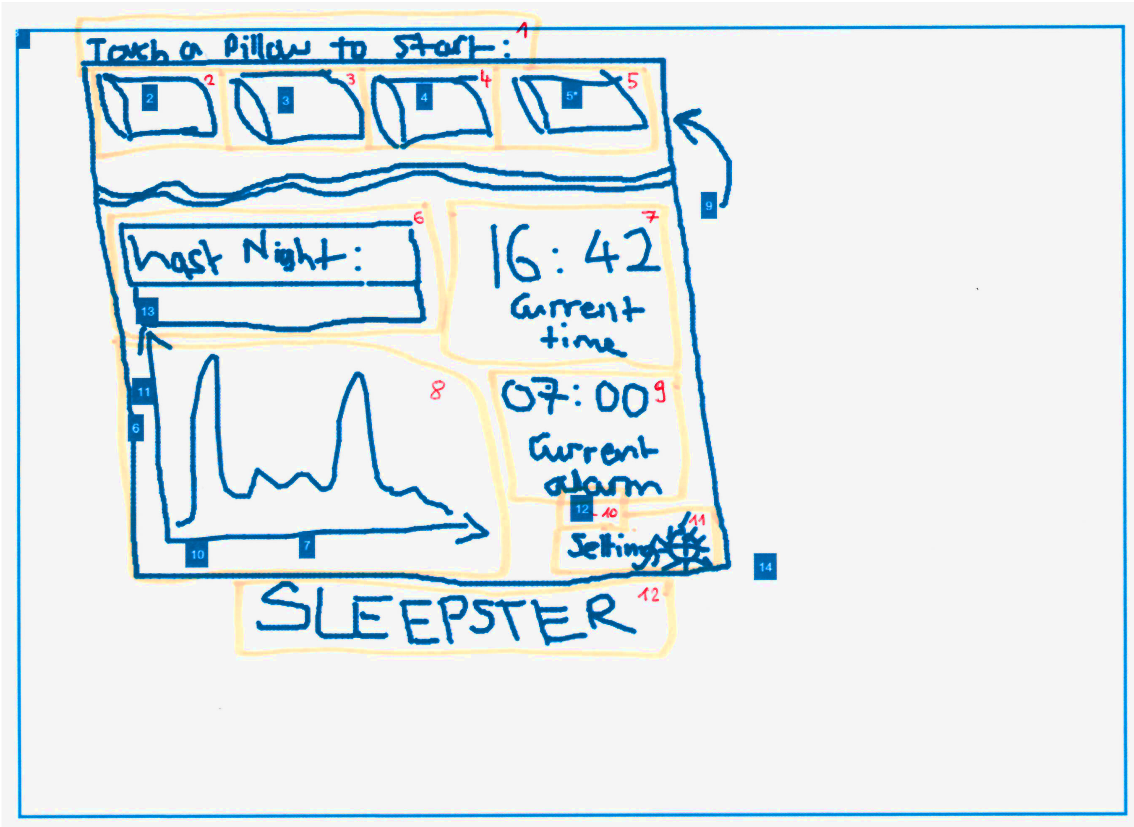


Figure 6.14: Example of highlighting the twelve design elements identified in the design presented in Figure 6.8 through orange lines (and red numbers to identify the elements for the rating table, see Figure 6.15).

To rate the relevancy of the design elements created by the participants the experimenter identified the different elements in each design and indicated them by an orange line separating them from the rest of the design (Figure 6.14). For each of these elements the raters then specified if it was relevant to the task (i.e. one would expect to find such an element in the prototype to be designed) or not. An example Relevancy rating for one of the designs can be seen in Figure 6.15.

To test for independence of the number of relevant screen elements from the task again a Chi square test was used, comparing number of relevant and not relevant ratings for the university dashboard and sleeping tool. The result is that the relevancy of design elements (relevant or not relevant) is independent from the task ($\chi 2$ (1, n=239)=0.655, $p>.05$ (ns)).

No.	Relevant	Not relevant
1		✓
2	✓	
3	✓	
4		✓
5	✓	
6	✓	
7	✓	
8	✓	
9	✓	
10	✓	
11	✓	
12	✓	

Figure 6.15: Example Relevancy rating for the design in Figure 6.14.

6.3.3 Results

Aesthetics

Quality	Layered Elaboration		PDotCapturer	
	Mean	SD	Mean	SD
Clean	4.27	0.96	4.92	0.64
Pleasant	3.64	0.88	4.42	0.76
Symmetrical	3.82	0.94	4.33	1.03
Aesthetic	3.27	0.75	3.75	0.83
Original	2.64	0.64	2.92	0.95
Sophisticated	2.45	0.66	2.25	0.60
Spectacular	2.00	0.60	1.83	0.55
Creative	2.73	0.75	2.42	0.86

Table 6.18: Results of the Aesthetic ratings for paper and tool (mean and standard deviation (SD) on a 7-point scale).

Table 6.18 shows the results of the Aesthetic rating. To compare the two approaches applied, the same inferential statistics described above for the task were applied, only this time for the approach used (paper-based or PDotCapturer). Only one item showed a significant statistical difference between paper-based approach and PDotCapturer: "Pleasant" with $U=34.00$, $p<0.05$, paper-based mean rank=9.09, PDotCapturer mean rank=14.67. Thus the designs created with PDotCapturer (mean=4.42) were rated as more pleasant than the designs created with the paper-based approach (mean=3.64).

However, as no other results showed a significant difference between the approaches (Table 6.19) it suggested that Layered Elaboration and PDotCapturer could be used interchangeably to collect PD ideas. The approach used only slightly (in one out of eight aesthetics aspects) influenced the design outcome regarding aesthetics.

Although mostly not statistically significant, two observations can be made when comparing the results for paper and tool in Table 6.18. On the one hand, the classical aesthetics qualities (all means above 3) have been rated higher than the expressive aesthetics qualities (all means below 3) for all designs. On the other hand, PDotCapturer has been rated higher than paper in the classical aesthetic qualities ('clean', 'pleasant', 'symmetrical', 'aesthetic') while Layered Elaboration has mostly been rated higher than the tool in the expressive aesthetic qualities (['original', being the exception] 'sophisticated', 'spectacular', 'creative').

Quality	**Mean rank Layered Elaboration**	**Mean rank PDotCapturer**	
Clean	9.77	14.04	U = 41.5, p = .106
Pleasant	9.09	14.67	U = 34.0, p = .035*
Symmetrical	10.18	13.67	U = 46.0, p = .185
Aesthetic	10.18	13.67	U = 46.0, p = .185
Original	10.77	13.13	U = 52.5, p = .378
Sophisticated	12.68	11.38	U = 58.5, p = .593
Spectacular	12.82	11.25	U = 57.0, p = .511
Creative	13.00	11.08	U = 55.0, p = .470

Table 6.19: Mann-Whitney U test results for differences in aesthetics qualities between the two approaches. *p<.05.

Usability

None of the usability heuristic rating results (Table 6.20) showed a significantly difference between the paper-based approach and PDotCapturer in the Chi square tests performed (Table 6.21). However, as described above three heuristics were rated as 'not applicable' for all prototypes. This can be explained by looking more closely at the tasks and prototypes designed. Because of the time restrictions for

Heuristic	**Layered Elaboration**			**PDotCapturer**		
	NV	**V**	**N/A**	**NV**	**V**	**N/A**
Visibility of system status	8	3	0	7	5	0
Match between system and the real world	10	1	0	10	2	0
User control and freedom	6	5	0	6	6	0
Consistency and standards	8	3	0	5	7	0
Error prevention	0	0	11	0	0	12
Recognition rather than recall	9	2	0	10	1	1
Flexibility and efficiency of use	0	0	11	0	0	12
Aesthetic and minimalist design	3	8	0	4	8	0
Help users recognize, diagnose, and recover from errors	0	0	11	0	0	12
Help and documentation	4	7	0	2	10	0

Table 6.20: Results of the Usability ratings for paper and tool (NV = not violated, V = violated, N/A = not applicable).

the activity the participants only designed one-page (of their) applications. It is therefore reasonable that the heuristic "Help users recognize, diagnose, and recover from errors" could not be rated for the prototypes. They were neither interactive (so that an error message could have been evoked) nor did the participants design interaction flows that would have contained error messages. In the same way the designs did not include confirmation dialogues, making the "Error prevention" heuristic not applicable. Accelerators, like keyboard shortcuts, were invisible in the (non-interactive) design. Thus, again, it made sense that this heuristic could not be rated and was consequently given the 'not applicable' rating. As the observation of these three heuristic being inapplicable was consistent over all created designs it could be said that this was not caused by the task or approach used, but rather by the general task of designing a (non-interactive) one-page application.

Relevancy

The results of the Relevancy ratings is shown in Table 6.22. With the paper-based approach 128 design elements were identified, 45 of which were rated as not relevant for the respective kind of prototype. With PDotCapturer 111 design elements were created by the participants, 27 of which were rated as not relevant. Therefore paper resulted in an average of 7.55 relevant design elements and PDotCapturer in an average of 7 relevant design elements per design. As for the test for indepen-

Heuristic	**Result of Pearson's $\chi 2$ analysis**
Visibility of system status	$\chi 2$ (1, n=23)=0.524, p>.05 (ns)
Match between system and the real world	$\chi 2$ (1, n=23)=0.290, p>.05 (ns)
User control and freedom	$\chi 2$ (1, n=23)=0.048, p>.05 (ns)
Consistency and standards	$\chi 2$ (1, n=23)=2.253, p>.05 (ns)
Error prevention	N/A
Recognition rather than recall	$\chi 2$ (2, n=23)=1.345, p>.05 (ns)
Flexibility and efficiency of use	N/A
Aesthetic and minimalist design	$\chi 2$ (1, n=23)=0.100, p>.05 (ns)
Help users recognize, diagnose, and recover from errors	N/A
Help and documentation	$\chi 2$ (1, n=23)=1.155, p>.05 (ns)

Table 6.21: $\chi 2$ values showing if the usability rating is independent from the approach or not.

dence of task and relevance of screen elements, a Chi square test was performed. The result shows that the number of relevant to irrelevant design elements is not significantly different between the paper-based approach and PDotCapturer ($\chi 2$ (1, n=239)=3.314, p>.05 (ns)).

Relevancy	**Layered Elaboration**	**PDotCapturer**
Not relevant	45	27
Relevant	83	84

Table 6.22: Results of the Relevancy rating for paper and tool.

Questionnaire

To get subjective ratings of the two approaches from the participants, in addition to the comparison based on the data gathered from the designs and presented above, the participants were asked to express their opinion on the approaches and their comparison in a questionnaire (an example statement to compare Paper and PDotCapturer is shown in Figure 6.11). The results are presented in Table 6.23. For the data analysis the five points on the scale between paper and PDotCapturer were given the values 1 (for the point next to Paper) to 5 (for the point next to PDotCapturer). The point in the middle between paper and tool was thus assigned the value 3. Thus all values of less than three indicated favour of paper (the further away from three, and thus closer to 1, the more), where all values larger than three indicated a preference of PDotCapturer over the Layered Elaboration approach (the bigger, and thus closer to 5, the more).

Statement	Mean	SD
Overall I preferred this method	2.61	1.61
This method was easier to learn	2.25	1.55
This method was easier to use	2.46	1.48
This method was less cumbersome	2.39	1.18
I felt more confident using this method	2.32	1.34
I preferred this method for creating textual feedback	3.32	1.60
I preferred this method to create freehand drawings	1.82	1.28
I preferred this method when expressing my mood	2.96	1.40
I preferred this method to check out the design of others	3.39	1.52
I preferred this method to respond to the design of others	3.32	1.31
I think this method produced the better result	2.39	1.32

Table 6.23: Replies regarding the statements comparing the paper-based approach and PDotCapturer on a scale from 1=paper to 5=PDotCapturer (i.e. a mean of 3 is neutral between paper and tool).

When analysing the comparison of paper and PDotCapturer by the participants presented in Table 6.23 it could be seen that PDotCapturer was rated slightly better than paper for three of the statements (mean >3). The first one was providing textual feedback. The participants preferred typing to handwriting on paper and acetates. This is also supported by the comments specified in the questionnaire, e.g. "Faster and easier to type on pdot and it's condensed into sticky note." The other two statements were the ones based on the PD tasks PDotCapturer was initially designed for (check out an existing design and annotate it). Again, this result from the data analysis is also supported through the comments provided in the questionnaire regarding why and for what which method was preferred: "Evaluating was easier with PDot with the contextual sticky notes." and "Really quick when giving feedback and like, dislike, neutral emotions are useful". Whereas the tool was preferred for textual feedback, non-verbal feedback such as drawings and sketches was clearly preferred to be given using the paper-based approach. This could be explained by the observation that drawing with a mouse on the screen is not common. This was also expressed by the participants in the freeform text fields regarding why they preferred one of the methods to the other: "easier to draw by pen/pencil" or "Hard to draw on computer". All other ratings were between 2 and 3 and therefore in (slight) favour of paper to tool.

To summarize the section on the comparison of PDotCapturer with paper-based approaches: Different empirical studies have been performed evaluating PDotCapturer and how it compares to its paper-based counterparts. With maturing versions of the tool, the difference in numbers of ideas collected using paper or tool converge. Although the results are not directly comparable, because elicited from different studies, these numbers and the other findings indicate, that a user-friendly and customised tool can elicit similar results as paper-based approaches. End-users preferred PDotCapturer over paper for PD tasks it was initially designed for, but in general they mostly preferred paper.

Chapter 7

General Discussion

This chapter presents specific implications regarding the six research questions and general implications derived from the findings presented in this thesis.

7.1 Implications Regarding the Research Questions

RQ1: *To what extent can software tools* ***replace*** *current paper-based participatory design approaches in* ***gathering*** *ideas from end-users?*

It was shown that an appropriate and dedicated software tool could elicit comparable quantity and quality of ideas in participatory design. This is in line with similar findings in research areas related to PD, for example for usability testing of websites (Walker et al., 2002). However, in the process the surprising result was found that end-users preferred the PDotCapturer tool for low-fidelity prototypes but a paper-based approach for high-fidelity prototypes. The expectation had been that the tool would be particularly suited for the evaluation of an interactive prototype (high-fidelity). With the prototype being integrated in the idea elicitation environment, the end-user is enabled to directly express ideas while interacting, without a focus shift from the screen to the paper and back. A possible explanation for this surprising finding might be that the participants are occupied with getting a grasp of the scrutinized prototype if it is interactive and then prefer a method to provide their ideas, which they do not have to learn as well. This assumption is supported by the (comparably) high rating results for the items 'I found this PDotCapturer unnecessarily complex', 'I found this method very cumbersome to use', and 'I needed to learn a lot of things before I could get going with this method' in the questionnaire for PDotCapturer used on the high-fidelity prototype.

Another possible explanation for this finding might lie in the different affordances of paper and tool. "The physical properties of paper ... afford many different human actions, such as ... manipulating ... and in combination with a marking tool, writing on." (Sellen and Harper, 2003, p. 17). A PC on the other hand has different affordances from paper: "... because it dynamically displays information, it affords the viewing of moving images" (Sellen and Harper, 2003, p. 18), or in our case, interactive prototypes. This could explain why the PC was preferred to display the interactive prototype whereas the paper was preferred to express ideas by writing (and scribbling) on. Having a non-interactive prototype might have shifted the 'affordance of viewing moving images' from the prototype to the interactive tool.

With the keyboard, digital technology "... affords the creation of regular, geometric, uniform marks" (Sellen and Harper, 2003, p. 18), or 'text'. Together with the 'affordance of viewing moving images' this could explain the end-user preference for PDotCapturer, as a tool using digital technology, for the PD task of reviewing the design of others and providing textual input.

The findings presented show that for low-fidelity prototypes PDotCapturer can replace similar paper-based approaches. This allows PD practitioners to user a tool over paper and benefit from the features of the tool (for example undo and redo features for end-users and analysis support for designers). For PD activities with high-fidelity prototypes implications regarding the tool-usage can be drawn. To avoid possible confusion between the prototype under scrutiny and PDotCapturer, the instructions on how to use the tool could be altered. Currently the end-users are asked to express their ideas as soon as they occur. This leads to an interruption of the user's interaction with the prototype through an interaction with PDotCapturer. By asking the end-users to interact with the prototype first and later express their ideas, a clearer separation between prototype and tool could be achieved. This change would need to be evaluated though, as not expressing ideas immediately might result in a drop of quantity (they might be forgotten) or even quality of ideas.

As end-users preferred paper over tool for creating their designs it seems advisable for HCI specialists to use a paper-based approach for the initial design task. For the annotation task of low-fidelity prototypes PDotCapturer was preferred. To use the tool for this purpose the initial designs, being drawings and thus low-fidelity prototypes, could be digitalized, for example by taking a picture using a digital camera or mobile phone. They could then be presented in PDotCapturer for the following rounds of PD activities to further refine the design. PDotCapturer could be enhanced to facilitate and ease this process, e.g. a mobile app could be developed to take a picture of a design on paper and integrate it into the tool. A scenario for this in practice could be applying the different approaches over several workshops: an initial design workshop with paper followed by one or more refinement workshops using PDotCapturer. This would not only allow the end-users to work with their preferred method for each step of the PD design and refinement activities, but also support designers through analysis support features in the PDotAnalyser tool. Like in document management, with paper and digital documents, where "introduction of a new technology can stimulate a synergy between old and new" (Liu and Stork, 2000, p. 97) PDotCapturer and PDotAnalyser as new tools can induce new ways to work with and handle paper.

RQ2: *To what extent can software tools* ***replace*** *current paper-based participatory design approaches in* ***analysing*** *the ideas of end-users by designer?*

An evaluation study with designers showed that they would benefit from a tool supporting their analysis task, including the application of CAt+ rating for evaluating end-user ideas. PDotAnalyser is therefore a good starting point, but in addition to the available aggregation and filtering options more automation of the data analysis is desired. Tool support for designers in PD activities in the future should therefore provide further automation to utilise the full potential of software. A tool "opens up new research opportunities with the field of Natural Language Processing and Machine Learning to develop visualization techniques and automated organization of the ideas" (Walsh and Foss, 2015, p. 106). Given the strong user-centred focus of PD, such technology-oriented research is traditionally not part of the PD research area. Optimizing the algorithms to automatically analyse the PD ideas gathered would therefore require the expertise and knowledge of researchers from data-driven research areas.

Besides using algorithms to improve the data analysis process, a change of the data retrieval process can also facilitate the data analysis task. With a more structured retrieval of information from the end-users, it becomes easier to make sense of, categorise, and sort the data later on. The open paper-based approach of scribbling ideas does not guarantee structured retrieval of specific information. Clear instructions, guidelines, and examples can be given, but might complicate the approach (e.g. having to fill out a kind of form on paper for each idea) and could still be bypassed by the end-user (e.g. answer options on paper can be skipped where in a software tool mandatory fields can be defined). Appropriately designed tools can structure the idea specification process and provide a more seamlessly integrated guidance to end-users than paper-based approaches. This is to some degree already done in PDotCapturer in having the 'emotional response' as a mandatory meta information to be specified for each idea. As a result PDotAnalyser is able to generate a heatmap, visualising colour-coded areas of cumulated positive or negative emotional response towards the initial prototype design. This heatmap was perceived helpful by the participating designers of the studies.

Further enhancing and structuring the data gathering activity could on one hand support and guide end-users in expressing their ideas (e.g. the CAt+ categories could be shown to the end-user to select from in addition to providing drawing and text). This could not only be done through tool usage but also through adaptation of existing paper-based PD approaches. On the other hand further structuring of the data gathering process could simplify or partly even obviate the necessity of performing the CAt+ rating for designers or through algorithms.

RQ3 and RQ4: *To what extent is the* ***number of ideas*** *captured by tool-based PD activities different from that by their paper-based counterparts using an* ***off-the-shelf tool*** *and using a newly created, dedicated* ***PD tool****?*

Differences in the number of ideas elicited might be attributed to the learning time and effort needed to get to know a tool, which is not necessary for the already familiar paper-based approach (Sellen and Harper, 2003, p. 18). Therefore assigning fifteen minutes of each tool-based PD workshop to familiarize the end-users with using the tool to provide their ideas, could be beneficial. This way the participants would be more familiar with the tool before attempting the actual idea gathering task. An issue would then be that the tool-based approach is less cost-effective than paper, because of the learning time associated with the tool. This could be mitigated on the other hand by saving time for data digitalisation.

The presented findings suggested that a dedicated tool can close the (quantitative) gap between a tool and paper, performing better than a general purpose tool. Although the tool was dedicated to the task, some requirements were necessary (e.g. has to work in a browsers, mouse input) to ensure it is still generic enough to be widely applicable, like paper, which can be used in many different circumstances. However these restrictions might at some point prevent the tool from further improvements, like stylus input or audio recording, making it even more comparable or even outperform paper. Easing some of the initial restrictions might then be required to allow for further tailoring. In this case a balance needs to be found between optimal support and real world applicability (e.g. economic conditions, like cost of equipment and available budget). To make use of some tool benefits, designers might even accept some disadvantages over paper. To inform practitioners about the approach options available to them and to allow them to make informed decisions PD researchers need to compare currently applied approaches with potential new alternatives.

RQ5 and RQ6: *How are the ideas captured by tool-based PD activities* ***qualitatively different*** *from those by their paper-based counterparts using an* ***off-the-shelf tool*** *and using a newly created, dedicated* ***PD tool****?*

Whereas the quantitative difference between two alternatives is rather easy and straightforward to specify, it is more complicated for qualitative differences. Nevertheless this is very important to do to determine if and when an off-the-shelf tool can be used to support or replace paper or other non-digital PD approaches. Following the example of performing the comparison for one tool, presented in this thesis, other comparisons should be performed to create a database of different tool and paper-based approach qualities. Based on these findings a set of guidelines when to use which approach could be created to guide practitioners and researchers in the selection process. The research presented in this thesis can thus be seen as a starting point and should be expanded to cover more and more application domains of PD.

Quality-wise the results of the comparison of paper and PDotCapturer are somewhat inconclusive. There is a tendency of paper eliciting more ideas on the functionality of the prototype. Other than that the comparison shows that the qualities of the ideas elicited with paper and tool are comparable and thus researchers and practitioners can use the tool-approach to benefit from its advantages (e.g. analysis support).

7.2 General Implications

CAt+ coding scheme

For the study evaluating PDotAnalyser, CAt+ was integrated into the tool for designers. However, CAt+ can be used to rate PD ideas independent of the gathering approach (i.e. also for ideas gathered with non-tool approaches). The data enhanced with CAt+ information could then support the analysis process for designers. Improvements of the data analysis task lead to a better utilisation of PD activities, not by ensuring that end-users are supported in specifying their ideas, but by ensuring designers are supported in considering and appropriately representing these ideas (Read et al., 2016).

Possible sources of differences in results other than the approach used

This thesis focussed on qualitative studies to evaluate the approaches and compare the PD ideas gathered. Individual differences of participants were either addressed through a fully crossed study design (counterbalancing differences in end-users through having them use both approaches) or randomization of the participants, where a fully crossed study design was not feasible. However, differences in the results of comparing paper and tool could be caused by confounding variables other than the tool or paper-based approach itself. Experience (for example with PCs or drawing on paper) and other factors (for example age, attitude towards each approach) could influence the result of the comparison. Large-scale, quantitative studies should be performed to identify these variables and therefore building the groundwork for recommending paper or tool-based approaches.

Roles in PD idea analysis

Currently PDotAnalyser is a generic tool offering one GUI to support the broad role of 'designer'. The tool and support it offers have been perceived positively by the participants evaluating its usability and user experience. However, as this broad and generic support is only one of the possible options, one open question is, what the best way to address PD idea analysis conceptually would be.

It could be that having these roles combined works best; such a combination seems necessary at least in small projects and SMEs due to a small size of staff. Or it might be beneficial to have separate roles: HCI specialist for conducting the PD activities, graphic designer for creating a visual redesign based on the ideas gathered, and developer for implementing the redesign. The latter approach could be further supported by PDotAnalyser through different, slightly adapted versions of the GUI.

Unfortunately, an answer to this question cannot be given yet, as PDotAnalyser has not been used 'in the wild' so far. An advantage of the 'combined roles' approach could be that one person follows through the entire process of idea generation and analysis. This can mitigate the risk of information loss (or misinterpretation) caused by knowledge transfer through a chain of people. A disadvantage of combining the roles in one person might be that, for example, a developer without proper graphic design and HCI training might not get the most out of the PD activities (cf. Bruun and Stage, 2014).

Logically speaking, an advantage of having dedicated people for each role would be the opportunity to make the most of the experience of each role and person. In collaboration with the graphic designers and developers, the HCI specialists can plan the target and content of the PD activity. HCI specialists would then be responsible to recruit end-users as participants, to set up and to conduct the activity (using PDotCapturer). After the PD event the HCI specialist could prepare the ideas gathered for further analysis through graphic designers and developers (using PDotAnalyser). The graphic designers and developers can then focus on the analysis and addressing of the ideas (using their dedicated GUIs of PDotAnalyser). The cognitive workload of idea analysis can thus be split between HCI specialists and designers/developers. For example HCI specialists can use the CAt+ coding scheme to rate the ideas. This would provide cluster and filter options for the designers/developers to make their task of considering and addressing the ideas less cognitively demanding.

That such a separation would be useful was also expressed by the participants in the third study evaluating version 2 of PDotAnalyser. Although the integrated CAt+ coding scheme was perceived as useful, the participants did not consider the coding itself as an inherently meaningful task for developers. They would prefer having the coding already done automatically or at least automated where possible. As developers they would like to focus on using the categories and attributes to filter and find relevant ideas, when working on the improvement of the prototype.

With clearly separate roles and tasks using specific approaches and tools, the areas in which people work and research can be distinctly defined. Even if several roles are combined in the same person, clear task separation (for example by using different tools for different tasks) could help practitioners and researchers to focus more clearly on the aspects and expertise of the role they assume. Combining roles for the task of PD idea analysis, on the other hand, makes the lines defining areas of interest and expertise even more blurry (cf. the implications of Sanders' Design map (Sanders, 2006, 2008) discussed in Section 2.1). In fact, such role/area defining lines might not be that meaningful anymore, because they are 'crossed', transitioned, and negotiated constantly. This could provide one explanation why the definition of PD is becoming less clear cut and changing over time.

The combination of roles in PD also has possible implications for the education of future 'designers'. To address the need for filling a versatile rather than a specialised position they need a broader instead of very specific education, combining expertise from graphic design, software development and HCI research.

Democratisation in software design

As described in the positioning of this research in the PD research area section (Section 2.1) PDotCapturer focusses on collecting end-user ideas to improve the

GUI design of software. It has been criticized that this diverges from the initial democratisation of the work conditions idea of PD (ACMTOCHI, 2016; Gregory, 2003).

However, the tenet of participatory design, to 'let all voices be heard', is also the motivation to apply PD in software design. By shaping not only the GUI design, but also the functionality of software, end-users actively influence the features of the software they will have to work with. Especially for knowledge workers, who mainly or even solely work with software, software changes are changes to a major part of their working conditions. Tools can support the democratisation process by facilitating sharing and discussion of ideas within and between different groups of end-users. Unlike with paper-based approaches end-users do not have to provide their ideas all at the same time and while being in the same location. This allows for the inclusion of a broader set of end-users which might otherwise not be heard (Walsh et al., 2012). But also in co-located settings tools can ensure that no voice is lost, for example through digital storage of the ideas, allowing for easy access later.

Generalisability of the findings for other PD areas

In this work the comparison of two tools (myBalsamiq and different versions of PDotCapturer) and different comparable paper-based approaches (Layered Elaboration (Walsh et al., 2010) and Paper Booklets) was performed. Besides paper-based methods, there are plenty of other, non-paper-based, non-digital methods in PD, e.g. using foam, clay, or Lego bricks to form a three dimensional prototype (Sanders et al., 2010). PDotCapturer might be applicable to support some of them (e.g. by annotating pictures of a three-dimensional product or by loading a 3D model instead of pictures into PDotCapturer). PDotCapturer would in this case be a different approach rather than a way to replace the existing one and software tools that look different from PDotCapturer might be needed. A method creating three-dimensional prototypes might also require specialised technology, like virtual reality (VR).

With the CAt+ rating scheme being dedicated to PD ideas elicited on webapps, PDotAnalyser cannot necessarily be used for other PD target areas. However, by adapting CAt+ if and where necessary or changing the rating scheme, PDotAnalyser could be adapted for other PD areas.

PDotCapturer and PDotAnalyser can be applied in any setting where comparable paper-based methods (like Layered Elaboration) are applied. It does not necessarily have to be webapps in particular or more general software design.

7.3 Limitations

Having performed several rather small empirical studies to find answers to the different research questions posed in this thesis, results in several possible threats of validity. Besides the threats for individual studies, like small sample sizes, there are some threats caused by the context and setting of the research in general, presented here as well.

Small sample size

Most of the studies performed in the scope of the research presented in this thesis only had a relatively small sample size. Thus the results are not necessarily representative and can not necessarily be generalized. More studies with more participants would be useful to further substantiate the findings presented in this thesis.

Missing background information on participants

The tight schedule for most of the PD events and considerations of questionnaire length most of the time did not allow for the collection of substantial background information on the participants. Therefore assumptions on possible causes and reasoning for observed results can not always be backed up with information regarding the participants. Additionally, such background information could have been used to eliminate the influence of possible confounding variables between participants (e.g. when putting different participants in groups).

CAt+ integration in PDotAnalyser

PDotAnalyser only supports the CAt+ coding scheme for rating of the PD ideas. CAt+ was initially developed to compare paper- and tool-based results and integrated based on the assumption that it might also be beneficial when analysing the results. Other methods exist and might be equally or maybe even more appropriate (e.g. RAId, Read et al., 2016).

Confounding variables

Although the different studies were carefully designed to account for possible confounding variables, the settings and differences between studies can have introduced confounding variables influencing the results and findings.

Computer literacy of participants

A possible limitation of this research might be that most of the participants in the above described studies had good computer literacy. Results of comparing paper and software tools might differ if non-PC-experts were to be involved.

Limited experience of participants with PDotCapturer tool

Most of the studies presented above were performed with participants that had not used PDotCapturer before. Giving the learning of interface and interaction associated with the tool, studies with experienced or even expert PDotCapturer users might have yielded different results.

Focus on annotating existing prototypes

Although Participatory Design is a broad field with a great variety of approaches, methods, tools, and artefacts to be designed, this thesis mainly focuses on one particular usage scenario of PD: annotating existing designs. This resulted in a specific focus of the work and the resulting findings, thus being a limitation of this research.

Go-Lab project context

Focussing on PD activities performed in the context of the Go-Lab project restricted the research. Undoubtedly the project context introduced some special requirements, for example that the tool had to work in a constrained school environment (where for example installation of external software is not allowed). The participants and applications mostly came from an educational background. Therefore the generalizability of the findings outside the educational sector might be constrained.

Focus on usability and UX of the tools

The presented research focussed on the usability of the tools and the user experience of end-users and designers when using them. Other factors, like cognitive load, have not been taken into consideration, but it would be interesting to measure them as well and compare paper and tool based on these factors.

Chapter 8

Conclusion and Future Work

This chapter presents conclusions and future work suggestions for the different topics covered in this thesis.

Tool-support in PD

For end-users a tool-based PD sessions could offer several advantages over a paper-based one (as confirmed by the responses of participants in the free text fields in the questionnaire comparing paper and tool when creating a design from scratch): When annotating the initial designs of other end-users participating in the same event, there is no issue with illegible handwriting, as most of the text is typed. The understandability of initial designs for other end-users annotating as well as designers analysing the ideas could be improved through elaborate textual descriptions and explanations. While these could clutter the interface design on paper and could cause space issues, as they are written inside the design space, they were 'hidden' in small markers not interfering with the prototype design with the tool. Typing was perceived as advantageous over writing by hand by some participants. Additionally, PDotCapturer was preferred over paper to explore and annotate low-fidelity prototypes. These benefits from the end-users perspective should motivate more tool-usage in PD.

Besides end-users also designers could benefit from tool-usage in PD. The formal evaluations presented in this thesis showed that for example the heatmap visualisation generated by PDotAnalyser was perceived as helpful. In addition, informal observations when PDotAnalyser was used to analyse PDotCapturer data, provided insights into potential benefits of the tool: One anecdotal finding gathered this way was that the quicker and easier retrieval of the original expressed idea in its context, by following an automatically generated link from the Excel datasheet compared to searching through piles of paper, was identified as one of the benefits of the tool-supported approach over paper. Another one was that, due to typed textual comments in the tool, there was no issue with illegible handwriting, which sometimes posed a problem with comments written on paper. Additionally, PDotAnalyser allowed the designer to filter and explore the ideas in different ways, e.g. based on distribution on the prototype or chronological. This helped to identify reoccurring themes and patterns in the data. The emotional response indicated by the end-user when expressing an idea sometimes added to its understandability. Especially for short comments it could help to decide if an issue was highlighted (red smiley to indicate negative mood) or approval was expressed (green colour indicating a positive response).

Practical benefits of PDotCapturer were that the ideas were stored digitally (easy access to the raw data from any browser) and each PD idea was automatically assigned meta-data like time and date, event ID, and participant ID, when it was exported from the database. With paper these information either had to be added manually to each idea dataset (e.g. participant ID) or could sometimes not be retrieved at all (e.g. exact time). These benefits from the designer-perspective should motivate them to promote tool-usage in PD.

Comparison of paper and tool in PD

To address the open question which influences the experience with PDotCapturer has on the comparison results, additional studies could be conducted that first get participants familiar with the tool and then compare it with paper. This could be done by using PDotCapturer in a series of PD workshops, with the comparison with paper being conducted after several tool-supported sessions, not in the first one.

PDotCapturer

To allow for a general applicability of PDotCapturer the technical requirements have been restricted to a minimum (any browser, Internet access, mouse and keyboard as input devices). By loosening these restrictions more technically advanced options could be explored. For example one extension suggested by participants in the studies was audio input. By requiring headsets or at least microphones at the PD location, instead of typing textual feedback the end-user could just talk to the tool. The designers could then either work with the audio recordings or text-to-speech technology could be used to transform the audio input to textual data.

By requiring touch- or stylus-supporting devices PDotCapturer could offer an input modality closer to the paper-based experience than drawing with a mouse. It would thus be interesting to compare PDotCapturer on a touch-screen device with a paper-based approach. Future studies and tool extensions could explore these options and evaluate how these more advanced modifications of a tool change its comparison with paper.

The studies presented in this thesis mainly focussed on features, usability, and user experience of PDotCapturer, although it would also be interesting to measure other factors such as cognitive load while using the tool. Future studies should look into that and also in comparing paper and tool regarding these factors.

PDotAnalyser

Based on the latest evaluation results, PDotAnalyser can be further improved to support designers in the PD idea analysis task. These changes then need to be evaluated, leading to possible further enhancements. The results can also be used to further substantiate the evaluation results presented in this thesis.

Explore different usage scenarios

To broaden the research started in this thesis PDotCapturer could be used in different use case scenarios. For example as a communication tool between end-user and designer. Currently the tool has mostly been applied in time critical situations, where idea gathering and analysis had to be performed in sequence. However, PDotCapturer could be used in a synchronous or asynchronous setting where the end-user draws an idea, the designer changes the prototype, the end-user expresses the next idea, and so on.

Although PDotCapturer and PDotAnalyser have only been used in co-located settings so far, their design is feasible for using them in a distributed setting. Gathering of PD ideas with PDotCapturer and analysis of the ideas using PDotAnalsyser can be independent and does not have to be performed in the same place and at the same time.

Not all (new) issues arising when changing from a co-located to a distributed PD setting can be addressed by tool usage. For example, motivating end-users to participate in activities is not automatically achieved by applying a software tool to gather ideas. But it can help to mitigate some of the barriers: providing ideas with a tool can be more flexible and less cumbersome (e.g. ideas are automatically transferred to the designer, the end-user does not have to send them manually). It would be interesting to further investigate, how tools could be enhanced to address the motivation issue. For example gamification could be used to make providing ideas more playful, engaging, and stimulative.

For the development of PDotCapturer and PDotAnalyser software design in general and webapps in particular were chosen. However, the resulting tools could also be used in other application domains. For architecture and city planning instead of a software prototype, an architectural drawing or a town map could be displayed. The functionality of PDotCapturer could then be used straight away to annotate these.

For other domains the tool or method would need to be slightly adapted. For example, for 3D models in product development the current version of PDotCapturer could only work with pictures of the model to be annotated. But it could be extended with a model viewer, which would allow the 3D model to be presented directly in the tool. For other application domains, like politics, no purposeful usage scenario might be found. It would be interesting to explore the applicability and changes necessary for PDotCapturer and PDotAnalyser in these areas.

CAt+

Through several evaluations comparing PD ideas gathered with different approaches and on different maturity of prototypes, the completeness of the coding scheme CAt+ has been evaluated and established. It can now be recommended to be used in the coding of PD ideas on webapps. Even if PD ideas are elicited with paper-based approaches, coding them using CAt+ can still be beneficial for the further analysis process. The CAt+ coding scheme should be applied to other datasets, collected in different contexts than webapps so as to further validate its robustness and generalisability.

In addition, it could be further evaluated how designers use and benefit from the CAt+ coding of ideas to utilise possible improvement potentials. The initial evaluation performed in conjunction with the PDotAnalyser evaluation is promising. But PDotAnalyser and CAt+ will have to be used 'in the wild' to evaluate them and their applicability further. Additionally they should be compared with other alternatives to support designers in their task.

To summarize, this thesis adds to the sparse research literature regarding tool-support for PD activities and its comparability with established paper-based PD approaches. An off-the-shelf tool suitable for PD was identified and compared with paper. Furthermore PDotCapturer, a dedicated PD tool, was created and also compared with paper-based approaches in several empirical studies in different settings. To not only support end-users but also designers in PD activities with a tool, PDotAnalsyer was designed, developed and evaluated. Finally, with CAt+ a coding scheme was created that not only allows for the qualitative comparison of different PD approaches, but can also be used to further enrich PD ideas gathered to provide analysis support for designers.

Appendix A

Appendices

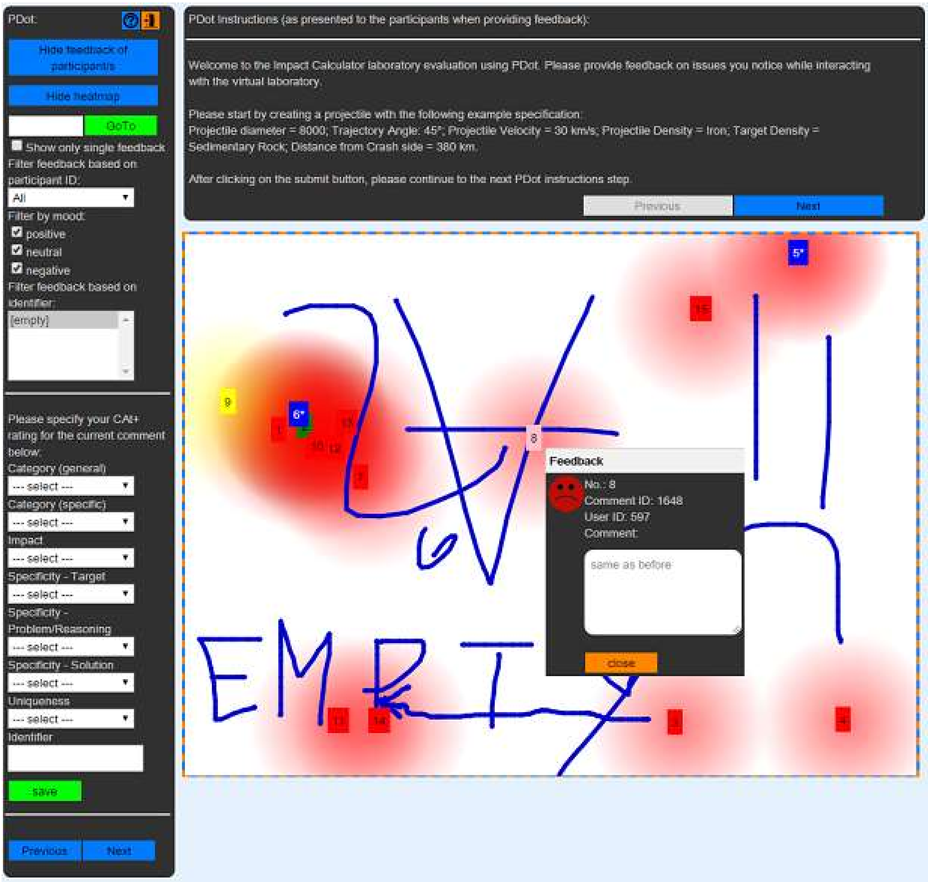
A.1 Questionnaire regarding CAt+ for designers

PDot Researcher View questionnaire for final evaluation at EPFL

Final evaluation of Researcher View of PDot with developers.

Welcome to the PDot Researcher View evaluation for developers.

The questions in this questionnaire are about the **Researcher View** of PDot that you just used to analyse and rate feedback (see screenshot below).



You can start by clicking "Next" at the bottom.

There are 18 questions in this survey

General information

1 [PRVFEPDOTUSERID]PDot username *

Please write your answer here:

2 [PRVFEAGE]Age: *

Please write your answer here:

3 [PRVFEG]Gender: *

Please choose **all** that apply:

- ☐ Female
- ☐ Male
- ☐ Prefer not to say
- ☐ Other:

4 [PRVFEICTC]How do you rate your ICT competence (from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

5 [PRVFESD]How experienced are you in software development (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

6 [PRVFEUE]How experienced are you in usability evaluations (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

7 [PRVFEPREV]Have you used a previous version of the PDot Researcher View before? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

User Experience

8 [AttrakDiff2_short]

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of the PDot researcher view.

An example: **disagreeable** ○ ○ ○ ○ ◉ ○ ○ **likeable**

This choice tells us that the PDot researcher view is somewhat likeable, but there is still room for improvement
***(Note: There is no right or wrong answer. Your personal opinion is what counts.)* ***

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

9 [Finstad]For each of the following statements, please indicate your extent of agreement by selecting the number of choice. *

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot researcher view capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using the PDot researcher view is a frustrating experience.	○	○	○	○	○	○	○
3. The PDot researcher view is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with the PDot researcher view.	○	○	○	○	○	○	○

Usability of PDot (researcher view)

10 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree):

Please choose the appropriate response for each item:

	1	2	3	4	5
Features of the PDot researcher view were sufficient for the task.	○	○	○	○	○
Design of PDot researcher view was visually appealing.	○	○	○	○	○
The visualization of the content of the comment in a PopUp was a good way to provide this information.	○	○	○	○	○
I would prefer to have the comment information displayed in the PDot tool on the left hand side instead of the current PopUp solution.	○	○	○	○	○
The information given in the comment PopUp was sufficient.	○	○	○	○	○
More sophisticated ways to filter the feedback (besides by PDot instruction step, UserID, or DrawingID) would be needed.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to learn.	○	○	○	○	○
The PDot researcher view showing single comments was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to learn.	○	○	○	○	○

11 [PRVEUOPRVD]Please elaborate on your ratings above where appropriate (indicating the rating element [Arbitrary example as there is no username: I think the information in the comment PopUp is not sufficient, besides the User ID also the username should be shown.]):

Please write your answer here:

SUS

12 [PRVFESUS]Please rate your agreement to the following statements from Strongly disagree (1) to Strongly agree (5).

Please choose the appropriate response for each item:

	1	2	3	4	5
I think that I would like to use the PDot Researcher View frequently	○	○	○	○	○
I found the PDot Researcher View unnecessarily complex	○	○	○	○	○
I thought the PDot Researcher View was easy to use	○	○	○	○	○
I think that I would need the support of a technical person to be able to use the PDot Researcher View	○	○	○	○	○
I found the various functions in the PDot Researcher View were well integrated	○	○	○	○	○
I thought there was too much inconsistency in the PDot Researcher View	○	○	○	○	○
I would imagine that most people would learn to use the PDot Researcher View very quickly	○	○	○	○	○
I found the PDot Researcher View very cumbersome to use	○	○	○	○	○
I felt very confident using the PDot Researcher View	○	○	○	○	○
I needed to learn a lot of things before I could get going with the PDot Researcher View	○	○	○	○	○
I think that I would like to use the PDot Researcher View again	○	○	○	○	○

CAt+ rating

13 [PRVJTELCRQ]For each of the following statements, please indicate your extent of agreement by selecting the number of choice (from 1 = "strongly disagree" to 5 = "strongly agree").

Please choose the appropriate response for each item:

	1	2	3	4	5
I understand the concept behind the categories for rating feedback.	○	○	○	○	○
Categories are a good instrument to rate feedback.	○	○	○	○	○
Having the categories rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "impact".	○	○	○	○	○
The rating attribute "impact" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the impact rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Target".	○	○	○	○	○
The rating attribute "Specificity - Target" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Target" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Problem/Reasoning".	○	○	○	○	○
The rating attribute "Specificity - Problem/Reasoning" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Problem/Reasoning" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Solution".	○	○	○	○	○
The rating attribute "Specificity - Solution" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Solution" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "uniqueness".	○	○	○	○	○
The rating attribute "uniqueness" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the uniqueness rating would be helpful when addressing the feedback.	○	○	○	○	○

Features

14 [FEATURES_LIKE]What did you like about the researcher view of PDot?

Please write your answer here:

15 [FEATURES_DISLIKE]What did you not like about the researcher view of PDot?

Please write your answer here:

16 [FEATURES_ADD]Which features would you like additionally in the researcher view of PDot?

Please write your answer here:

17 [FEATURES_REM]

Which features would you like to have removed from the researcher view of PDot?

(Please give reasons for your answer.)

Please write your answer here:

18 [Usage]If you would like to use PDot in one of your projects to gather feedback please enter your e-mail address below:

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.2 Questionnaire and Interview questions for Initial Prototype of PDotCapturer

Questionnaire:

Your participation is voluntary and your responses will be handled confidentially. The data will be processed anonymised and will not be identified individually.

--- General information ---

1) Age: o <20 o 20-24 o 25-29 o 30-34 o >34

2) Gender: o Male o Female o Other

3) Field of study/research: ____________________

4) Please specify how you are adopting new technology:

o As one of the first, even if it still has some issues
o Early, but after all of the first issues have been resolved
o When most of the other people use/get it
o After most of the other people already got it
o Very later after all the others or never

5) How would you rate your experience in requirements engineering?
Very low o o o o o Very high

--- Regarding the feedback methods ---

6) Which of the feedback methods did you like best? Please rank them, by putting the numbers from 1 (for "liked best") to 3 ("liked least") into the squares in front of the method:

[ ] paper-based not at all o o o o o very much
[ ] tool-based (myBalsamiq) not at all o o o o o very much
[ ] tool-based (PDtool) not at all o o o o o very much

7) How would you rate the following qualities of the interaction with the different tools from 1 ("not at all") to 5 ("very much")?

	paper	tool (myBalsamiq)	tool (PDtool)
natural	1 o o o o o 5	1 o o o o o 5	1 o o o o o 5
easy to use	1 o o o o o 5	1 o o o o o 5	1 o o o o o 5
easy to learn	1 o o o o o 5	1 o o o o o 5	1 o o o o o 5
flexible	1 o o o o o 5	1 o o o o o 5	1 o o o o o 5

8) Please specify the most critical feature you found missing with each feedback method:

paper: ____

myBalsamiq: ____

PDtool: ____

Thank you very much.

--- Interview ---

9) Please describe briefly the reasons for your choice when ranking the feedback methods:

10) Could you imagine using (one of) those methods in the future, if you have access to the tool?

a) paper

b)myBalsamiq

c) PDtool

Context:

a) to give feedback

b) to gather feedback

c) recommend it to peers

d) other

Why (not)?

Do you have any suggestions how to improve the feedback methods (paper, myBalsamiq, PDtool)?

A.3 Questionnaire for Version 2 of PDotCapturer for Students

PDot User View evaluation (ULeic students)

Evaluation of User View of PDot for ULEIC students.

Welcome to the PDot user view evaluation for University of Leicester students.

Your participation is voluntary and your responses will be handled confidentially. The data will be processed in batches and will not be identified individually. You can start by clicking "Next" at the bottom.

There are 9 questions in this survey

General information

1 [PUVEGISI]StudentID *

Please write your answer here:

[ ]

2 [PUVEGIG]Gender: *

Please choose **only one** of the following:

- ○ Female
- ○ Male

3 [PUVEGIICTC]How do you rate your ICT competence (from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

4 [PUVEGIFWE]How do you rate your familarity with the topic of electricity as presented in the learning content of the Inquiry Learning Space (ILS) you just evaluated (before working with the ILS; from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

Usability of PDot (user view)

5 [PUVEUOPUVQ1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree): *

Please choose the appropriate response for each item:

	1	2	3	4	5
The tool used to conduct the evaluation (PDot) was easy to learn.	○	○	○	○	○
PDot was easy to use.	○	○	○	○	○
Features of PDot were sufficient for the task.	○	○	○	○	○
Design of PDot was visually appealing.	○	○	○	○	○
Having instructions to follow included in PDot was useful.	○	○	○	○	○
Sticky notes marking the spot of my feedback were useful.	○	○	○	○	○
Freehand drawing functionality was useful.	○	○	○	○	○
Textual feedback option was useful.	○	○	○	○	○
Option to specify mood was useful.	○	○	○	○	○
It was easy to switch between different modes (instructions, interact with app, give feedback).	○	○	○	○	○
PDtool was useful to give feedback.	○	○	○	○	○

Features

6 [PUVEF_FEATURES_LIKE]What did you like about PDot? *

Please write your answer here:

7 [PUVEFEATURES_DISLIKE]What did you not like about PDot? *

Please write your answer here:

8 [PUVEF_FEATURES_ADD]Which features would you like additionally in PDot? *

Please write your answer here:

9 [PUVEF_FEATURES_REM]

Which features would you like to have removed from PDot?

(Please give reasons for your answer.)

*****

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.4 Questionnaire for Version 2 of PDotCapturer for Teachers

PDot evaluation (teacher)

Evaluation of PDot for teachers.

Welcome to the PDot evaluation for teachers.

Your participation is voluntary and your responses will be handled confidentially. The data will be processed in batches and will not be identified individually. You can start by clicking "Next" at the bottom.

There are 11 questions in this survey

General information

1 [PUVETGIA]Age

Please write your answer here:

2 [PUVETGIG]Gender:

Please choose **only one** of the following:

- Female
- Male

3 [PUVETGIST]School type: *

Please choose **only one** of the following:

- Primary School
- Secondary School
- Other

4 [PUVETGIC]Country: *

Please choose **only one** of the following:

- Austria
- Belgium
- Cyprus
- Estonia
- France
- Germany
- Greece
- Netherlands
- Portugal
- Spain
- Switzerland
- United Kingdom
- Other

User Experience

5 [AttrakDiff2_short]

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of PDot.

An example: **disagreeable** ○ ○ ○ ○ ◉ ○ ○ **likeable**

This choice tells us that PDot is somewhat likeable, but there is still room for improvement
(Note: There is no right or wrong answer. Your personal opinion is what counts.)

*

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

6 [Finstad]For each of the following statements, please indicate your extent of agreement by selecting the number of choice. *

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using PDot is a frustrating experience.	○	○	○	○	○	○	○
3. PDot is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with PDot.	○	○	○	○	○	○	○

Usability of PDot

7 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree):

Please choose the appropriate response for each item:

	1	2	3	4	5
The tool used to conduct the evaluation (PDot) was easy to learn.	◯	◯	◯	◯	◯
PDot was easy to use.	◯	◯	◯	◯	◯
Features of PDot were sufficient for the task.	◯	◯	◯	◯	◯
Design of PDot was visually appealing.	◯	◯	◯	◯	◯
Having a scenario to follow included in PDot was useful.	◯	◯	◯	◯	◯
Sticky notes marking the spot of my feedback were useful.	◯	◯	◯	◯	◯
Freehand drawing functionality was useful.	◯	◯	◯	◯	◯
Textual feedback option was useful.	◯	◯	◯	◯	◯
Option to specify mood was useful.	◯	◯	◯	◯	◯
It was easy to switch between different modes (instructions, interact with app, give feedback).	◯	◯	◯	◯	◯
PDot was useful to give feedback.	◯	◯	◯	◯	◯

Features

8 [FEATURES_LIKE]What did you like about PDot?

Please write your answer here:

9 [FEATURES_DISLIKE]What did you not like about PDot?

Please write your answer here:

10 [FEATURES_ADD]Which features would you like additionally in PDot?

Please write your answer here:

11 [FEATURES_REM]

Which features would you like to have removed from PDot?

(Please give reasons for your answer.)

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.5 Questionnaire for Version 2 of PDotAnalyser for Students

PDot Researcher View evaluation (ULeic students)

Evaluation of Researcher View of PDot for ULEIC students.

Welcome to the PDot researcher view evaluation for University of Leicester students.

You can start by clicking "Next" at the bottom.

There are 10 questions in this survey

General information

1 [PRVEGISI]StudentID *

Please write your answer here:

User Experience

2 [AttrakDiff2_short]

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of the PDot researcher view.

An example: **disagreeable** ○ ○ ○ ○ ◉ ○ ○ **likeable**

This choice tells us that the PDot researcher view is somewhat likeable, but there is still room for improvement
(Note: There is no right or wrong answer. Your personal opinion is what counts.)

*****

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

3 [Finstad]For each of the following statements, please indicate your extent of agreement by selecting the number of choice. *

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot researcher view capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using the PDot researcher view is a frustrating experience.	○	○	○	○	○	○	○
3. The PDot researcher view is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with the PDot researcher view.	○	○	○	○	○	○	○

Usability of PDot (researcher view)

4 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree): *

Please choose the appropriate response for each item:

	1	2	3	4	5
Features of the PDot researcher view were sufficient for the task.	○	○	○	○	○
Design of PDot researcher view was visually appealing.	○	○	○	○	○
The visualization of the content of the comment in a PopUp was a good way to provide this information.	○	○	○	○	○
I would prefer to have the comment information displayed in the PDot tool on the left hand side instead of the current PopUp solution.	○	○	○	○	○
The information given in the comment PopUp was sufficient.	○	○	○	○	○
More sophisticated ways to filter the feedback (besides by PDot instruction step, UserID, or DrawingID) would be needed.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to learn.	○	○	○	○	○
The PDot researcher view showing single comments was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to learn.	○	○	○	○	○

5 [PRVEUOPRVD]Please elaborate on your ratings above where appropriate (indicating the rating element [Arbitrary example as there is no username: I think the information in the comment PopUp is not sufficient, besides the User ID also the username should be shown.]):

Please write your answer here:

Features

6 [FEATURES_LIKE]What did you like about the researcher view of PDot? *

Please write your answer here:

7 [FEATURES_DISLIKE]What did you not like about the researcher view of PDot? *

Please write your answer here:

8 [FEATURES_ADD]Which features would you like additionally in the researcher view of PDot? *

Please write your answer here:

9 [FEATURES_REM]

Which features would you like to have removed from the researcher view of PDot?

(Please give reasons for your answer.)

*****

Please write your answer here:

10 [Interview]To gather more detailed information regarding PDot (user and researcher view), interviews with more detailed questions regarding functionality and usability will be conducted. If you would like to participate in such an interview (which would also be a good opportunity to get an insight and first hand experience in this user evaluation technique), please enter your e-mail address below:

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

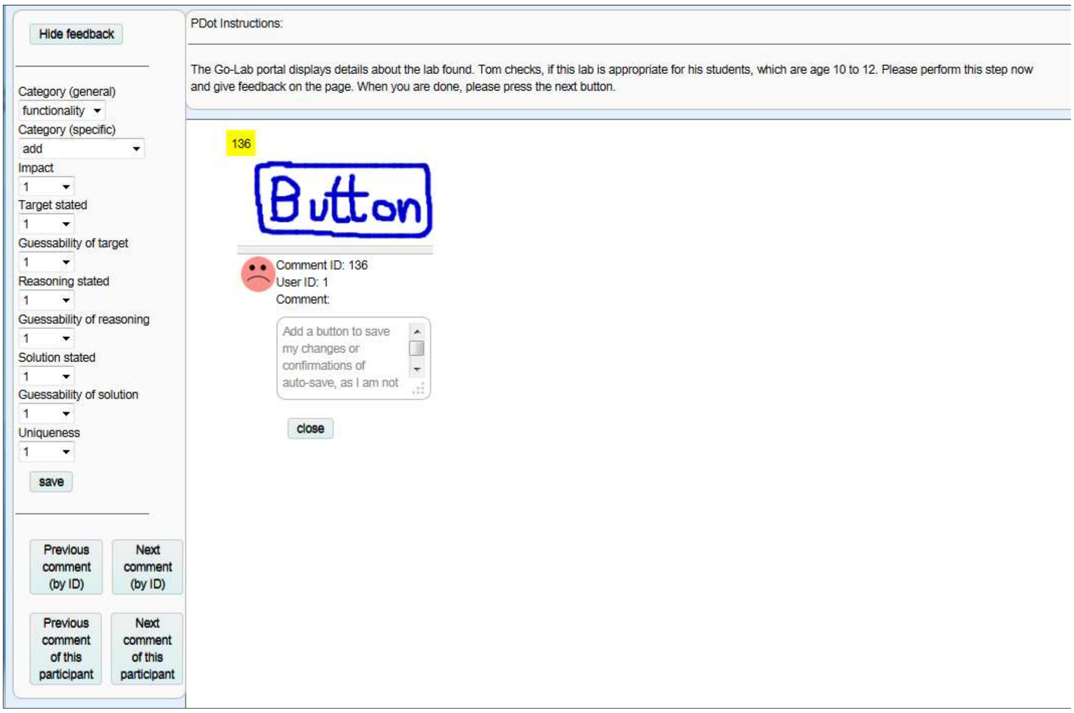
A.6 Questionnaire for Version 2 of PDotAnalyser for Researchers and Practitioners

PDot Researcher View questionnaire for JTEL summer school 2015

Evaluation of Researcher View of PDot for PhD students at JTEL summer school 2015.

Welcome to the PDot Researcher View evaluation for PhD students at the JTEL summer school 2015.

The questions in this questionnaire are about the **Researcher View** of PDot that you just used to analyse and rate feedback (see screenshot below).



You can start by clicking "Next" at the bottom.

There are 11 questions in this survey

General information

1 [PRVJTELGIPDOTUSERID]PDot username *

Please write your answer here:

[ ]

User Experience

2 [AttrakDiff2_short]

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of the PDot researcher view.

An example: **disagreeable** ○ ○ ○ ○ ◉ ○ ○ **likeable**

This choice tells us that the PDot researcher view is somewhat likeable, but there is still room for improvement
(Note: There is no right or wrong answer. Your personal opinion is what counts.)

*****

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

3 [Finstad]For each of the following statements, please indicate your extent of agreement by selecting the number of choice. *

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot researcher view capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using the PDot researcher view is a frustrating experience.	○	○	○	○	○	○	○
3. The PDot researcher view is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with the PDot researcher view.	○	○	○	○	○	○	○

Usability of PDot (researcher view)

4 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree):

Please choose the appropriate response for each item:

	1	2	3	4	5
Features of the PDot researcher view were sufficient for the task.	○	○	○	○	○
Design of PDot researcher view was visually appealing.	○	○	○	○	○
The visualization of the content of the comment in a PopUp was a good way to provide this information.	○	○	○	○	○
I would prefer to have the comment information displayed in the PDot tool on the left hand side instead of the current PopUp solution.	○	○	○	○	○
The information given in the comment PopUp was sufficient.	○	○	○	○	○
More sophisticated ways to filter the feedback (besides by PDot instruction step, UserID, or DrawingID) would be needed.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to learn.	○	○	○	○	○
The PDot researcher view showing single comments was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to learn.	○	○	○	○	○

5 [PRVEUOPRVD]Please elaborate on your ratings above where appropriate (indicating the rating element [Arbitrary example as there is no username: I think the information in the comment PopUp is not sufficient, besides the User ID also the username should be shown.]):

Please write your answer here:

Features

6 [FEATURES_LIKE]What did you like about the researcher view of PDot?

Please write your answer here:

7 [FEATURES_DISLIKE]What did you not like about the researcher view of PDot?

Please write your answer here:

8 [FEATURES_ADD]Which features would you like additionally in the researcher view of PDot?

Please write your answer here:

9 [FEATURES_REM]

Which features would you like to have removed from the researcher view of PDot?

(Please give reasons for your answer.)

Please write your answer here:

10 [Interview]To gather more detailed information regarding PDot (user and researcher view), interviews with more detailed questions regarding functionality and usability will be conducted. If you would like to participate in such an interview (which would also be a good opportunity to get an insight and first hand experience in this user evaluation technique), please enter your e-mail address below:

Please write your answer here:

CAt+ rating

11 [PRVJTELCRQ]For each of the following statements, please indicate your extent of agreement by selecting the number of choice (from 1 = "strongly disagree" to 5 = "strongly agree").

Please choose the appropriate response for each item:

	1	2	3	4	5
I understand the concept behind the categories for rating feedback.	○	○	○	○	○
Categories are a good instrument to rate feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "impact".	○	○	○	○	○
The rating attribute "impact" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "target stated".	○	○	○	○	○
The rating attribute "target stated" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "guessability of target".	○	○	○	○	○
The rating attribute "guessability of target" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "reasoning stated".	○	○	○	○	○
The rating attribute "reasoning stated" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "guessability of reasoning".	○	○	○	○	○
The rating attribute "guessability of reasoning" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "solution stated".	○	○	○	○	○
The rating attribute "solution stated" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "guessability of solution".	○	○	○	○	○
The rating attribute "guessability of solution" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
I understand the meaning of the rating attribute "uniqueness".	○	○	○	○	○
The rating attribute "uniqueness" is a good measurement to quantify the quality of a comment.	○	○	○	○	○

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

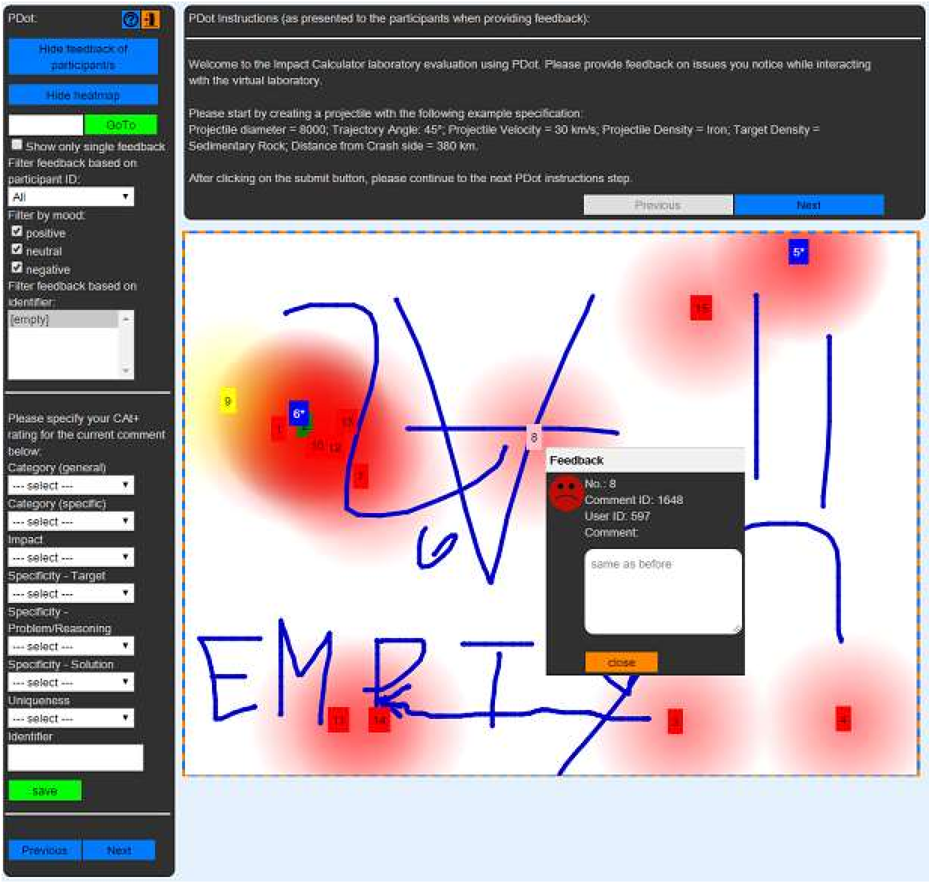
A.7 Questionnaire for Version 2 of PDotAnalyser for Designers

PDot Researcher View questionnaire for final evaluation at EPFL

Final evaluation of Researcher View of PDot with developers.

Welcome to the PDot Researcher View evaluation for developers.

The questions in this questionnaire are about the **Researcher View** of PDot that you just used to analyse and rate feedback (see screenshot below).



You can start by clicking "Next" at the bottom.

There are 18 questions in this survey

General information

1 [PRVFEPDOTUSERID]PDot username *

Please write your answer here:

2 [PRVFEAGE]Age: *

Please write your answer here:

3 [PRVFEG]Gender: *

Please choose **all** that apply:

- ☐ Female
- ☐ Male
- ☐ Prefer not to say
- ☐ Other:

4 [PRVFEICTC]How do you rate your ICT competence (from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

5 [PRVFESD]How experienced are you in software development (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

6 [PRVFEUE]How experienced are you in usability evaluations (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

7 [PRVFEPREV]Have you used a previous version of the PDot Researcher View before? *

Please choose **only one** of the following:

- Yes
- No

User Experience

8 [AttrakDiff2_short]

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of the PDot researcher view.

An example:	**disagreeable**	○	○	○	○	●	○	○	**likeable**

This choice tells us that the PDot researcher view is somewhat likeable, but there is still room for improvement
***(Note: There is no right or wrong answer. Your personal opinion is what counts.)* ***

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

9 [Finstad]For each of the following statements, please indicate your extent of agreement by selecting the number of choice. *

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot researcher view capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using the PDot researcher view is a frustrating experience.	○	○	○	○	○	○	○
3. The PDot researcher view is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with the PDot researcher view.	○	○	○	○	○	○	○

Usability of PDot (researcher view)

10 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree):

Please choose the appropriate response for each item:

	1	2	3	4	5
Features of the PDot researcher view were sufficient for the task.	○	○	○	○	○
Design of PDot researcher view was visually appealing.	○	○	○	○	○
The visualization of the content of the comment in a PopUp was a good way to provide this information.	○	○	○	○	○
I would prefer to have the comment information displayed in the PDot tool on the left hand side instead of the current PopUp solution.	○	○	○	○	○
The information given in the comment PopUp was sufficient.	○	○	○	○	○
More sophisticated ways to filter the feedback (besides by PDot instruction step, UserID, or DrawingID) would be needed.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing all comments given in one phase was easy to learn.	○	○	○	○	○
The PDot researcher view showing single comments was useful to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to use to analyse feedback.	○	○	○	○	○
The PDot researcher view showing single comments was easy to learn.	○	○	○	○	○

11 [PRVEUOPRVD]Please elaborate on your ratings above where appropriate (indicating the rating element [Arbitrary example as there is no username: I think the information in the comment PopUp is not sufficient, besides the User ID also the username should be shown.]):

Please write your answer here:

SUS

12 [PRVFESUS]Please rate your agreement to the following statements from Strongly disagree (1) to Strongly agree (5).

Please choose the appropriate response for each item:

	1	2	3	4	5
I think that I would like to use the PDot Researcher View frequently	○	○	○	○	○
I found the PDot Researcher View unnecessarily complex	○	○	○	○	○
I thought the PDot Researcher View was easy to use	○	○	○	○	○
I think that I would need the support of a technical person to be able to use the PDot Researcher View	○	○	○	○	○
I found the various functions in the PDot Researcher View were well integrated	○	○	○	○	○
I thought there was too much inconsistency in the PDot Researcher View	○	○	○	○	○
I would imagine that most people would learn to use the PDot Researcher View very quickly	○	○	○	○	○
I found the PDot Researcher View very cumbersome to use	○	○	○	○	○
I felt very confident using the PDot Researcher View	○	○	○	○	○
I needed to learn a lot of things before I could get going with the PDot Researcher View	○	○	○	○	○
I think that I would like to use the PDot Researcher View again	○	○	○	○	○

CAt+ rating

13 [PRVJTELCRQ]For each of the following statements, please indicate your extent of agreement by selecting the number of choice (from 1 = "strongly disagree" to 5 = "strongly agree").

Please choose the appropriate response for each item:

	1	2	3	4	5
I understand the concept behind the categories for rating feedback.	○	○	○	○	○
Categories are a good instrument to rate feedback.	○	○	○	○	○
Having the categories rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "impact".	○	○	○	○	○
The rating attribute "impact" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the impact rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Target".	○	○	○	○	○
The rating attribute "Specificity - Target" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Target" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Problem/Reasoning".	○	○	○	○	○
The rating attribute "Specificity - Problem/Reasoning" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Problem/Reasoning" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "Specificity - Solution".	○	○	○	○	○
The rating attribute "Specificity - Solution" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the "Specificity - Solution" rating would be helpful when addressing the feedback.	○	○	○	○	○
I understand the meaning of the rating attribute "uniqueness".	○	○	○	○	○
The rating attribute "uniqueness" is a good measurement to quantify the quality of a comment.	○	○	○	○	○
Having the uniqueness rating would be helpful when addressing the feedback.	○	○	○	○	○

Features

14 [FEATURES_LIKE]What did you like about the researcher view of PDot?

Please write your answer here:

15 [FEATURES_DISLIKE]What did you not like about the researcher view of PDot?

Please write your answer here:

16 [FEATURES_ADD]Which features would you like additionally in the researcher view of PDot?

Please write your answer here:

17 [FEATURES_REM]

Which features would you like to have removed from the researcher view of PDot?

(Please give reasons for your answer.)

Please write your answer here:

18 [Usage]If you would like to use PDot in one of your projects to gather feedback please enter your e-mail address below:

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.8 Questionnaire Comparing Paper and PDot-Capturer for Paper Groups

Paper-based method questionnaire for final evaluation with HCI students in March 2016

Evaluation of paper-based method for HCI students in March 2016.

Welcome to the paper-based method evaluation for HCI students.

You can start by clicking "Next" at the bottom.

There are 17 questions in this survey

General information

1 [FEGIGroup]Group: *

Please choose **only one** of the following:

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4

2 [FEGIEvaluationID]Evaluation ID: *

Please write your answer here:

3 [FEGIStudentNumber]Student number: *

Please write your answer here:

4 [FEGIStudentUsername]Student username: *

Please write your answer here:

5 [FEGIA]Age: *

Please write your answer here:

6 [FEGIG]Gender: *

Please choose **only one** of the following:

- Female
- Male
- Prefer not to say
- Other

7 [FEGIICT]How do you rate your ICT competence (from 1 = "very low" to 5 = "very high"; ICT=Information and communications technology)? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

8 [FEGITOPIC]How do you rate your familiarity with creating online quizzes (before working with the Quiz Tool; from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

9 [FEGISD]How experienced are you in software development (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

10 [FEGIUXE]How experienced are you in usability evaluations (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5

Usability and User Experience

11 [FEASUS]Thinking about the method of giving feedback by annotating screen images on paper, please rate your agreement to the following statements from Strongly disagree (1) to Strongly agree (5). *

Please choose the appropriate response for each item:

	1	2	3	4	5
I think that I would like to use this method again	○	○	○	○	○
I found this method unnecessarily complex	○	○	○	○	○
I thought this method was easy to use	○	○	○	○	○
I would imagine that most people would learn to use this method very quickly	○	○	○	○	○
I found this method very cumbersome to use	○	○	○	○	○
I felt very confident using this method	○	○	○	○	○
I needed to learn a lot of things before I could get going with this method	○	○	○	○	○
Being able to create freehand drawings was useful to express my opinion	○	○	○	○	○
Being able to provide textual feedback was useful to express my opinion	○	○	○	○	○

Features

12 [FEATURES_LIKE]What did you like about the way in which you provided feedback?

Please write your answer here:

13 [FEATURES_DISLIKE]What did you not like about the way in which you provided feedback?

Please write your answer here:

14 [FEATURES_IMPROVE]Do you have recommendations how the way in which you provided feedback might be improved?

Please write your answer here:

15 [FEATURES_FROMPAPER]Imagine an online software tool that would allow you to provide feedback. What features would you like to have in this tool, based on your experience of giving feedback on paper?

Please write your answer here:

16 [FEATURES_ADDITIONAL]Are there additional features you can think of that an online software tool could usefully provide?

Please write your answer here:

17 [FEATURES_PREFERENCE]Do you think you would prefer to provide feedback by using an online software tool or the paper-based method? Why?

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.9 Questionnaire Comparing Paper and PDot-Capturer for PDotCapturer Groups

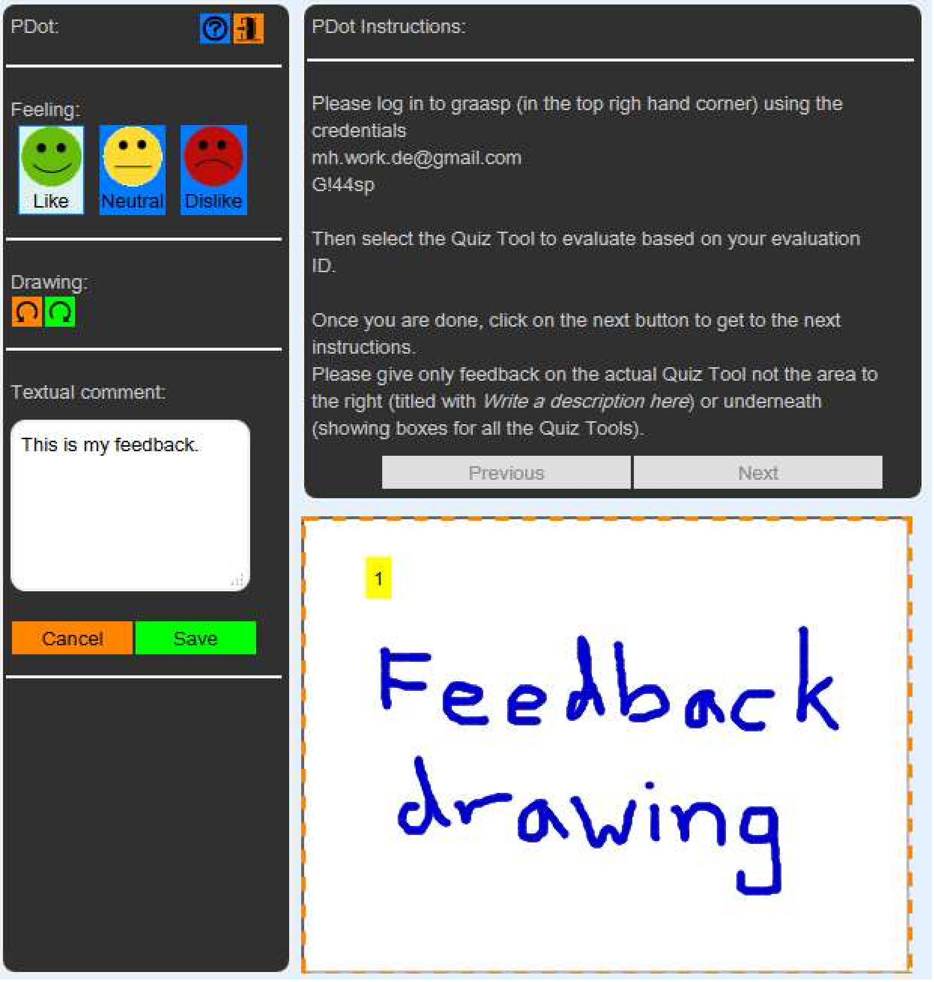
PDot User View questionnaire for final evaluation with HCI students in March 2016

Evaluation of PDot for HCI students in March 2016.

Welcome to the PDot User View evaluation for HCI students.

Note:

The questions in this questionnaire are about the User View of PDot,
the **tool that you just used to give feedback** (see screenshot below)
and **not** about the Quiz Tool you commented on.



You can start by clicking "Next" at the bottom.

There are 20 questions in this survey

General information

1 [FEGIGroup]Group: *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4

2 [FEGIEvaluationID]Evaluation ID: *

Please write your answer here:

3 [FEGIPDOTUSERID]PDot username: *

Please write your answer here:

4 [FEGIStudentNumber]Student number: *

Please write your answer here:

5 [FEGIStudentUsername]Student username: *

Please write your answer here:

6 [FEGIA]Age *

Please write your answer here:

7 [FEGIG]Gender: *

Please choose **only one** of the following:

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other

8 [FEGIICT]How do you rate your ICT competence (from 1 = "very low" to 5 = "very high"; ICT=Information and communications technology)? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

9 [FEGITOPIC]How do you rate your familarity with creating online quizzes (before working with the Quiz Tool; from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

10 [FEGISD]How experienced are you in software development (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

11 [FEGIUXE]How experienced are you in usability evaluations (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- 1
- 2
- 3
- 4
- 5

User Experience

12 [AttrakDiff2_short]

The following attributes are used to rate the PDot User View, the tool you just used to provide feedback.

In the table below you will find 10 pairs of contrasting attributes. Where you place your choice between two attributes indicates your view about the quality of the PDot User View.

An example: **disagreeable** ○ ○ ○ ○ ◉ ○ ○ **likeable**

This choice tells us that PDot is somewhat likeable, but there is still room for improvement
(Note: There is no right or wrong answer. Your personal opinion is what counts.)

*****

Please choose the appropriate response for each item:

Confusing	○	○	○	○	○	○	○	Structured
Practical	○	○	○	○	○	○	○	Impractical
Predictable	○	○	○	○	○	○	○	Unpredictable
Simple	○	○	○	○	○	○	○	Complicated
Dull	○	○	○	○	○	○	○	Captivating
Stylish	○	○	○	○	○	○	○	Tacky
Cheap	○	○	○	○	○	○	○	Premium
Unimaginative	○	○	○	○	○	○	○	Creative
Good	○	○	○	○	○	○	○	Bad
Ugly	○	○	○	○	○	○	○	Beautiful

13 [Finstad]

The following statements are used to further rate the PDot User View, the tool you just used to provide feedback.

For each of the following statements, please indicate your extent of agreement by selecting the number of choice.

*

Please choose the appropriate response for each item:

	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1. The PDot User View capabilities meet my requirements.	○	○	○	○	○	○	○
2. Using the PDot User View is a frustrating experience.	○	○	○	○	○	○	○
3. The PDot User View is easy to use.	○	○	○	○	○	○	○
4. I have to spend too much time correcting things with the PDot User View.	○	○	○	○	○	○	○

SUS

14 [FESUS]Please rate your agreement to the following statements from Strongly disagree (1) to Strongly agree (5). *

Please choose the appropriate response for each item:

	1	2	3	4	5
I think that I would like to use the PDot User View frequently	○	○	○	○	○
I found the PDot User View unnecessarily complex	○	○	○	○	○
I thought the PDot User View was easy to use	○	○	○	○	○
I think that I would need the support of a technical person to be able to use the PDot User View	○	○	○	○	○
I found the various functions in the PDot User View were well integrated	○	○	○	○	○
I thought there was too much inconsistency in the PDot User View	○	○	○	○	○
I would imagine that most people would learn to use the PDot User View very quickly	○	○	○	○	○
I found the PDot User View very cumbersome to use	○	○	○	○	○
I felt very confident using the PDot User View	○	○	○	○	○
I needed to learn a lot of things before I could get going with the PDot User View	○	○	○	○	○
I think that I would like to use the PDot User View again	○	○	○	○	○

Usability of PDot

15 [Q1]Please indicate your opinion regarding the following statements from 1 (Strongly Disagree) to 5 (Strongly agree): *

Please choose the appropriate response for each item:

	1	2	3	4	5
The tool used to conduct the evaluation (PDot User View) was easy to learn.	○	○	○	○	○
The PDot User View was easy to use.	○	○	○	○	○
Features of the PDot User View were sufficient for the task.	○	○	○	○	○
Design of the PDot User View was visually appealing.	○	○	○	○	○
Having a scenario to follow included in the PDot User View was useful.	○	○	○	○	○
Sticky notes marking the spot of my feedback were useful.	○	○	○	○	○
Freehand drawing functionality was useful.	○	○	○	○	○
Textual feedback option was useful.	○	○	○	○	○
Option to specify mood was useful.	○	○	○	○	○
It was easy to switch between different modes (instructions, interact with app, give feedback).	○	○	○	○	○
The PDot User View was useful to give feedback.	○	○	○	○	○

Features

16 [FEATURES_LIKE]What did you like about the PDot User View?

Please write your answer here:

17 [FEATURES_DISLIKE]What did you not like about the PDot User View?

Please write your answer here:

18 [FEATURES_ADD]Which features would you like additionally in the PDot User View?

Please write your answer here:

19 [FEATURES_REM]

Which features would you like to have removed from the PDot User View?

(Please give reasons for your answer.)

Please write your answer here:

20 [FEATURES_PREFERENCE]Do you think you would prefer to provide feedback by using an online software tool or the paper-based method? Why?

Please write your answer here:

Thank you very much for your feedback.

Submit your survey.
Thank you for completing this survey.

A.10 Questionnaire Comparing Paper and PDot-Capturer when Creating a Design from Scratch

Questionnaire comparing paper-based and PDot-based Layered Elaboration with HCI students in March 2016

1 PDot username: *

2 Age: *

3 Gender: *

Please choose **only one** of the following:

- ◯ Female
- ◯ Male
- ◯ Prefer not to say
- ◯ Other:

4 How do you rate your ICT competence (from 1 = "very low" to 5 = "very high"; ICT=Information and communications technology)? *

Please choose **only one** of the following:

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5

5 How do you rate your familiarity with creating prototypes (before today; from 1 = "very low" to 5 = "very high")? *

Please choose **only one** of the following:

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5

6 How experienced are you in software development (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

7 How experienced are you in usability evaluations (from 1 = "not at all" to 5 = "very experienced")? *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4

8 Please choose the appropriate response for each item:

	Paper					PDot
Overall I preferred this method		○	○	○	○	○
This method was easier to learn		○	○	○	○	○
This method was easier to use		○	○	○	○	○
This method was less cumbersome		○	○	○	○	○
I felt more confident using this method		○	○	○	○	○
I preferred this method for creating textual feedback		○	○	○	○	○
I preferred this method to create freehand drawings		○	○	○	○	○
I preferred this method when expressing my mood		○	○	○	○	○
I preferred this method to check out the design of others		○	○	○	○	○
I preferred this method to respond to the design of others		○	○	○	○	○
I think this method produced the better result		○	○	○	○	○

9 Please elaborate a little bit on why and for what you preferred which method:

Bibliography

Abras, C., Maloney-Krichmar, D., and Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456.

ACMTOCHI (2016). Re-imagining participatory design. retrieved january 13, 2017 from http://tochi.acm.org/re-imagining-participatory-design/.

Altman, D. G. (1990). *Practical statistics for medical research*. CRC press.

Andriole, S. J. (1989). *Storyboard Prototyping: A New Approach to User Requirements Analysis*. QED Information Sciences, Inc., Wellesley, Massachusetts, USA.

Atterer, R. and Schmidt, A. (2007). Tracking the interaction of users with ajax applications for usability testing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1347–1350. ACM.

Bailey, B. P. and Konstan, J. A. (2003). Are informal tools better?: comparing demais, pencil and paper, and authorware for early multimedia design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 313–320. ACM.

Bergvall-Kåreborn, B. and Ståhlbrost, A. (2008). Participatory design: one step back or two steps forward? In *Proceedings of the tenth anniversary conference on participatory design 2008*, pages 102–111. Indiana University.

Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Bruun, A. and Stage, J. (2014). Barefoot usability evaluations. *Behaviour & Information Technology*, 33(11):1148–1167.

Campos, P. and Nunes, N. (2007). Towards useful and usable interaction design tools: Canonsketch. *Interacting with Computers*, 19(5-6):597–613.

Carmel, E., Whitaker, R. D., and George, J. F. (1993). Pd and joint application design: a transatlantic comparison. *Communications of the ACM*, 36(6):40–48.

Carroll, J. M., Chin, G., Rosson, M. B., and Neale, D. C. (2000). The development of cooperation: five years of participatory design in the virtual school. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '00, pages 239–251, New York, NY, USA. ACM.

Clement, A. and Van den Besselaar, P. (1993). A retrospective look at pd projects. *Communications of the ACM*, 36(6):29–37.

Crewe, K. (2001). The quality of participatory design: The effects of citizen input on the design of the boston southwest corridor. *Journal of the American Planning Association*, 67(4):437–455.

Danielsson, K., Naghsh, A. M., Gumm, D., and Warr, A. (2008). Distributed participatory design. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 3953–3956. ACM.

De Angeli, A., Bordin, S., and Blanco, M. M. (2014). Infrastructuring participatory development in information technology. In *Proceedings of the 13th Participatory Design Conference: Research Papers - Volume 1*, PDC '14, pages 11–20, New York, NY, USA. ACM.

De Jong, T., Sotiriou, S., and Gillet, D. (2014). Innovations in stem education: the go-lab federation of online labs. *Smart Learning Environments*, 1(1):3.

Druin, A. (1999). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 592–599. ACM.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, 21(1):1–25.

Druin, A., Bederson, B. B., Hourcade, J. P., Sherman, L., Revelle, G., Platner, M., and Weng, S. (2001). Designing a digital library for young children. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '01, pages 398–405, New York, NY, USA. ACM.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5):323–327.

Fischer, G. (2013). Meta-design: Empowering all stakeholder as codesigners. *Handbook on Design in Educational Computing*, pages 135–145.

Fleiss, J. L. and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, pages 613–619.

Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical methods for rates and proportions*. John Wiley & Sons.

Forlano, L. and Mathew, A. (2014). From design fiction to design friction: Speculative and participatory design of values-embedded urban technology. *Journal of Urban Technology*, 21(4):7–24.

Francis, P., Balbo, S., and Firth, L. (2009). Towards co-design with users who have autism spectrum disorders. *Universal Access in the Information Society*, 8(3):123–135.

Garrett, J. J. (2010). *The elements of user experience: User-centered design for the Web and beyond*. Pearson Education.

Gartner, J. and Wagner, I. (1996). Mapping actors and agendas: Political frameworks of systems design and participation. *Human-Computer Interaction*, 11(3):187–214.

Gennari, J. H. and Reddy, M. (2000). Participatory design and an eligibility screening tool. In *Proceedings of the AMIA symposium*, pages 290–294. American Medical Informatics Association.

Govaerts, S., Cao, Y., Vozniuk, A., Holzer, A., Zutin, D. G., Ruiz, E. S. C., Bollen, L., Manske, S., Faltin, N., Salzmann, C., Tsourlidaki, E., and Gillet, D. (2013). *Towards an Online Lab Portal for Inquiry-Based STEM Learning at School*, pages 244–253. Springer Berlin Heidelberg, Berlin, Heidelberg.

Greenbaum, J. (1993). A design of one's own: towards participatory design in the united states. *Participatory design: Principles and practices*, pages 27–37.

Greenbaum, J. and Kyng, M., editors (1991). *Design at Work: Cooperative Design of Computer Systems*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Gregory, J. (2003). Scandinavian approaches to participatory design. *International Journal of Engineering Education*, 19(1):62–74.

Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., and Farber, A. (2004). Mixing ideas: a new technique for working with young children as design partners. In *Proceedings of the 2004 conference on Interaction design and children: building a community*, pages 35–42. ACM.

Gumm, D. C. (2006). Distributed participatory design: An inherent paradoxon? *Proceedings of IRIS29*.

Guo, F., Zhang, C., and Cui, L. (2007). Sketching interfaces for remote collaboration. In *2007 11th International Conference on Computer Supported Cooperative Work in Design*, pages 63–68.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23–34.

Halskov, K. and Hansen, N. B. (2015). The diversity of participatory design research practice at pdc 2002-2012. *International Journal of Human-Computer Studies*, 74:81 – 92.

Hassenzahl, M. and Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction*, 25(3):235–260.

Heintz, M. and Law, E. L.-C. (2015). Solution-based requirements capture with pdot in an e-learning context. In *15th IFIP TC. 13 International Conference on Human-Computer Interaction, INTERACT 2015, online http://ivu.di.uniba.it/LandscapeOfUXRequirements/.*

Heintz, M. and Law, E. L.-C. (2016). Challenges and resolutions for engaging teachers and students in participatory design of online science learning resources. In *Proc. of Fourth Workshop on Cultures of Participation in the Digital Age, CoPDA'16, Gothenburg, Sweden, October 23, 2016, CEUR-WS.org, online CEUR-WS.org/Vol-1776/paper7.pdf.*

Heintz, M., Law, E. L.-C., Govaerts, S., Holzer, A., and Gillet, D. (2014a). Pdot: participatory design online tool. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2581–2586. ACM.

Heintz, M., Law, E. L.-C., and Heintz, S. (2014b). Review of online tools for asynchronous distribute online participatory design. In Renzel, D. and Klamma, R., editors, *"Large-Scale Social Requirements Engineering", IEEE Computer Society Special Technical Community on Social Networking E-Letter, vol. 2, no. 3, September 2014.*

Heintz, M., Law, E. L.-C., Manoli, C., Zacharia, Z., and van Riesen, S. A. (2015a). A survey on the usage of online labs in science education: Challenges and implications. In *Global Engineering Education Conference (EDUCON), 2015 IEEE*, pages 827–835. IEEE.

Heintz, M., Law, E. L.-C., and Soleimani, S. (2015b). *Paper or Pixel? Comparing Paper- and Tool-Based Participatory Design Approaches*, pages 501–517. Springer International Publishing, Cham.

Helps, R. and Helps, C. (2016). Measuring stylus and tablet performance for usability in sketching. In *Proceedings of the 5th Annual Conference on Research in Information Technology*, RIIT '16, pages 19–24, New York, NY, USA. ACM.

Hendriks, N., Huybrechts, L., Wilkinson, A., and Slegers, K. (2014). Challenges in doing participatory design with people with dementia. In *Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium Papers, and Keynote Abstracts - Volume 2*, PDC '14, pages 33–36, New York, NY, USA. ACM.

Hornbæk, K. (2008). Usability evaluation as idea generation. In *Maturing Usability*, pages 267–286. Springer.

Hornbaek, K. and Stage, J. (2006). The interplay between usability evaluation and user interaction design. *International Journal of Human-Computer Interaction*, 21(2):117–123.

Hundhausen, C., Trent, S., Balkar, A., and Nuur, M. (2008). The design and experimental evaluation of a tool to support the construction and wizard-of-oz testing of low fidelity prototypes. In *2008 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 86–90. IEEE.

Kelly, S. R., Mazzone, E., Horton, M., and Read, J. C. (2006). Bluebells: a design method for child-centred product development. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pages 361–368. ACM.

Kensing, F. and Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, 7(3):167–185.

Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., and Möller, S. (2011). I'm home: Defining and evaluating a gesture set for smart-home control. *International Journal of Human-Computer Studies*, 69(11):693 – 704.

Kindred, J. and Mohammed, S. N. (2005). "he will crush you like an academic ninja!": Exploring teacher ratings on ratemyprofessors.com. *Journal of Computer-Mediated Communication*, 10(3):00–00.

Klemmer, S., Newman, M., Farrell, F., Meza, R., and Landay, J. A. (2000). A tangible difference: Participatory design studies informing a designers' outpost.

In *CSCW2000 Workshop on Shared Environments to Support Face-to-Face Collaboration*.

Könings, K. D., Brand-Gruwel, S., and van Merriënboer, J. J. (2010). An approach to participatory instructional design in secondary education: an exploratory study. *Educational Research*, 52(1):45–59.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage.

Landay, J. A. and Myers, B. A. (1995). Interactive sketching for the early stages of user interface design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 43–50, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Landay, J. A. and Myers, B. A. (2001). Sketching interfaces: toward more human interface design. *Computer*, 34(3):56–64.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Law, E. L.-C., van Schaik, P., and Roto, V. (2014). Attitudes towards user experience (ux) measurement. *International Journal of Human-Computer Studies*, 72(6):526–541.

Lin, J., Newman, M. W., Hong, J. I., and Landay, J. A. (2000). Denim: finding a tighter fit between tools and practice for web site design. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 510–517. ACM.

Lin, J., Thomsen, M., and Landay, J. A. (2002). A visual language for sketching large and complex interactive designs. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 307–314. ACM.

Liu, Z. and Stork, D. G. (2000). Is paperless really more? *Communications of the ACM*, 43(11):94–97.

Lohmann, S., Ziegler, J., and Heim, P. (2008). Involving end users in distributed requirements engineering. In *Engineering Interactive Systems*, pages 221–228. Springer.

MacDonald, F. and Miller, J. (1998). A comparison of tool-based and paper-based software inspection. *Empirical Software Engineering*, 3(3):233–253.

Madden, A., Ruthven, I., and McMenemy, D. (2013). A classification scheme for content analyses of youtube video comments. *Journal of documentation*, 69(5):693–714.

Maltby, J. and Day, L. (2002). *Early success in statistics*. Pearson Education.

Marois, L., Viallet, J.-E., Poirier, F., and Chauvin, C. (2010). Experimenting introductory tools for innovation and participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*, PDC '10, pages 259–262, New York, NY, USA. ACM.

Molich, R., Jeffries, R., and Dumas, J. S. (2007). Making usability recommendations useful and usable. *Journal of Usability Studies*, 2(4):162–179.

Monteiro, P., Borges, P., Machado, R. J., and Ribeiro, P. (2012). A reduced set of rup roles to small software development teams. In *Proceedings of the International Conference on Software and System Process*, pages 190–199. IEEE Press.

Moraveji, N., Li, J., Ding, J., O'Kelley, P., and Woolf, S. (2007). Comicboarding: Using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1371–1374, New York, NY, USA. ACM.

Muller, M. J. (1991). Pictive-an exploration in participatory design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 225–231. ACM.

Muller, M. J. (2007). *Participatory design: The third space in HCI*. Erlbaum, Mahway, NJ, USA.

Muller, M. J., Wildman, D. M., and White, E. A. (1993). Taxonomy of pd practices: A brief practitioner's guide. *Communications of the ACM*, 36(6):26–28.

Naghsh, A. M. and Andy, D. (2004). Gabbeh: a tool to support collaboration in electronic paper prototyping. In *CSCW 2004 the ACM Conference on Computer Supported Cooperative Work, Chicago, USA*.

Neustaedter, C. and Bernheim Brush, A. J. (2006). Linc-ing the family: the participatory design of an inkable family calendar. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 141–150. ACM.

Newman, M. W., Lin, J., Hong, J. I., and Landay, J. A. (2003). Denim: An informal web site design tool inspired by observations of practice. *Human-Computer Interaction*, 18(3):259–324.

Nielsen, J. (1994a). Heuristic evaluation. *Usability inspection methods*, 17(1):25–62.

Nielsen, J. (1994b). *Usability engineering*. Elsevier.

Porat, T. and Tractinsky, N. (2012). It's a pleasure buying here: The effects of web-store design on consumers' emotions and attitudes. *Human–Computer Interaction*, 27(3):235–276.

Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M. E., Alcorn, A., Wass, S., Bernadini, S., Avramides, K., Keay-Bright, W., Chen, J., Waller, A., Guldberg, K., Good, J., and Lemon, O. (2012). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, 16(2):117–127.

Puri, S. K., Byrne, E., Nhampossa, J. L., and Quraishi, Z. B. (2004). Contextuality of participation in is design: A developing country perspective. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1*, PDC 04, pages 42–52, New York, NY, USA. ACM.

Read, J., Gregory, P., MacFarlane, S., McManus, B., Gray, P., and Patel, R. (2002). An investigation of participatory design with children-informant, balanced and facilitated design. In *Interaction design and Children*, pages 53–64. Shaker Maastricht.

Read, J. C., Fitton, D., Sim, G., and Horton, M. (2016). How ideas make it through to designs: Process and practice. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, pages 16:1–16:10, New York, NY, USA. ACM.

Rettig, M. (1994). Prototyping for tiny fingers. *Communications of the ACM*, 37(4):21–27.

Rogers, Y., Sharp, H., and Preece, J. (2011). Interaction design: beyond human-computer interaction.

Sanders, E. B.-N. (2006). Design research in 2006. *Design Research Quarterly*, 1(1).

Sanders, E. B.-N. (2008). On modeling: An evolving map of design practice and design research. *interactions*, 15(6):13–17.

Sanders, E. B.-N., Brandt, E., and Binder, T. (2010). A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*, pages 195–198. ACM.

Scariot, C. A., Heemann, A., and Padovani, S. (2012). Understanding the collaborative-participatory design. *Work*, 41(Supplement 1):2701–2705.

Schuler, D. and Namioka, A. (1993). *Participatory design: Principles and practices.* CRC Press.

Segura, V. C., Barbosa, S. D., and Simões, F. P. (2012). Uiskei: a sketch-based prototyping tool for defining and evaluating user interface behavior. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 18–25. ACM.

Sellen, A. J. and Harper, R. H. (2003). *The myth of the paperless office.* MIT press.

Snyder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces.* Morgan Kaufmann.

Stolterman, E. (2008). A design research map. retrieved january 16, 2017 from http://transground.blogspot.co.uk/2008/11/design-research-map.html.

Stumpf, S., Rajaram, V., Li, L., Burnett, M., Dietterich, T., Sullivan, E., Drummond, R., and Herlocker, J. (2007). Toward harnessing user feedback for machine learning. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 82–91. ACM.

Sundar, S. S., Oh, J., Bellur, S., Jia, H., and Kim, H.-S. (2012). Interactivity as self-expression: a field experiment with customization and blogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 395–404. ACM.

Svanaes, D. and Seland, G. (2004). Putting the users center stage: Role playing and low-fi prototyping enable end users to design mobile systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 479–486, New York, NY, USA. ACM.

Teo, H.-H., Oh, L.-B., Liu, C., and Wei, K.-K. (2003). An empirical study of the effects of interactivity on web user attitude. *International journal of human-computer studies*, 58(3):281–305.

Van de Kant, M., Wilson, S., Bekker, M., Johnson, H., and Johnson, P. (1998). Patchwork: A software tool for early design. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, CHI '98, pages 221–222, New York, NY, USA. ACM.

Vilbergsdottir, S. G., Hvannberg, E. T., and Law, E. L.-C. (2014). Assessing the reliability, validity and acceptance of a classification scheme of usability problems (cup). *Journal of Systems and Software*, 87:18–37.

Walker, M., Takayama, L., and Landay, J. A. (2002). High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(5):661–665.

Walsh, G. (2011). Distributed participatory design. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1061–1064. ACM.

Walsh, G., Druin, A., Guha, M. L., Bonsignore, E., Foss, E., Yip, J. C., Golub, E., Clegg, T., Brown, Q., Brewer, R., et al. (2012). Disco: a co-design online tool for asynchronous distributed child and adult design partners. In *Proceedings of the 11th International Conference on Interaction Design and Children*, pages 11–19. ACM.

Walsh, G., Druin, A., Guha, M. L., Foss, E., Golub, E., Hatley, L., Bonsignore, E., and Franckel, S. (2010). Layered elaboration: a new technique for co-design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1237–1240. ACM.

Walsh, G. and Foss, E. (2015). A case for intergenerational distributed co-design: The online kidsteam example. In *Proceedings of the 14th International Conference on Interaction Design and Children*, pages 99–108. ACM.

Walsh, G., Foss, E., Yip, J., and Druin, A. (2013a). Facit pd: a framework for analysis and creation of intergenerational techniques for participatory design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2893–2902. ACM.

Walsh, G., Foss, E., Yip, J., and Druin, A. (2013b). Octoract: An eight-dimensional framework for intergenerational participatory design techniques.

Weibel, N., Signer, B., Norrie, M. C., Hofstetter, H., Jetter, H.-C., and Reiterer, H. (2011). Papersketch: a paper-digital collaborative remote sketching tool. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 155–164. ACM.

Wixon, D. (2003). Evaluating usability methods: why the current literature fails the practitioner. *interactions*, 10(4):28–34.

Wong, Y. Y. (1992). Rough and ready prototypes: Lessons from graphic design. In *Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 83–84, New York, NY, USA. ACM.

Zhao, L. and Lu, Y. (2012). Enhancing perceived interactivity through network externalities: An empirical study on micro-blogging service satisfaction and continuance intention. *Decision Support Systems*, 53(4):825–834.