Understanding Innovation

Hasso Plattner
Christoph Meinel
Larry Leifer *Editors*

Design Thinking Research

Building Innovators

Springer

Understanding Innovation

Series Editors
Christoph Meinel
Larry Leifer

For further volumes:
http://www.springer.com/series/8802

Hasso Plattner • Christoph Meinel • Larry Leifer
Editors

Design Thinking Research

Building Innovators

Springer

Editors
Hasso Plattner
Christoph Meinel
Hasso Plattner Institute for Software
Systems Engineering
Campus Griebnitzsee
Potsdam
Germany

Larry Leifer
Center for Design Research (CDR)
Stanford University
Stanford
CA, USA

ISSN 2197-5752 ISSN 2197-5760 (electronic)
ISBN 978-3-319-06822-0 ISBN 978-3-319-06823-7 (eBook)
DOI 10.1007/978-3-319-06823-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014943108

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The ever-increasing complexity of today's world poses equally daunting challenges for all kind of organizations in business and society. These challenges can range from fast changing and growing customer demands, with increased pressure on companies to innovate, to fundamental social problems, such as securing a water supply in third world countries. All of these issues have one thing in common: in order to come up with innovative solutions, we need a fundamental shift in the way we address problems and approach challenges.

In recent years, many people and organizations have discovered the innovative power of Design Thinking. This method combines users' perspectives, technological feasibility, and business perspectives to work out innovative solutions beyond the typical expectation. And it works—as I witness myself in everyday life.

Finding out how and why Design Thinking works and systematically assessing it as an innovation approach is the intention behind the HPI-Stanford Design Thinking Research Program (DTRP). The DTRP was started in 2008 as a joint venture between Stanford University in Palo Alto, California, USA, and the Hasso Plattner Institute for Software Systems Engineering in Potsdam, Germany, and allows multidisciplinary research teams to investigate existing frameworks, tools, systems, and successful practice methods, while at the same time creating new ones. The aim is to scientifically understand basics and principles of Design Thinking, why and how this method works and also the reasons when it doesn't. In this new volume of the Springer-series "Understanding Innovation," the research teams' findings—the result of the fifth program year—have again been compiled in order to share the latest scientific insights with the interested public.

The researchers themselves are endowed with diverse backgrounds in fields, such as engineering, neurology, social science, or economics. They study topics dealing with the complex interaction between Design Thinking team members, creativity building and data handling, as well as conducting long-term field studies in real business environments.

Design Thinking starts at a point that too often tends to be forgotten or ignored: the genuine understanding of users and their expectations. Thus, the core strength of this method is that it offers space to explore and discover user insights. By developing strong empathy through interviews and observations, the real needs of the clients are revealed. These can then be addressed to come up with innovative and desirable ideas, which are subsequently prototyped and tested in iterative circles. Design Thinking provides the framework, tools, and mind-set to create breakthrough ideas, inspired by a deep understanding of the user's needs. All that's required is an open mind.

Design Thinking has the power to transform the way we work by transforming the way we think, approach problems, and develop products and services or even processes and strategies. Ultimately, Design Thinking has the power to transform a whole organization and make it sustainably innovative and fit for the future. It is my hope and belief that through the findings of our Design Thinking Research Program we can contribute to a better understanding of this method's functioning—and its further dissemination in companies and society.

Palo Alto, CA
Winter 2013/2014

Hasso Plattner

Contents

Part III Supporting Information Transfer

Introduction – Design Thinking Is Mainly About Building Innovators

Christoph Meinel and Larry Leifer

1 Is It Really About Building People?

There is mounting evidence that the engineering design thinking paradigm works when applied with diligence and insight, but is it really only about products and services? While profits are typically associated with goods and services, we really must ask, *who* made that happen? Who was responsible for their conception and implementation? Are we too pre-occupied with the innovation when the real story is about the innovators?

Design thinking is mainly about building innovators who can use the design thinking paradigm to transform ideas into reality, to transform organization, and to transform all aspects of life.

When hunting for the "next big idea" the journey to the solution is initially undefined. Every hunt has its unique path, and those who take it learn and discover the unknown. They have to find their way by reading the context, observing and interpreting the signals, understanding and making choices. People indeed face many challenges during their innovation journey. The path is constantly changing as are the activities and roles people play. Thus, we have to find out how are and how can people be best prepared and equipped for a successful journey.

In this volume, we seek to re-focus the attention of the reader on the human innovators. While the design thinking paradigm has always been about people, we are often distracted by the pursuit of "big product ideas" versus "big people." In this phrasing, we seek to get beyond ideas to the creative diligence required to transform

C. Meinel (✉)
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, Potsdam, Germany
e-mail: meinel@hpi.uni-potsdam.de

L. Leifer
Stanford Center for Design Research, Stanford University, Panama Mall 424, Stanford 94305-2232, CA, USA
e-mail: leifer@cdr.stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_1, © Springer International Publishing Switzerland 2015

ideas into realities: real companies, real products and services and real organizational transformation. Building on Volume 4 of this series, “Building Innovation Eco-Systems”,[1] we will retain the structure of Design Thinking Rules while transforming the conversation to focus on the design requirements for the people we build.

2 What Are the Rules for Building Design Thinkers Who Innovate?

We have evidence (Ju et al. 2014) supporting the role of several design thinking activities that have long been considered important, but were too often perceived through the lens of the product and service versus those who create them. Of these, the over-arching truth lies in the fact that every physical product and/or service is actually owned by the people who make it a reality.

The “***rules of design thinking***” are actually the “***design requirements***” for the behavior of innovators. The challenge and goal of this introductory section is to formulate some new rules for design thinking and to translate them into the design requirements for building innovators.

I. The Human Rule All innovator activity is ultimately social in nature. Never go hunting alone.

Our studies substantiate the assertion that successful innovation through design thinking will always bring us back to the “***human-centric point of view.***” This is the imperative to solve technical problems in ways that satisfy human needs and acknowledge the human element in all technologies and organizations. The innovators we build must have and implement this core value and behavior.

To find “big ideas” we have to learn how to hunt again. Hunting is all about the people we hunt with. However, we are in a system that is based on individuals, just as education is focused on individuals. But a team is necessary. Go hunting with a team that is diverse and agile. People are the most valuable asset in the design process.

Innovator Design Requirements for the Human Rule

Be aware that every human eco-system is unique, as is every business scenario. Thus, observe and document your context carefully. Where are you hunting? Deeply internalize to keep people at the center of all things:

- Cover the walls with images of people you seek to actualize. Celebrate their successes and failures.
- Preserve the “human scale.” Forget the organization scale and focus on the innovation team—typically 3–4 core individuals who are co-creating over time.

[1] Meinel and Leifer (2013).

- Envision how the last big innovator in your eco-system delivered winning products and services with “empathy-in-action.”
- Strive to become an expert, maybe an Olympian example of empathy-in-action; for yourself and others.[2]

II. The Ambiguity Rule Innovators must preserve ambiguity. Never go home with a lone idea.

There is no chance for your organization to “***discover***” your contribution if you only have one idea. Innovation demands experimentation at the limits of your knowledge, at the limits of your ability to control events, and with the freedom for you to see things differently. The innovators we build must always be in a rebuilding mode.

The hunting path for the “big idea” might be long and the ambiguity sometimes frustrating—but we need ambiguity. This is how we design possibilities to create alternative futures. We want a future with more ambiguity and more options. Keep hunting with ambiguity—the “next big idea” is just around the corner.

Innovator Requirements for the Ambiguity Rule

- Keep track of assumptions.
- Place them boldly in your design space for every constraint you are coping with.
- List a competing opportunity.
- Check your thinking: are you looking for the global fix, or, are you keenly aware that most everything in design and business is context dependent.
- Take time to define the problem and solutions space context.
- Understand the user.

III. The Re-design Rule All innovation is re-innovation. Who is the innovator that preceded you?

The human needs that we seek to satisfy have been with us for millennia. When looking to the future it is always helpful to look to the past. How did people hunt in the past? Try to understand them, learn from them. Never leave them out of your consideration. Through time and evolution there have been many provisionally successful innovators. Do you know who they are and how they got there? Because technology and social circumstances change constantly, it is imperative to understand how needs have been addressed in the past and by whom. Then we can apply “***foresight tools and methods***” to better estimate the social and technical conditions we will encounter 5, 10, 20 years from now.

Innovator Requirements for the Re-design Rule

Hunting is hard work. Taking it home is harder and more dangerous. Nothing beats a prepared mind.

- Be sure your team is well informed about the history of organizational change and context. How did others effect change? How did they circumnavigate the skeptics? In which ways did they satisfy needs?

[2] Kress (2012).

- List the pros and cons—concentrate on the former.
- Take advantage of foresight thinking tools and the foresight playbook.[3]

IV. The Tangible Rule Make innovation tangible. Make your "innovator story" tangible.

Communication within the hunting team is crucial—being tangible is essential because we have to learn rapidly in order to produce well. Make ideas tangible and learn from them. Communicate via prototypes. Conceptual prototyping has been a central activity in design thinking during the entire period of our research, yet it is only in the past few years that we have come to realize that "***prototypes are tangible stories***." Seen as stories, we now have fresh insights regarding the nature of their structure, their narrative and the suspense and surprise they deliver. We are also mindful of the listener's context, the user. The "***make it tangible***" rule becomes, "***make it a good story***."

Innovator Requirements for the Tangible Rule

There are more great ideas out there in the world than those inside our heads.

- Put differently, searching in the world tangibly is a great way to get new ideas, unplanned associations, undreamed metaphors and serendipity squared.
- Show me, don't tell me.[4]

We have summarized, and in some cases paraphrased, the design requirements in the following table. Take the framework and apply it to your project, your organization, and your team. This is not a tool of physics. Everything about it is context dependent. Define your context.

Innovator design requirements

Requirement	Context	Metric	Rationale
The Human Rule: All innovator activity is ultimately social in nature. Never go hunting alone	Every human eco-system is unique. Every business scenario is unique. Take time to observe and document your context. Where are you hunting?	Count the people in your framework. Count your team's linkages. Cover the walls with images of your team, the users, and their team. Count their success. Count their failures. Count the innovators	Capture the narrative about how the last big innovator in your eco-system managed "empathy-in-action" to deliver winning products and services that addressed user needs in compelling ways
The Ambiguity Rule: Innovators must preserve ambiguity. Never go home with just one idea	Check your thinking; are you looking for the global fix, or are you keenly aware that most everything in	Count the last innovator's sense of assumptions, opportunities, and constraints. How	Did the last innovator in your segment really use ambiguity to afford creativity? Did that innovator "get"

(continued)

[3] Carleton and Cockayne (2013).

[4] Edelman (2012).

Requirement	Context	Metric	Rationale
	design and business is "who dependent." Are you really thinking like the customer you seek to take home?	many ways did they define them?	"creative self efficacy"[5]
The Re-design Rule: All innovation is re-innovation. Who is the innovator that preceded you?	Most human needs have been satisfied before. Who did the last innovation? How did they map the foreseeable future? Understand past hunters and the hunted	Count the number of ways this need has been satisfied in the past. Enumerate the pros and cons. Position your team to absolutely nail just one of the cons without losing the pros	Foresightful innovations tend to last. Understanding the past prepares you for the future. Never leave home without it[6]
The Tangible Rule: Make innovation tangible. Make your "innovator story" tangible	There are more great ideas out there in the world than those inside the head of the last innovator. Searching tangibly is a great way to learn from those who have already done so	Count tangible encounters. Make note of who they were with. Who was that innovator? Is their picture on your wall?	Show me, don't tell me[7]

3 The HPI-Stanford Design Thinking Research Program

Started in 2008, the HPI-Stanford Design Thinking Research Program (DTRP) between Hasso Plattner Institute for Software Systems Engineering and Stanford University is financed and supported by the Hasso Plattner Foundation.

3.1 Program Vision

The HPI-Stanford Design Thinking Research Program engages multidisciplinary research teams to scientifically investigate the phenomena of innovation in all its holistic dimensions. Researchers are especially encouraged to develop ambitious,

[5] Albert Bandura: http://en.wikipedia.org/wiki/Albert_Bandura, Carleton et al. (2008).

[6] Carleton and Cockayne (2013).

[7] Edelman (2012), Lübbe (2011).

long-term explorations related to the innovation method of design thinking in its technical, business, and human aspects. The program strives to apply rigorous academic methods to understand the scientific basis for how and why the innovation method of design thinking works and fails.

Researchers in the program study, for example, the complex interaction between members of multi-disciplinary teams challenged to deliver design innovations. The need for creative collaboration across spatial, temporal, and cultural boundaries is an important feature of the domain. In the context of disciplinary diversity researchers explore how design thinking methods mesh with traditional engineering and management approaches, specifically, why the structure of successful design teams differs substantially from traditional corporate structures. The overall goal of the program is to discover metrics that determine the success of challenges approached with design thinking methods. A special interest of the program is to explore the use of design thinking in the field of Information Technology and IT systems engineering.

3.2 Program Priorities

The focus of the Design Thinking Research Program is the collaboration between researchers at Stanford University, USA, and those at Hasso Plattner Institute in Potsdam, Germany. Projects that set new research priorities for this emergent knowledge domain are favorably funded. Furthermore, in this context, field studies in real business environments are considered especially important to assess the impact and/or needed transformations of design thinking in organizations. Project selection is also based on intellectual merit and evidence of open collaboration.

Special interest lies in the following points of view and guiding questions:

- What are people really THINKING and DOING when they are engaged in creative design innovation? How can new frameworks, tools, systems, and methods augment, capture, and reuse successful practices?
- What is the IMPACT of design thinking on human, business, and technology performance? How do the tools, systems, and methods really work to create the right innovation at the right time? How do they fail?

3.3 Road Map Through This Book

Divided into three parts, this book compiles the outcomes of the 5th year’s projects which have again covered diverse facets of design thinking.

Aspects such as empathy, creativity, personality, culture, and people’s actions in their context, play a significant role when approaching challenges with design thinking. Thus, the chapters in Part I, "***Assessing Influential Factors in Design Thinking***," examine the impact of those factors on design thinking and vice versa.

Design thinking only works in teams. Collaboration is essential for innovative outcomes. Part II, "***Empowering Team Collaboration***," presents insights on how to support teams in their design work.

The question on how to optimally ensure knowledge transfer and avoid information loss during the innovation process and afterwards is addressed in the last part, "***Supporting Information Transfer***."

3.4 Part I: Assessing Influential Factors in Design Thinking

In "**Empathy via Design Thinking: Creation of Sense and Knowledge**," the authors Eva Köppen and Christoph Meinel assess the growing demand to be empathic that can be witnessed in organization studies and management advice literature; a requirement not only for leadership but also for the whole staff. Design thinking has ultimately provided methods and techniques for fostering empathy in teamwork settings. With the help of a study, the article addresses the question of whether design thinking indeed delivers helpful empathy-techniques that will assist employees in their daily routine.

Creativity stands in the focus of "**Developing Novel Methods to Assess Long-Term Sustainability of Creative Capacity Building and Applied Creativity**." The team of Manish Saggar, Grace Hawthorne, Eve-Marie Quintin, Nick Bott, Eliza Keinitz, Ning Lui, Yin-Hsuan Chien, Daniel Hong, Adam Royalty, and Allan L Reiss, investigates the ability to create novel and useful outcomes, which has been widely recognized as an essential skill for both entrepreneurial and every-day success. Their research proposes to examine the impact and sustainability of creative capacity building using targeted training.

Design thinking asserts that individuals and teams have the ability to build their innovative capacity through various tools and methods no matter their predispositions to creativity and innovation. The contexts of design thinking attempt to alter design process towards more innovative ideas. "**The Personal Trait Myth: A Comparative Analysis of the Innovation Impact of Design Thinking Tools and Personal Traits**," by Nikolas Martelaro, Shameek Ganguly, Martin Steinert, and Malte Jung attempts to experimentally disentangle the impact of disposition and situation during design activity. The authors present a variety of design contexts intended to be tested against dispositional factors during an experimental design task. They then present a pilot study exploring how process-priming impacts design process during a problem-solving task and an open-ended design task.

In "**Theaters of Alternative Industry: Hobbyist Repair Collectives and the Legacy of the 1960s American Counterculture**," Daniela K. Rosner and Fred Turner describe initial results from an ethnographic study of design and engineering engagements in community-operated sites at which hobbyists mend and repair mass-produced goods. They conducted participant observation at seven repair events and two collectives in the San Francisco Bay area where consumer electronics are reassembled. In their study they spoke with approximately eighty repair practitioners. Here they describe surprising connections between repair and social movements that, in turn, reveal deep ties between contemporary hobbyist repair and countercultural design practices of the 1960s. These links, they argue, open new and important areas for design research.

3.5 Part II: Empowering Team Collaboration

Increasingly organizations are turning to off-site design thinking professional development programs as a way to grow design competencies in their workforce. Therefore "**Assessing the Development of Design Thinking: From Training to Organizational Application**," by Adam Royalty, Karen Ladenheim, and Bernard Roth has two main goals (1) To develop an initial assessment tool that helps identify how well organizations support employees' continued learning and application of design thinking. (2) To describe a process for constructing design thinking assessment tools. The assessment created is informed by an exploration of existing design thinking executive education programs and tested in a large organization committed to using design thinking.

Joel Sadler and Larry Leifer contributed "**TeamSense: Prototyping Modular Electronics Sensor Systems for Team Biometrics**." Electronic sensors systems can be used to unobtrusively gather real-time measurements of human interaction and biometrics. However, developing custom sensor systems can be costly, time intensive and often requires high technical expertise in embedded mechatronic systems. The authors present a prototyping case study of a real world system, TeamSense, with the scenario of a manager who wishes to use embedded sensors to develop data-driven insights on team performance. Team Biometrics is a term used here to refer to a sensor system that measures some physical characteristic of a group of individuals. This work has broad implications for design thinking and the importance of toolkits in reducing entry barriers for rapid prototyping with sensors.

In "**Tele-Board MED: Supporting Twenty-First-Century: Medicine for Mutual Benefit**," Julia von Thienen, Anja Perlich, and Christoph Meinel present a medical documentation system designed to support patient-doctor cooperation at eye level. In particular, Tele-Board MED tackles the challenge of turning the task of documenting a patient's medical records—which can disturb the treatment flow—into a curative process in and of itself. With its focus on cooperative documentation, Tele-Board MED embraces patient empowerment and, at the same time, the project is deeply rooted in the culture of design thinking. Results from an initial feedback study with 34 behavior psychotherapists are presented.

Peer and self assessment offer an opportunity to scale both assessment and learning to global classrooms. In "**Peer and Self Assessment in Massive Online Classes**," the team of Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer, reports its experiences with two iterations of the first large online class to use peer and self assessment. The team performed three experiments and introduces a data-driven approach that highlights high-variance items for improvement.

In user-centered design processes, one of the most important tasks is to synthesize information from user research into insights and a shared point of view among team members. "**Tagging User Research Data: How to Support the Synthesis of Information in Design Teams**," by Raja Gumienny, Steven Dow, Matthias Wenzel, Lutz Gericke, and Christoph Meinel explores the synthesis process and opportunities for providing computational support. Based on interviews on common practices and challenges of information synthesis, they developed digital whiteboard software for sorting individual segments of user research. Through a case study, they explore the differences between computer-supported group interaction and an individual clustering condition.

3.6 Part III: Supporting Information Transfer

The last chapter in Part II already indicated that information transfer and data handling is crucial for a successful innovation process. Thus the third part of this book picks up the issue of how to avoid losing information and how to properly secure and transfer it.

In "**Embodied Design Improvisation: A Method to Make Tacit Design Knowledge Explicit and Usable**," the authors David M. Sirkin and Wendy Ju present a design generative and evaluative technique that they call embodied design improvisation. It incorporates aspects of storyboarding, Wizard of Oz prototyping, domain expert improvisation, video prototyping and crowd sourced experimentation, to elicit tacit knowledge about embodied experience. They have been developing this technique for their research on physical interaction design over time. On the other hand, practitioners often rely on subtle, shared cues that are difficult to codify, and, as a result, are often left underexplored. Their current technique provides an approach to understanding how everyday objects can transition into mobile, actuated, robotic devices, and prescribes how they should behave while interacting with humans. By codifying and providing an example of this technique, the authors hope to encourage its adoption in other design domains.

Information transfer is explored in "**Connecting Designing and Engineering Activities II**," by Thomas Beyhl and Holger Giese. The transition from designing innovative products or services to implementing them is challenging since innovators and engineers are seldom the same people. A knowledge transfer between both groups is inevitable, but in practice seldom goes smoothly since usually only the final innovative product or service is subject to the handover process. The design path and design decisions need to be recovered later on. The authors introduce their inference engine. It infers the design path and design decisions of design thinkers with the help of their design thinking inference rule set.

The design work for programmers stands in the focus of "**How Cost Reduction in Recovery Improves Performance in Program Design Tasks**," by Bastian Steinert and Robert Hirschfeld. Changing source code often leads to undesired implications, raising the need for recovery actions. Programmers need to manually keep recovery costs low by working in a structured and disciplined manner and regularly performing practices such as testing and versioning. While additional tool support can alleviate this constant need, the question is whether it affects programming performance. The authors present their controlled lab study and their recovery tool called CoExist that makes it possible to easily revert to previous development states and also allows forgoing test runs.

In the last chapter, "**DT@Scrum: Integrating Design Thinking with Software Development Processes**," Franziska Häger, Thomas Kowark, Jens Krüger, Christophe Vetterli, Falk Übernickel, and Matthias Uflacker tackle the problem of what happens when design thinking activities are not properly integrated into production processes, e.g. software development. In this case, handovers become necessary and potentially prevent great ideas from becoming real products. A seamless integration of design thinking into the regular development processes of software development companies is still subject to research. The authors present DT@Scrum, a process model that uses the Scrum framework to integrate design thinking into software development. We are introduced to the results of their experiments as well as possible future applications.

4 Summary

Design thinking is about people. It is about finding innovative solutions for people based on their needs. With this book and the underlying research projects we aim to understand the innovation process of design thinking and the people behind it—the innovators. Discover the unknown and learn. This is not only central in design thinking as a whole, but for our Research Program as well. These contributions shed light on and show deeper insights of how to support the work of design teams in order to systematically and successfully develop innovations and design progressive solutions for tomorrow.

Multi-faceted topics were investigated, studies conducted and experiments conceived. With the help of constant exchange between all research groups, joint workshops and community building activities, the different projects were discussed and enhanced within the research community. By sharing the insights from our research program with you we also invite you to engage in dialogue with us on your ideas, insights, and questions on design thinking. We hope you enjoy and benefit from the content presented and strongly welcome and encourage feedback and further scholary debates. To further deep-dive into design thinking research we invite you to the "Electronic Colloquium on Design Thinking Research" on http://www.ecdtr.hpi-web.de where you can find more materials from the design thinking research community and share your own.

We would like to thank all authors—researchers from the Design Thinking Research Program—for contributing their research results. Additionally we are also thankful to many helping hands from Stanford and HPI who have supported this program with regard to its community building activities and workshops which made this program special and successful, a vivid, inspirational community. Special thanks go to Claudia Koch for preparing this book and supporting the authors and editors as well as Dr. Sharon Nemeth for her constant support in reviewing the chapters.

We strongly hope to inspire our readers with this book and to have contributed to a better understanding of this method. It is our sincere wish that with the help of our findings we might support you in hunting down your big ideas and bringing them home.

References

Carleton T, Cockayne W (2013) Playbook for strategic foresight and innovation. Stanford University, Stanford, CA. http://foresight.stanford.edu/index.html

Carleton T, Cockayne W, Leifer L (2008) An exploratory study about the role of ambiguity during complex problem solving. In: Proceedings for the 2008 Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium, Stanford, CA

Edelman J (2012) Understanding radical breaks: media and behavior in small teams engaged in redesign scenarios. Dissertation, Stanford University, California, http://purl.stanford.edu/ps394dy6131

Ju W, Aquino Shluzas L, Leifer L (2014) People with a paradigm: the Center for Design Research's Contributions to Practice. In: Chakrabarti A, Lindemann U (eds) Impact of design research on practice, Springer

Kress G (2012) The effects of team member intrinsic differences on emergent team dynamics and long-term innovative performance in engineering design teams. Dissertation, Stanford University, California. http://purl.stanford.edu/hm975hz5458

Lübbe A (2011) Tangible business process modeling—design and evaluation of a process model elicitation technique. Dissertation, Hasso Plattner Institute for IT Systems Engineering, University of Potsdam. http://ecdtr.hpi-web.de/static/books/tangible_business_process_modelling/

Meinel C, Leifer L (2013) Introduction. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Building innovation eco-systems. Springer, Berlin, pp 3–10

Part I
Assessing Influential Factors in Design Thinking

Empathy via Design Thinking: Creation of Sense and Knowledge

Eva Köppen and Christoph Meinel

Abstract A growing demand to be empathic can be witnessed in organization studies and management advice literature. This requirement does not only focus on the leadership anymore, but rather on the whole staff. Design Thinking has ultimately provided methods and techniques for fostering empathy in teamwork settings. From these developments two questions arise that shall be addressed by this article: How could empathy have become one of the most important things for the economy today? And second: Does Design Thinking indeed deliver useful empathy-techniques that will help employees in their daily routine? For this study we used a documentary analysis approach. The results show that empathy in organizations is a creator of sense and knowledge, but misconceptions of it may also lead to unintentional costs for employees.

1 Introduction

Empathy has gained much attention in recent years within the realm of management studies and advice literature (see e.g. Leonard and Rayport 1997; Miyashiro 2011; Postma et al. 2012; Pavlovich and Krahnke 2012; Cameron and Spreitzer 2012; Goleman 2003). A frequency analysis showed that the number of empathy-related publications for the area of business and economics has been growing constantly over the past 20 years. The database JSTOR registers more empathy articles in economics and business than in the areas of psychology and philosophy, where the term "empathy" was actually rooted. Why is the concept of empathy suddenly of

E. Köppen (✉) • C. Meinel
HPI-Stanford Design Thinking Research Program, Hasso Plattner Institute for Software Systems Engineering, Prof.-Dr.-Helmert-Street 2-3, 14482 Potsdam, Germany
e-mail: Eva.Koeppen@hpi.uni-potsdam.de; Christoph.Meinel@hpi.uni-potsdam.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_2, © Springer International Publishing Switzerland 2015

interest for the economic sector? And how does Design Thinking contribute to the growing demand to be empathic?

Design theorists as well as practitioners describe empathy as a crucial impact factor of Design Thinking (e.g. Brown 2008; Kouprie and Visser 2009; Kolko 2011; Carlgren et al. 2013; GE Reports 2011). Design Thinking authors are of the opinion that empathic insights are a form of extremely important knowledge that stems from concrete interaction with other people. This knowledge is therefore not the result of a solely analytical process (Grotz and Creuznacher 2012). Indeed, three types of knowledge characterize design (Utterback et al. 2006 in Rylander 2009: 10): technological knowledge, knowledge about user needs, and knowledge about product language (e.g. which signs are to be used to deliver a message to the user and the cultural context in which the user will give meaning to those signs). As will be proved later on, the two last forms of knowledge are rooted in an empathic understanding of other people. In order to achieve this specific knowledge, elaborate strategies are described by Design Thinking.

In this article, we ask what empathy in the context of Design Thinking and organizations actually means. We thereby challenge a positive but rather fuzzy view of it, which can be found in the management texts on empathy. To put it in the words of philosopher Jesse Prinz:

> Empathy is a thick concept, and it connotes praise. But an endorsement of empathy requires more than a warm fuzzy feeling. (Prinz 2011: 214).

We suggest to viewing empathy in organizations via Design Thinking as a form of knowledge construction. The analysis of empathy techniques in Design Thinking will further show that empathy can be divided in two forms: internal and external empathy. The specific techniques in these two areas will be analyzed. Paradox and problematic issues arising from them will be discussed.

We will conclude by (a) suggesting reasons for the important role that empathy plays in contemporary innovation strategies and (b) highlighting why Design Thinking is the answer to this demand by facilitating the integration of empathic techniques in the organizational context and (c) pointing to misleading empathy conceptions that are more likely to be a risk than a solution. A documentary research approach was chosen for this study.

2 What Is Empathy?

We understand the term empathy in its broadest sense as perspective-taking, including both the involuntary act of feeling with someone else as well as the cognitive act of placing oneself into someone else's position and adopting their perspective (see also Köppen et al. 2011). As a basic form of social cognition, empathy is the capacity "to share, to experience the feelings of another person" (Greenson 1960). Empathy is an ability that allows us to comprehend the situations and the perspectives of others, both imaginatively and affectively (Rogers 1975). It is therefore not about how I would feel in the certain situation of the other. Empathy is the attempt to reconstruct the specific perspective of the other and how *he* perceives the situation. The aim of empathy is to construe mutual understanding.

3 Method

In this chapter, we want to create access to the provided empathy techniques as well as to the normative expectations that are raised by Design Thinking regarding the empathic behavior of employees. For various reasons we decided to use a qualitative approach for this work. Firstly, quantitative methods of collecting data in the field of empathy research, such as questionnaires or scales, are generally used in the study of psychopathological groups (e.g. sociopaths, narcissists, people with autism). That means almost no effects arise from these methods for non-clinical groups. Quantitative measurements are also highly problematic because they do not deliver information about the circumstances and challenges of certain interactions in companies (Rastetter 2008: 160). Second, these methods try to measure the actual amount of empathy in people as a static psychological construct, while of interest here are the empathic techniques required by modern work and how Design Thinking delivers a framework and tools for these techniques. From this follows that empathy is not seen as something static within a person but rather as something that changes according to the social situation or context.

For these reasons, a qualitative documentary research approach was chosen for this study. This is a method of observation that analyzes documents and archives of cultures in order to provide a description of, for example, the self-descriptions and agenda levels of organizations (Aronson et al. 2004). These text fragments are a symbolic interaction of organizations with their environment (Rastetter 2008: 167). Our text material consisted of (a) programmatic descriptions of Design Thinking from Design Thinking facilities in companies and “schools of Design Thinking” and (b) descriptions on websites of companies that implement Design Thinking.

The use of textual material stemming from websites has the disadvantage that these materials are not reproducible. Furthermore, they may be changed by the editors of the webpage after the request in carrying out this study. This does not necessarily need to be a problem, for

> (. . .) documents need to be considered as situated products, rather than as fixed and stable ‘things’ in the world. (Prior 2003: 26)

The text fragments were chosen in an open selection process that did not follow a structured approach. The important criterion was that the documents need to show certain discursive similarities, like the modeling of specific empathic practices and conventions about how to work with empathy. Furthermore, the documents needed to demonstrate an analogical vocabulary and follow the same “story line”. A similar structure and a certain line of argument regarding empathy in fact became apparent.

From these traits we were able to extract the distinction between internal and external empathy as will be elaborated later on.

4 Empathy in Design Thinking

Within a modern corporate world, design-driven techniques, intercultural and multidisciplinary teamwork as well as the term "user-centeredness" are becoming more and more relevant. The question has to be raised, which new forms of non-technical, interpersonal knowledge are being created and how they can be managed and carried on. In the course of this development, the working world of the last decades has also witnessed a growing demand for access to personality-bound and emotional capabilities of employees (in the research literature known as "subjectifying" processes in the workplace, see e.g. Schönberger and Springer 2003; Voswinkel 2002). Accordingly, the social skill of empathy has also grown more important for companies, management and advice literature (see e.g. Miyashiro 2011; Postma et al. 2012; Pavlovich and Krahnke 2012; Goleman 2003).

However, there still seems to exist a lack of concrete techniques that facilitate the enhancement of empathy and empathic knowledge of the daily work in companies. At this point, the Design Thinking process can be seen as the attempt of utilizing empathy methods from the realm of design in order to generate empathic perspective taking (a) among team-members and (b) toward the user. Both cases are about generating access to the perspectives of other persons and to create an interpersonal knowledge from these insights that shall be useful in the further development of a product.

The whole Design Thinking process should guide the non-designer, who is supposed to work on creativity-related topics in teams iteratively, from a vague understanding of a problem to an appropriate solution. Design Thinking relies on five iterative working modes: "Empathize" is about exploring the nature of the problem and understanding the users and their needs. The findings of this phase are then categorized in a "Define" step, which synthesizes the main findings and acts as a "persona" (an ideal user) to validate decisions later in the process. The remaining three modes are "Ideate", "Prototype" and "Test". These modes deal with generating ideas that are expressed in prototypes, in order to test them with users, who are close to the persona.

The role of empathy in Design Thinking is not only highlighted by the process itself (remember the first step "Empathize"), but also by studies on Design Thinking. For example, authors like Tim Brown explain that the most important skill for a Design Thinker is to

> (. . .) imagine the world from multiple perspectives – those of colleagues, clients, end users, and customers. (Brown 2008: 87)

Case studies on the use of Design Thinking, as well as self-descriptions from companies, also demonstrate that empathy is the most basic and most desired principle for companies as to why Design Thinking should be implemented:

> In the interviews, it was striking how essentially all interviewees stressed the importance of empathy as part of a mindset, as a way of relating to the customer, and as an outcome of user research. (Carlgren et al. 2013: 13)

> (. . .), design thinking is really about seeing the world through the eyes of people. . . We don't design products for customers, we design experiences for people. (GE Reports 2011)

The set-up of a multidisciplinary team is furthermore seen as a crucial element in Design Thinking:

> The principle of diversity also includes diversity in team members and networks. The importance of teamwork and making teams as diverse as possible were central themes in the interviews. (Carlgren et al. 2013: 13)

For this kind of cooperation, empathy is said to be mandatory. Grotz and Creuznacher (2012: 20) remark that a Design Thinker needs to be empathic because otherwise he will not be able to acknowledge his teammates who probably have other cultural or disciplinary backgrounds. He has to gain empathic knowledge about the strengths and weaknesses of a colleague and needs to know which thoughts or feelings stakeholders have.

Obviously, empathy is of high relevance for the concept of Design Thinking. We now want to dig deeper and look for the meaning of empathy. During our analysis we found that there exist two areas where empathy takes place: in user research and in teamwork. We call the two specific empathy forms external and internal empathy. In the course of the following two sections we will gain a clearer picture about what empathy is by using this division. We will also discuss the respective advantages and weaknesses of both forms.

5 External Empathy

The goal of the empathic approach is to find out what users need. What sounds banal at first, points to a modern understanding of product development: While in the past products evolved from technical progress and intellectual and analytical knowledge work, the production in the Design Thinking paradigm should not start until the hidden wishes and needs of users or customers are analyzed.

The work of a Design Thinker therefore includes an unequivocal customer and user orientation. The highest goal for a Design Thinker is to conceive and design something useful. Whether he has really achieved this goal has to be proven in cooperation with the user himself:

> Empathy for the people you are designing for and feedback from these users is fundamental to good design. (d.school bootcamp bootleg 2011, introduction)

For empathic practice in the daily working routine, three guidelines are given for the successful completion of this empathy requirement. First, there is the observation of users in their "natural environment"—so to speak in the context of their living environment. To find out something about the target group by solely doing a market-oriented analysis is apparently not sufficient anymore. The second aspect is the interviewing of and interaction with the user. Being communicative and gaining access to the social world of the user may still not be part of the traditional curriculum of, for example, a technical education. It nevertheless seems to be an indispensable part of modern creative work. Third, putting oneself in the position of someone else by tracing the experience of that user's world (a classic example is the simulation of being in the situation of elderly and frail people by wearing glasses that are intended for this purpose etc.) can be helpful to foster empathy.

These techniques already give information about how empathy is being understood in this case: not as something that comes to you spontaneously and automatically but as something that can be achieved by an active and conscious focus on the counterpart. It is about gaining knowledge of other people, which means that

> (...) problems you are trying to solve are rarely your own – they are those of particular users. (d.school bootcamp bootleg 2011: 1)

Empathy is possible if one's own perspective is rejected in favor of the observed user. This clearly concentrates on the rather non-spontaneous and more cognitive-analytical aspects of empathy. Empathy functions as a bridge between people and needs to be something that stems from self-reflection and attentive observation of the user.

> Note that thoughts/beliefs and feelings/emotions cannot be observed directly. They must be inferred by paying careful attention to various clues. (d.school bootcamp bootleg 2011: 15, underlined in original)

The term "infer" strongly relates to the analytical skills of a person. The required capabilities do not refer to forms of "emotional resonating" or "emotional contagion". Basically, this ability expresses the mindset of the therapist. These capabilities can also be compared with the viewpoint of a qualitative researcher, who not only takes into consideration what people *say* but also takes into account the ways people *do* things and the implicit meanings of their actions.

In any case, this rather rational empathic approach should be adopted by employees working with Design Thinking in order to unfold hidden patterns of user action via interviews and observation

But interestingly enough, it is also possible to convert problems of others to your own problems in a far more emotional way. For example, with the method of the "bodystorm" the Design Thinker acts out a certain situation in which a user may find herself in order to test how it feels to be the other person. In the words of the Design Thinker:

> What you're focused on here is the way you interact with your environment and the choice you make while in it. (...) We bodystorm to help create empathy in the context of possible solutions for prototyping. (d.school bootcamp bootleg 2011: 31)

The method of the “prototype for empathy” contains a similar background. Prototypical environments are created that are tested to check the insights into the real-life environment of the user that have been fostered so far (d.school bootcamp bootleg 2011: 33). To be able to personally feel oneself into the situation of another person is, of course, far more emotional than some of the cognitive techniques described above. In line with these techniques, another quotation also shows that the affective quality of empathy in Design Thinking plays a role:

> Lose your agenda and let the scene soak into your psyche. Absorb what users say to you, and how they say it, without thinking about the next thing you’re going to say. (d.school bootcamp bootleg 2011: 6)

Contrary to the traditional image of the rational, tactical, controlled employee, Design Thinking pursues the strategy of actively letting go to be able to even better place oneself in another person’s position. These methods for the optimization of personal empathy are based on intuition as well as on the uncontrolled and emotional engaging with the other.

We conclude that even though the former descriptions and recommendations of empathy tend to describe the conscious and controllable components of empathy, the just mentioned method for an enhancement of empathy is applied to one’s intuition and the uncontrolled emotional engagement with the other person. The necessary empathic attitude appears paradoxical because an analytical and controlled position is being intertwined with a spontaneous and unconstrained state of mind.

5.1 Contradictory Requirements

From what has been said so far, we can now derive two aspects about external empathy that might be the source of misconceptions during the integration of Design Thinking:

First, empathy as a technique is something cognitive as well as something emotional. As a requirement, this might be a source of confusion for employees. Should I keep a rational distance or should I get emotionally lost in the situation? When nobody tells them, employees are likely to be frustrated because they don’t know if they are doing things right.

This uncertainty about emotional versus cognitive aspects of empathy is nothing new and can be traced back to scientific studies on empathy. Some scientists claim empathy is an emotion (Pavlovich and Krahnke 2012) some say it’s not a feeling at all (Stein 1980; Prinz 2011). Some divide between cognitive perspective taking and emotional empathy (Geulen 1982; Ekman 2004; Goleman 2003). Others assume that empathy is both: emotional and at the same time cognitive (Bischof-Köhler 1989). So called multi-level-theories are of the opinion that emotional contagion, mimicry and cognitive perspective-taking are all forms of empathy (Davis 2007; de Waal 2011; Rizzolatti et al. 2008; Lamm et al. 2007).

The diverse discussion on empathy has obviously expanded into the Design Thinking paradigm. If organizations want to implement Design Thinking, they should therefore keep in mind that the requirement of being empathic is twofold and not explicit at all. Employees might need support in deciding if they should use their cognitive or emotional skills while building empathy.

Second, depending on the context it can be emotionally difficult and exhausting to actually feel with another person (e.g. a homeless, ill or a suffering person). Studies on "emotional dissonances" resulting from "emotional labour" (Hochschild 2003) or the burnout syndrome (Neckel and Wagner 2013) have shown that "feeling into" another person can cause emotional suffering if the barriers between the own self and the other self are blurred. Managers need to keep in mind that being empathic is not just fun but also a "demanding way of being" (Rogers 1975). For some employees this might result in an extra work load.

5.2 Positive Identity Construction

The perception, documentation and interpretation of the experiences of a user make it possible for the Design Thinker to extract a form of implicit knowledge from these experiences. This is the promise of empathy in Design Thinking. From the hidden knowledge that slumbers in the user and can be dissected by the Design Thinker, really innovative ideas will be designed. For the employee who practices Design Thinking this means that he might find a new meaning in his daily work. He now knows who he is designing for.

> Designers engage with users (people!) to understand their needs and gain insights about their lives. (d.school bootcamp bootleg 2011: 11)

The narrative of empathy in the organization adds meaning to the daily work because it feels better to compose for people with feelings and needs rather than for anonymous and non-defined gray masses. What is more: Because of his empathic skills, the Design Thinker is able to find out needs that the user might not be aware of herself. The identity of the employee is thus strengthened in two ways. With her state of empathic knowledge she knows not only more about the user than the user himself, she also possesses a moral sovereignty which puts her before other the employees of other companies that are not taking into account the "true needs" of the consumers.

6 Internal Empathy

Another important "mindset" that can be found in Design Thinking aims at "radical collaboration". The object of this collaboration is to

> Bring together innovators with varied backgrounds and viewpoints. Enable breakthrough insights and solutions to emerge from diversity. (d.school bootcamp bootleg 2011: 3)

This "mindset" with its focus on multidisciplinary teamwork indirectly implies the requirement of empathy on side of the Design Thinker. If employees with distinct perspectives and backgrounds should "radically cooperate," this means that they have to learn to adjust their own point of view in favor of other perspectives. This is necessary in order to work on a collective solution that arises from a diversity of the team members.

Also, "radical collaboration" necessitates empathy from team members because it is the premise for the acceptance of the perspective of colleagues with different cultural or professional backgrounds. In Design Thinking, no explicit methods are described that focus on this operation area of empathy—maybe it is assumed that the disposition to be empathic within the team is a given.

As an indirect method to optimize empathy within the team, one can consider certain techniques that strengthen the shared identity and team spirit, for example a set of exercises to loosen up, the so called "warm-ups". These exercises may appear bizarre to external observers (d.school bootcamp bootleg 2011: 27), and hence may be the reason why they create a feeling of team spirit.

Apart from those methods that may help to change the team spirit in an ongoing project, there are also techniques that focus on the manipulation of the individual's attitude in order to optimize one's own empathy. One of these techniques is the principle of "building on the ideas of others". A method to generate ideas that relate to this principle allows a person to introduce only one idea. Beyond that she may only optimize or detail the ideas that were expressed by her teammates. In this way, one is forced to deal with the line of thought of another person. This method is used to create a high degree of empathic attention for team members with each other.

Another example is the behavior guideline "defer judgment": It means that colleagues should be perceived, asked and understood without being judged in a normative way. By this, one can create an empathic understanding between the teammates. Another guideline is to acquire a "beginner's mindset", which means that one's own experiences and the expert knowledge of individuals can be intercepted in due course:

> Your assumptions may be misconceptions and stereotypes, and can restrict the amount of real empathy you can build. (d.school bootcamp bootleg 2011: 5)

Interestingly enough, the implicit premise of this phrase is that there is a "real" empathy in contrast to an "unreal" empathy. That means there are different levels of understanding for other people. Empathy in this sense is something that can be enhanced via the reflection of one's own tendency to stereotype. It is useful to be permanently suspicious of one's own perspective and aware of personal prejudices, while remaining open and curious regarding the views of another person. This is the employee as we find him in literature about "subjectifying" in the workplace: The distance towards his own expertise is an important part of the employee's personality and is seen as a characteristic of an empathic personality.

6.1 Empathy or Sympathy?

The difficulty that arises from internal empathy, as described above, results from the thought that an expert—to a certain degree—should reject his own knowledge in favor of the team's decisions. It echoes the assumption that if I am empathic with another person *her* feelings and thoughts are suddenly *my* feelings and thoughts. But this is actually called "emotional contagion", something that occurs if, for example, one finds himself in a cheering crowd and all of a sudden feels happy himself without even knowing why. Transferred to the workplace this would mean that I give up my own opinions about something in order to vote for the team's solution. Superficiality is the obvious dangerous aspect of this "feeling the same way". The positive feeling of "finally we understand each other" is the reward of such a communication (Sennett 2012: 39). If teams relied more on this kind of harmonious cooperation than on their expertise nothing would be gained. A team discussion like this has a dialectic structure: I have an opinion (thesis), you have an opinion (antithesis) and we come together harmoniously in a shared opinion (synthesis). The aim of a dialectic conversation is consistency. That's why this type of teamwork is better expressed by the term "sympathy". Sympathy overcomes separation because in my mind I am trying to identify with you (ibid.: 38).

But the aim of empathy is not consistency and identification. It's mutual understanding. To gain this form of understanding, one has to be a careful listener and one has to accept the "otherness of others." While one has to be able to feel into the uniqueness of a person—it is precisely because the other is so unique that it will never be possible to simulate his feelings or thoughts in exactly the same way. The challenge is to understand him as fully as possible as an individual, rather than by empathizing with his inner experiences exactly. A conversation like this is marked by a strong emphasis on listening and discussing and not by consensus. Its structure is called dialogic and not dialectic (ibid.: 36). The required mindset is not so much described in terms of "I want to feel what you feel" but rather with the sentence "I'm curious to hear what you feel".

If this distinction becomes clear, people will not be forced to act like "beginners," because they have the right to stay who they are (experts, members of other cultures etc.). If they are open to other opinions and are able to listen carefully they may at the same time maintain their expert status. A beginner's mindset might on the contrary hinder them in their empathic cooperation.

6.2 Solidarity

The sociological work on the "subjectifying" of the working world conducted in recent years has shown that people are suffering more and more from the "competitive" atmosphere in their workplaces (Voß et al. 2013). The reasons for this are numerous: the introduction of excessive flexibility and the increased dismantling of hierarchical structures in the contemporary economy. Both lead to more freedom for the individual but also to more responsibility regarding one's own work and career. Many employees feel like they are on their own and have to fight against other competitors. This can lead to the feeling of insecurity or even burnout syndromes (Neckel and Wagner 2013).

The concept of internal empathy might provide a solution to this. Because empathic cooperation plays such a crucial role, the responsibility will be distributed on a team level. This means that it is not just one single person who will need to guarantee the success of a project or parts of a project. Not the individual, but the team is in charge. New forms of solidarity can arise from this "radical cooperation" that will counteract tendencies of isolation and separation.

7 Conclusion and Outlook

The first of the two initial questions examined the question of why empathy could become so important for the economic area of the western culture. We saw that in general empathy in Design Thinking signifies a modern product development paradigm. In this framework, first the requirements of the user are analyzed then one thinks about technical or financial feasibility. This is an emotion-driven worldview because it is assumed that the access to a person via her emotions is the most important and deepest one. This is because emotions guide behaviors in an unconscious way. But why is knowledge about the inner processes and emotions of users so important nowadays?

Traditional idea management or mere creativity techniques would be sufficient if modern products would only focus on cognitive contents. But this is not the case. In the contemporary economy it is not about innovative ideas that are based on cognitive insights. It is all about association and "esthetic events", which means that products and services are "experienced" in an emotional way (see Reckwitz 2012: 142, translation by the author.). New forms of working aim in their core at "esthetic innovation" and the creation of certain affective perceptions. This is why innovative forms of working need access to the emotionality of people. It is exactly this access that shall be provided by empathy. In order to be able to find out which emotional experience a consumer wants to have, his feelings and thoughts need to be recognized by the employee. From what has been said above, it follows that empathic capability should close the gap between producer and the emotional desires of the consumer. At the same time we have an explanation for the ever more highlighted role of empathy in business.

The second initial question asked to what extent Design Thinking contributes to this necessity of being an empathic employee. To sum it up, one can maintain that the claim for empathy within Design Thinking, on the one hand, creates knowledge about private, inner activities on the side of the user. This in turn can be used for the development of new products. In this sense, the emphasis on empathy serves the process of production. On the other hand, empathy was analyzed as a crucial part of the creation of sense within a project team, because the narrative of external empathy establishes new values and a new pride within the employee. It is a realization that he designs his ideas and products not only for "someone out there" but rather for real users with concrete needs. Furthermore, the internal empathy leads to the creation of a social and liable sphere within teamwork. We therefore conclude that empathy seems to be a means for social construction of the employee, because

> (...) on a social level, these constructions of knowledge influence how professionals construe their identities as either knowledge workers or designers. (Rylander 2009: 12)

In this view, empathy can be seen as a creator of value and sense. The human-centered rhetoric constructs identities—the designer sees himself as someone who works together closely with people and who satisfies their needs.

Because it's all about gaining knowledge about desires of people, we suggest describing empathy in the organizational context as a form of knowledge construction. In order to create this knowledge about other people's mind, one has to be empathic. The offered empathy techniques as provided in Design Thinking are a mixture of emotional and cognitive aspects. On the one hand, the Design Thinker shall see himself from a reflective distance in order to negate his own view in favor of the users' perspective. On the other hand he should maintain an open and non-analytical attitude. Therefore a conscious handling of these partly contradictory requirements and a clear picture of what empathy means to oneself is recommended before introducing empathy techniques to the workplace.

We see our contribution in the listing of empathic techniques for the construction of internal and external knowledge and in the demonstration of pitfalls and success-promising aspects. We hope that our findings may function as a starting point for (a) the comparison with traditional knowledge work and (b) the observation of the consequences for daily practice in companies. We also considered the "big picture" and suggested an explanation as to why empathic techniques have grown so important in the contemporary western economy.

For our further research, it will now be of interest to find out if empathy will indeed lead to innovation and positive change in companies that try out the Design Thinking approach. It will furthermore be of interest to observe how the "radical collaboration" between multidisciplinary team members and whether the implicit requirement of empathy will find its way into the organization.

Acknowledgement We thank Sharon Nemeth and Birga Schlottmann for their editorial support and Holger Rhinow for his inspiring comments.

References

Aronson E, Wilson TD, Akert RM (2004) Sozialpsychologie. 4. Aufl. Pearson Studium, München

Bischof-Köhler D (1989) Spiegelbild und Empathie. Die Anfänge der sozialen Kognition. Huber, Bern

Brown T (2008) Design Thinking. In: Harvard business review (6), S. 84–92

Cameron KS, Spreitzer GM (eds) (2012) The Oxford handbook of positive organizational scholarship. Oxford University Press, New York

Carlgren L, Elmquist M, Rauth I (2013) Demystifying design thinking: a conceptual framing of design thinking in use (Forthcoming)

d.school Stanford (2011) Bootcamp bootleg. Available on http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf. 28 Mar 2012

Davis MH (2007) Empathy. In: Stets JE und Turner JH (Hg) Handbook of the sociology of emotions. Handbooks of sociology and social research, Springer, New York

De Waal FBM (2011) Das Prinzip Empathie. Was wir von der Natur für eine bessere Gesellschaft lernen können; mit Zeichnungen des Autors. 1. Aufl. Hanser, München

Ekman P (2004) Gefühle lesen. Wie Sie Emotionen erkennen und richtig interpretieren. 1. Aufl. Elsevier, München

GE Reports (2011) Fusing science and empathy in design thinking at GE. Verfügbar unter http://www.gereports.com/fusing-science-and-empathy-in-design-thinking-at-ge/. zuletzt geprüft am 22 Nov 2011

Geulen D (1982) Perspektivenübernahme und soziales Handeln. Texte zur sozial-kognitiven Entwicklung. 1. Aufl. Suhrkamp, Frankfurt am Main (Suhrkamp-Taschenbuch Wissenschaft, 348)

Goleman D (2003) What makes a leader? In: Porter LW (Hg) Organizational influence processes. 2. Aufl. Sharpe, Armonk, NY

Greenson RR (1960) Empathy and its vicissitudes. Int J Psychoanal 41:418–424

Grotz A, Creuznacher I (2012) Design thinking—Prozess oder Kultur? In: Organisationsentwicklung. Zeitschrift für Unternehmensentwicklung und Change Management. S 14–21

Hochschild AR (2003) The managed heart. Commercialization of human feeling. 20. Aufl. University of California Press, Berkeley, CA. Available on http://www.loc.gov/catdir/description/ucal042/2003042606.html

Kolko J (2011) Exposing the magic of design. A practitioner's guide to the methods and theory of synthesis. Oxford University Press, New York

Köppen E, Rauth I, Schnjakin M, Meinel C (2011) The importance of empathy in it projects: a case study on the development of the German electronic identity card. In: Proceedings of international conference on engineering design, ICED11, 15–18 August 2011, Technical University of Denmark

Kouprie M, Visser S (2009) A framework for empathy in design: stepping into and out of the user's life. J Eng Des 20:437–448

Lamm CB, Batson CD, Decety J (2007) The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J Cogn Neurosci 19:42–58

Leonard DA, Rayport J (1997) Spark innovation through empathic design. Harv Bus Rev 75:102–113

Miyashiro (2011) Marie R: the empathy factor. Your competitive advantage for personal, team, and business success. Available on http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10497472

Neckel S, Wagner G (eds) (2013) Leistung und Erschöpfung. Burnout in der Wettbewerbsgesellschaft. Suhrkamp Verlag, Berlin

Pavlovich K, Krahnke K (2012) Empathy, connectedness and organisation. J Bus Ethics 105:131–137

Postma CE, Zwartkruis-Pelgrim E, Daemen E, Du J (2012) Challenges of doing empathic design: experiences from industry. Int J Des 6(1):59–70

Prinz J (2011) Against empathy. South J Philos 49:214–233

Prior L (2003) Using documents in social research. Sage, London

Rastetter D (2008) Zum Lächeln verpflichtet. Emotionsarbeit im Dienstleistungsbereich. Campus-Verl, Frankfurt am Main

Reckwitz A (2012) Die Erfindung der Kreativität. Zum Prozess gesellschaftlicher Ästhetisierung. Suhrkamp, Berlin

Rizzolatti G, Griese F, Sinigaglia C (2008) Empathie und Spiegelneurone. Die biologische Basis des Mitgefühls. 1. Aufl. Suhrkamp, Frankfurt am Main (edition unseld, 11)

Rogers CR (1975) Empathic: an unappreciated way of being. Psychologist 5:2–5

Rylander A (2009) Design thinking as knowledge work: epistemological foundations and practical implications. Des Manag J 4(1):7–19

Schönberger K, Springer S (2003) Subjektivierte Arbeit. Mensch, Organisation und Technik in einer entgrenzten Arbeitswelt. Campus-Verl, Frankfurt am Main

Sennett R (2012) Zusammenarbeit. Was unsere Gesellschaft zusammenhält. 1. Aufl. s.l. Carl Hanser Verlag, München

Stein E (1980) Zum Problem der Einfühlung. Kaffke, München

Voswinkel S (2002) Bewunderung ohne Würdigung? Paradoxien der Anerkennung doppelt subjektivierter Arbeit. In: Axel Honneth (Hg) Befreiung aus der Mündigkeit. Paradoxien des gegenwärtigen Kapitalismus. Frankfurter Beiträge zur Soziologie und Sozialphilosophie, vol 1. Campus-Verl, Frankfurt am Main

Voß et al (2013) Burnout und depression - Leiterkrankungen des subjektivierten Kapitalismus oder: Woran leidet der Arbeitskraftunternehmer? In: Neckel S, Wagner G (eds) Leistung und Erschöpfung. Burnout in der Wettbewerbsgesellschaft. Suhrkamp, Berlin

Developing Novel Methods to Assess Long-Term Sustainability of Creative Capacity Building and Applied Creativity

Manish Saggar, Grace Hawthorne, Eve-Marie Quintin, Eliza Kienitz, Nicholas T. Bott, Daniel Hong, Yin-Hsuan Chien, Ning Liu, Adam Royalty, and Allan L. Reiss

Abstract Creativity, the ability to create novel and useful outcomes, has been widely recognized as an essential skill for both entrepreneurial and every-day success. Given the vital import of creativity in our everyday lives, our research proposes to examine the impact and sustainability of creative capacity building using targeted training. In this chapter, we provide (a) a summary of behavioral results of creative capacity enhancement following 5-weeks of targeted training; (b) an unique experimental design to examine the long-term (after 1 year)

M. Saggar (✉) • E.M. Quintin • N. Liu • A.L. Reiss
Department of Psychiatry and Behavioral Sciences, Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA
e-mail: saggar@stanford.edu; quintin@stanford.edu; ningl@stanford.edu; areiss1@stanford.edu

G. Hawthorne • A. Royalty
Hasso Plattner Institute of Design (d.school), Building 550, 416 Escondido Mall, Stanford, CA 94305-3086, USA
e-mail: grace@dschool.stanford.edu; aroyalty@stanford.edu

E. Kienitz • N.T. Bott
Department of Psychiatry and Behavioral Sciences, Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA

Pacific Graduate School of Psychology-Stanford University Psy.D. Consortium, Stanford, CA, USA
e-mail: ekienitz@stanford.edu; nbott@stanford.edu

D. Hong
Institute of Biomedical Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan
e-mail: ddh0410@gmail.com

Y.H. Chien
Department of Pediatrics, Taipei City Hospital Zhong-Xing, Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan
e-mail: Dtpedr81@gmail.com

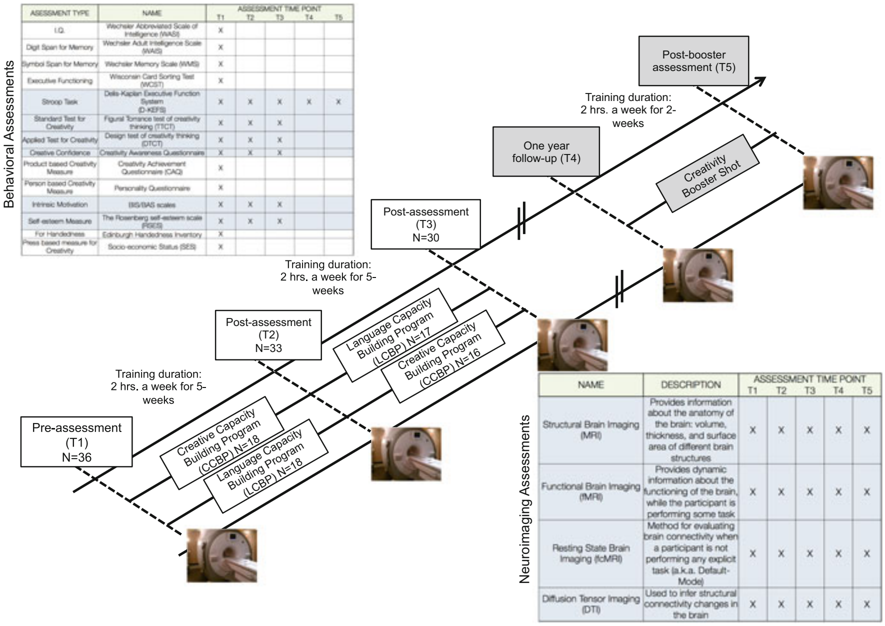
H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_3, © Springer International Publishing Switzerland 2015

sustainability of creative capacity building and the effect of a “booster-shot” of creativity training; and (c) preliminary insights and proposed work on the newly developed Design Test of Creativity Thinking (DTCT) to assess applied creativity. Altogether, we anticipate that our work will provide valuable insights into creative capacity building and assessment.

1 Introduction

Creativity has such a wide impact on our lives, ranging from entrepreneurial success (Amabile 1997; Kern 2010) to successful adaptation in addressing daily life-demands (Csikszentmihalyi 1996; Reiter-Palmon et al. 1998) and from promoting resilience (Metzl 2009) to psychological well-being (Cropley 1990; Runco 2004). Given this fundamental import and the fact that creativity is known to decline in early childhood, we argue that finding new ways to build and sustain creative capacity is required. Thus, not surprisingly, several methods have been devised to enhance creativity, namely provisioning effective incentives (Eisenberger and Shanock 2003), enhancing domain knowledge (Ericsson and Charness 1994), structuring group interactions (Kurtzberg and Amabile 2001), optimizing culture and environment (Ekvall and Ryhammar 1999), and targeted training (Scott et al. 2004). Among these approaches targeted training has been widely used across occupations and student populations. Although targeted training at the individual level has been effective in enhancing creative capacity, most of this work has been limited to academic settings in young children and adolescents (Scott et al. 2004). Thus, it is unclear whether targeted training is as successful in adults as in younger populations.

To address this gap and to examine the long-term sustainability of creative capacity, Stanford’s Hasso Plattner Institute of Design and the Center for Interdisciplinary Brain Research at Stanford conducted a randomized control trial where the efficacy of targeted design thinking skills were examined in enhancing creativity in adults (Hawthorne et al. 2013). In this study half of the participants ($n = 36$) initially received 5-weeks of Creative Capacity Building Program (CCBP) and the other half received a parallel control training for the same duration (Language Capacity Building Program (LCBP)). Both interventions lasted 5 weeks with weekly meetings of 2 h per week. We pseudo-randomly assigned participants to either intervention and hence formed two groups. We matched these groups on age, gender, and IQ. Following the initial intervention and data collection, the groups crossed-over to receive the second intervention, i.e. participants in the CCBP were assigned to the LCBP and vice versa. The second set of interventions was followed by third set of assessments and neuroimaging data collection (Hawthorne et al. 2013). Thus, overall we collected data at three time-points (T1–T3; Fig. 1).


Behavioral Assessments
Post-booster assessment (T5)
Training duration: 2 hrs. a week for 2-weeks
One year follow-up (T4)
Creativity Booster Shot
Post-assessment (T3) N=30
Training duration: 2 hrs. a week for 5-weeks
Post-assessment (T2) N=33
Language Capacity Building Program (LCBP) N=17
Creative Capacity Building Program (CCBP) N=16
Training duration: 2 hrs. a week for 5-weeks
Pre-assessment (T1) N=36
Creative Capacity Building Program (CCBP) N=18
Language Capacity Building Program (LCBP) N=18
Neuroimaging Assessments

ASESSMENT TYPE	NAME	ASSESSMENT TIME POINT T1	T2	T3	T4	T5
I.Q.	Wechsler Abbreviated Scale of Intelligence (WASI)	X				
Digit Span for Memory	Wechsler Adult Intelligence Scale (WAIS)	X				
Symbol Span for Memory	Wechsler Memory Scale (WMS)	X				
Executive Functioning	Wisconsin Card Sorting Test (WCST)	X				
Stroop Task	Delis-Kaplan Executive Function System (D-KEFS)	X	X	X	X	X
Standard Test for Creativity	Figural Torrance test of creativity thinking (TTCT)	X	X	X		
Applied Test for Creativity	Design test of creativity thinking (DTCT)	X	X	X		
Creative Confidence	Creativity Awareness Questionnaire	X	X	X		
Product based Creativity Measure	Creativity Achievement Questionnaire (CAQ)	X				
Person based Creativity Measure	Personality Questionnaire	X				
Intrinsic Motivation	BIS/BAS scales	X	X	X		
Self-esteem Measure	The Rosenberg self-esteem scale (RSES)	X	X	X		
For Handedness	Edinburgh Handedness Inventory	X				
Press based measure for Creativity	Socio-economic Status (SES)	X				

NAME	DESCRIPTION	ASSESSMENT TIME POINT T1	T2	T3	T4	T5
Structural Brain Imaging (MRI)	Provides information about the anatomy of the brain: volume, thickness, and surface area of different brain structures	X	X	X	X	X
Functional Brain Imaging (fMRI)	Provides dynamic information about the functioning of the brain, while the participant is performing some task	X	X	X	X	X
Resting State Brain Imaging (fcMRI)	Method for evaluating brain connectivity when a participant is not performing any explicit task (a.k.a. Default-Mode)	X	X	X	X	X
Diffusion Tensor Imaging (DTI)	Used to infer structural connectivity changes in the brain	X	X	X	X	X

Fig. 1 Overall study design and assessment schedule. The *gray-colored boxes* show long-term sustainability assessment (1-year follow-up) and the booster shot of creativity training

The CCBP was an abbreviated version of a highly popular class offered at the Stanford Hasso Plattner Institute of Design called ME266 Creative Gym (http://dschool.stanford.edu/classes/#creative-gym-a-design-thinking-skills-studio). We designed CCBP as an interactive studio where students can build their creative confidence and sharpen their individual design thinking skills through hands-on experiences, rapid prototyping, and other improvisational exercises. The LCBP intervention, on the other hand, consisted of many hands-on exercises to learn basic Chinese vocabulary, character writing, and commonly used phrases (e.g. How are you?), taught by a bilingual native Mandarin speaker. The LCBP intervention was intended to present a similar experience as in the CCBP class while reducing the opportunity to produce creative outcomes or reflect upon a creative process.

We have analyzed data from the first two time points (T1 and T2), i.e., before and after initial CCBP and LCBP training, and show that participation in CCBP indeed enhanced creativity in adults (Kienitz et al. 2014) and was associated with enhanced information processing ability (Bott et al. 2014). In this chapter, we provide a summary of these results.

Although, the efficacy of CCBP in enhancing creative capacity was shown with just 5-weeks of training, it is unclear whether such enhancement can sustain over longer period of time (e.g., 1 year) or if regular interventions are required to sustain creative capacity enhancement. To examine the long-term sustainability of creative capacity enhancement, in the second phase of our study, we invited our participants for a fourth behavioral and neuroimaging assessment 1 year after the last training session (T4, see Fig. 1). Further, a short "booster" dose of creative training (4 h) was provided to see if such abbreviated training could help sustain or further improve the creative capacity in adults. This booster session was followed by a fifth round of behavioral and neuroimaging assessments (T5, see Fig. 1). This chapter provides our unique experimental design and associated hypothesis to test the long-term sustainability of creative capacity enhancement.

Lastly, we argue that just as efforts are needed to foster creativity across lifespan, research is required to develop and test novel assessments for measuring creative capacity. Since J.P. Guilford's seminal lecture on creativity in 1950 (Guilford 1950), several tools have been developed to assess creative capacity across age groups and populations (Dietrich and Kanso 2010). Several psychometric instruments have been developed and standardized across age groups. For example, the Alternate Uses Test (AUT) has been widely used in literature since its inception (Guilford 1967). Like most of these instruments, the AUT is based on the foundation of divergent thinking, i.e., ability to produce a large number of unusual and novel responses to given problem. Although proposed as a starting point to study creativity, divergent thinking has often taken a primary role in the creativity scientific literature. However, some drawbacks might exist with using divergent thinking based assessments as the primary method for assessing creativity. In particular, research has shown that creative outcomes can also result from critical and convergent thinking (Nickerson 1999). Further, it is unclear how whether laboratory-based divergent thinking tests are truly relevant to applied creativity in real-world settings (Dietrich 2007). Considering these points, we developed a novel Design Thinking Creativity Test (DTCT), which measures individual creative capacity in solving real-world problems. In this chapter, we provide preliminary insights into assessing applied creativity using the DTCT.

Altogether, we provide a summary of results showing that creative capacity can be enhanced in adults using targeted training. We also present our experimental design for testing long-term sustainability of creative capacity enhancement. Lastly, we share insights regarding our novel DTCT for measuring individual creative capacity in a real-world setting.

2 Behavioral Correlates of Enhanced Creative Capacity

In this section, we briefly provide a summary of results from two behavioral outcomes collected before and after CCBP/LCBP training, i.e. from the first two time points (T1 and T2). First, we assessed the behavioral correlates of creativity using the standardized Torrance Test of Creativity Thinking Figural version (TTCT-F) (Torrance 1990). Although primarily based on divergent thinking, the TTCT has been widely described in the literature to assess creativity. In the TTCT-F, participants are given a set of incomplete figures and are asked to complete them so that each figure can tell a unique and complete story. The figures are later assessed by trained judges, who rate each figure primarily on the basis of fluency (number of items in the figure), originality (uniqueness of drawing), elaboration, abstraction of title, and resistance to premature closure (Torrance 1990). Using this widely used standardized measure of creativity, we found that participation in CCBP, as compared to LCBP, led to increased TTCT-F scores with moderate to large effect sizes (Kienitz et al. 2014). Specifically, we found that CCBP participants scored higher on the sub-scales of elaboration and resistance to premature closure after training as compared to LCBP participants. Increased elaboration scores, after CCBP training, suggests that participants generated more imaginative and detailed responses, while higher scores on resistance to premature closure suggests that CCBP participants considered more options in response to the presented stimuli and created wider associations after training (Kienitz et al. 2014). Altogether, these data provide evidence for the efficacy of targeted training to enhance creativity in healthy adults.

In addition to creativity, we also assessed changes in executive functioning with CCBP training as compared to LCBP training. Broadly defined, low-level executive functions of the brain include attention and information processing speed, while high-level executive functions include response inhibition, cognitive flexibility (ease of shifting between concepts/contexts), working memory, etc. (Bull and Scerif 2001). Investigating changes in executive functions associated with creativity training provided us with an opportunity to better understand how low- and high-level brain functioning supports and perhaps facilitates creative capacity enhancement. We administered three subtests of the Delis-Kaplan Executive Function System (D-KEFS) to measure executive functioning pre and post-training (Delis et al. 2001). The D-KEFS subtests used in this study were the Color-Word Interference, Verbal Fluency, and Design Fluency. Among these subtests, the color-word interference test (CWIT) was used as a primary outcome measure. The CWIT is based on the Stroop procedure (Stroop 1935) and has four conditions. The first two conditions assess "lower-level" goal-directed attention and processing speed, and the last two conditions assess "higher-level" inhibition and cognitive flexibility. As presented elsewhere (Bott et al. 2014), we showed that participation in CCBP led to increased low-level executive functioning, such that CCBP participants completed the task in less time (as compared to before training). No change was found in high-level executive functioning. These results are in line with the style of CCBP training, where participants were motivated to increase their bias towards action using fast-paced prototyping exercises (Hawthorne et al. 2013).

3 Experimental Design to Assess Long-Term Sustainability of Creative Capacity Enhancement

To measure the long-term sustainability of increased creative capacity, we invited our original set of participants to return 1-year after their third assessment (Fig. 1). At that time, we examined behavioral and neuroimaging measures of creativity and executive functioning for the fourth time (T4). Such examinations can provide crucial information regarding the estimated duration of targeted creativity training effects, and thus assist in the development of novel methods to efficiently enhance long-term creative capacity in the context of design thinking based curriculum. Figure 1 shows the proposed experimental design, where time point T4 depicts the 1-year follow-up.

In addition to long-term sustainability, we were also interested in examining whether enhanced creative capacity and the associated changes in brain and behavior require ongoing conditioning just as regular sit-ups are required in case of physical exercise to retain good health and physical fitness. To answer this question we designed a short creativity training session, henceforth referred to as "booster" session. After, 1-year follow-up assessment (at T4) participants took part in an abbreviated training protocol (two 2-h sessions based on the d.school's ME266 Creative Gym course methodology) and were assessed one more time at T5 (Fig. 1). By analyzing changes in behavioral and neuroimaging measures of creativity between T4 and T5, we can estimate the effects of a quick booster training.

Altogether, as a first-of-its-kind study, we collected a rich dataset at five time points. By analyzing these data in the future, we will be able to answer questions regarding the long-term sustainability of creativity, and whether ongoing conditioning is required for better retention of creative capacity.

4 Need for a Twenty-First Century Creative Capacity Assessment

A series of psychometric instruments are available today to assess individual creative capacity (Dietrich and Kanso 2010; Arden et al. 2010). However, most of these instruments only consider divergent thought processes. As noted previously, it has been previously shown that convergent and critical thinking can also result in creative outcomes (Nickerson 1999). Further, it is unclear whether laboratory-based instruments capture and relate to real-world issues and settings. While popular assessment tools like the TTCT have exercises to assess an individual's possession of mental creativity characteristics, they do not assess an individual's ability to overtly apply/exercise their creativity in a real world setting. Thus, we argue that a widely applicable creativity assessment tool, which also takes into account new information and concepts from recent behavioral and neuroscience research, is required to efficiently assess individual creative capacity.

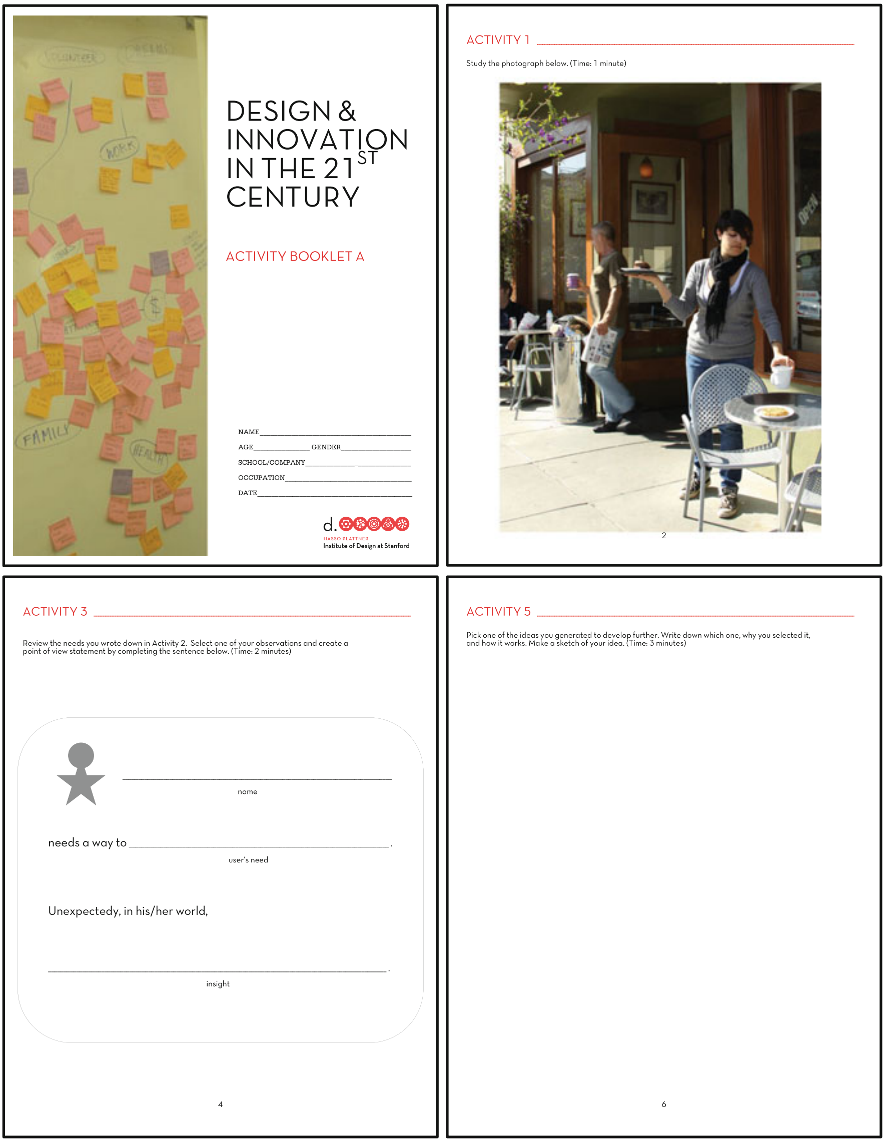
Design thinking methodology, when applied effectively, can help address prominent problems in industry, academia, politics and even interpersonal relationships (Kelley and Kelley 2012). In an attempt to develop a new instrument that can assess such real-word abilities, we prototyped the Design Thinking Creativity Test (DTCT) as a next generation creativity assessment that reflects problem solving needs of the twenty-first century. The DTCT emphasizes on assessment of case-based skills to directly measure the application of creative characteristics during an innovation event. As opposed to separately assessing convergent and divergent styles of thinking, the DTCT is based on the principles of design thinking that incorporate elements of both. In particular, the DTCT emphasizes assessment of case-based skills to directly measure an individual's application of creativity.

The DTCT prototype was structured to assess a subject's ability to apply their creative skills in a case study scenario within timed constraints, limited materials and changing conditions. These features help us measure a subject's flexibility, nimbleness and imagination aspects of creativity as applied in a convergent and divergent manner. For example, starting with a hypothetical scenario represented by an image (Fig. 2), each participant observed the scenario and identified the needs of the person depicted in the image. Next, each participant converged on one need and defined it. Each participant then created several possible solutions and physically prototyped the solutions (Fig. 3). Physically prototyping solutions helped participants translate the need into three-dimensional formats. To test participant's flexibility and adaptability, a pivot was inserted into the DTCT by changing specifics of the provided scenario. Based on the condition change, each participant altered the needs and made changes to their possible solutions by building upon original prototypes. The rapid flow of tasks throughout the activity/assessment is modeled after the design thinking methodology that Stanford's Hasso Plattner Institute of Design teaches and with the creativity skills building course Creative Gym ME366.

Our DTCT prototype was administered during the first three assessments (T1–T3) in the longitudinal study described above (Hawthorne et al. 2013). As also mentioned previously, in addition to the DTCT, we collected data from a battery of behavioral, cognitive, and neuroimaging evaluations. Thus, this large amount of multidimensional data will allow us to further refine and improve the DTCT by interrogating associations of different aspects of DTCT with these different aspects of brain functioning.

Although a promising start, in order to create a robust and standardized twenty-first century creativity assessment, more work is required. Specifically, the following questions need to be addressed:

- How reliable is a case-base format in capturing applied creativity in a simulated real-world setting?
- Which cognitive and behavioral constructs are applied while engaged in the DTCT assessment?



VOLUNTEER
WORK
FAMILY
HEALTH
OPEN

DESIGN & INNOVATION IN THE 21ST CENTURY

ACTIVITY BOOKLET A

NAME____________________
AGE__________ GENDER__________
SCHOOL/COMPANY____________________
OCCUPATION____________________
DATE____________________

d.
HASSO PLATTNER
Institute of Design at Stanford

ACTIVITY 1

Study the photograph below. (Time: 1 minute)

2

ACTIVITY 3

Review the needs you wrote down in Activity 2. Select one of your observations and create a point of view statement by completing the sentence below. (Time: 2 minutes)

____
name

needs a way to ____________________.
user's need

Unexpectedy, in his/her world,

____________________.
insight

4

ACTIVITY 5

Pick one of the ideas you generated to develop further. Write down which one, why you selected it, and how it works. Make a sketch of your idea. (Time: 3 minutes)

6

Fig. 2 A sample set of activities included in the DTCT

- Being based on the contemporary principles of design thinking, does DTCT better capture creative capacity enhancement as opposed to the standard TTCT assessment?

The creation of a statistically robust, well-standardized test relies on the acquisition of large amounts of data from large-scale studies in which representative

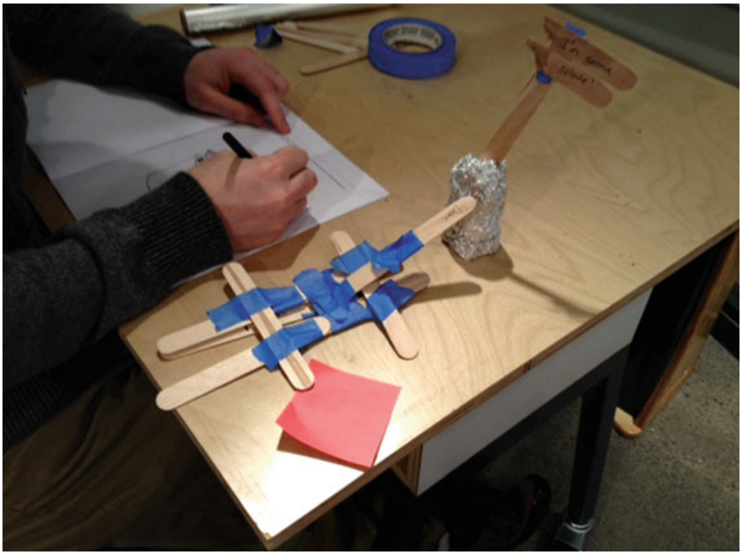


Fig. 3 Example of the DTCT task prototype from a representative participant

groups of individuals take the test under standardized conditions (DeVon et al. 2007). In this work, we would like to align the DTCT with creativity training goals from design thinking methodology as our initial attempt to establish construct validity. By doing so, we hope to make connections between creativity training, its effect on an individual's creative capacity, and the potential impact of increased capacity from training. By attaining this goal, we (and others) will be able to accurately evaluate the effects of training/teaching methodologies by utilizing a direct, twenty-first century assessment instrument.

Altogether, by creating a psychometrically robust creativity assessment measure as a companion to the TTCT, we will be able to create a measuring tool that can map training to development and practice to impact. These findings will help guide instruction content and training exercises for creativity training in a large number of settings as well as our associated institutions (Stanford and HPI). Teaching design thinking goes beyond classroom methodology to applied execution in real world scenarios for impactful change. The DTCT has the potential to become a new industry assessment norm for individuals, educators and executives across all disciplines and industries as the creativity assessment that matches problem solving in the twenty-first century.

5 Implications/Future Work

In this chapter, we provide a summary of results that show the efficacy of 5-weeks of targeted creativity training in enhancing creative capacity in adults using the principles of design thinking. We also provide a glimpse of our unique experimental design to examine the long-term (i.e., after 1 year) sustainability of creative capacity building and the effect of a "booster-shot" of creativity training. Finally, we provide preliminary insights and proposed work on the newly developed Design Test of Creativity Thinking (DTCT) to assess applied creativity. We believe that by examining whether individual creative capacity can be enhanced and the long-term sustainability of such enhancement, we can provide invaluable metrics and methods for improving creative and instructional effectiveness across disciplines and occupations.

References

Amabile TM (1997) Entrepreneurial creativity through motivational synergy. J Creat Behav 31:18–26. doi:10.1002/j.2162-6057.1997.tb00778.x

Arden R, Chavez RS, Grazioplene R, Jung RE (2010) Neuroimaging creativity: a psychometric view. Behav Brain Res 214:143–156. doi:10.1016/j.bbr.2010.05.015

Bott N, Quintin E-M, Saggar M, Kienitz E, Royalty A, Hong DW-C et al (2014) Creativity training enhances goal-directed attention and information processing. Thinking Skills Creativity. doi:10.1016/j.tsc.2014.03.005

Bull R, Scerif G (2001) Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. Dev Neuropsychol 19:273–293. doi:10.1207/S15326942DN1903_3

Cropley AJ (1990) Creativity and mental health in everyday life. Creat Res J 3:167–178. doi:10.1080/10400419009534351

Csikszentmihalyi M (1996) Creativity: flow and the psychology of discovery and invention. HarperCollins Publications, New York

Delis DC, Kaplan E, Kramer JH (2001) Delis-Kaplan executive function system (D-KEFS). Psychological Corporation, San Antonio, TX

DeVon HA, Block ME, Moyle-Wright P et al (2007) A psychometric toolbox for testing validity and reliability. J Nurs Sch 39:155–164. doi:10.1111/j.1547-5069.2007.00161.x

Dietrich A (2007) Who's afraid of a cognitive neuroscience of creativity? Methods 42:22–27. doi:10.1016/j.ymeth.2006.12.009

Dietrich A, Kanso R (2010) A review of EEG, ERP, and neuroimaging studies of creativity and insight. Psychol Bull 136:822–848. doi:10.1037/a0019749

Eisenberger R, Shanock L (2003) Rewards, intrinsic motivation, and creativity: a case study of conceptual and methodological isolation. Creat Res J 15:121–130. doi:10.1080/10400419.2003.9651404

Ekvall G, Ryhammar L (1999) The creative climate: its determinants and effects at a Swedish University. Creat Res J 12:303–310. doi:10.1207/s15326934crj1204_8

Ericsson KA, Charness N (1994) Expert performance. Am Psychol 49:725–747

Guilford JP (1950) Creativity. Am Psychol 5:444–454

Guilford JP (1967) The nature of human intelligence. McGraw-Hill, New York

Hawthorne G, Quintin E-M, Saggar M et al (2013) Impact and sustainability of creative capacity building: the cognitive, behavioral, and neural correlates of increasing creative capacity. In: Leifer L, Plattner H, Meinel C (eds) Design thinking research: understanding innovation. Springer, Heidelberg, pp 65–77. doi:10.1007/978-3-319-01303-9_5

Kelley T, Kelley D (2012) Reclaim your creative confidence. Harv Bus Rev 90:115–8, 135

Kern F (2010) What chief executives really want. Businessweek, Bloomberg

Kienitz E, Quintin E-M, Saggar M, Bott NT, Royalty A, Hong DW-C et al (2014) Targeted intervention to increase creative capacity and performance: a randomized controlled pilot study. Thinking Skills Creativity 13:57–66

Kurtzberg TR, Amabile TM (2001) From Guilford to creative synergy: opening the black box of team-level creativity. Creat Res J 13:285–294. doi:10.1207/S15326934CRJ1334_06

Metzl ES (2009) The role of creative thinking in resilience after hurricane Katrina. Psychol Aesthet Creat Arts 3:112–123

Nickerson RS (1999) Enhancing creativity. In: Sternberg RJ (ed) Handbook of creativity. Cambridge University Press, New York, pp 392–430

Reiter-Palmon R, Mumford MD, Threlfall KV (1998) Solving everyday problems creatively: the role of problem construction and personality type. Creat Res J 11:187–197. doi:10.1207/s15326934crj1103_1

Runco MA (2004) Creativity. Annu Rev Psychol 55:657–687. doi:10.1146/annurev.psych.55.090902.141502

Scott G, Leritz LE, Mumford MD (2004) The effectiveness of creativity training: a quantitative review. Creat Res J 16:361–388. doi:10.1080/10400410409534549

Stroop JR (1935) Studies of interference in serial verbal reactions. J Exp Psychol 18:643–662

Torrance EP (1990) Torrance tests of creative thinking. Figural forms A and B. Scholastic Testing Service, Benserille, IL

The Personal Trait Myth: A Comparative Analysis of the Innovation Impact of Design Thinking Tools and Personal Traits

Nikolas Martelaro, Shameek Ganguly, Martin Steinert, and Malte Jung

Abstract Design thinking asserts that individuals and teams have the ability to build their innovative capacity through various tools and methods no matter their predispositions to creativity and innovation. The contexts of design thinking attempt to alter design process towards more innovative ideas. This work attempts to experimentally disentangle the impact of disposition and situation during design activity. We present a variety of design contexts intended to be tested against dispositional factors during an experimental design task. We then present a pilot study exploring how process-priming impacts design process during a problem-solving task and an open-ended design task. Our preliminary results suggest that short process-priming activities may not be the most effective means for altering design process. Rather, more integrated contextual interventions may be better candidates for impacting design process and would be interesting test variables for future studies.

N. Martelaro (✉)
Stanford University, 424 Panama Mall, Building 560, Stanford, CA 94305, USA
e-mail: nikmart@stanford.edu

S. Ganguly
Apple Inc., Cupertino, CA 95014, USA
e-mail: shameekg@alumni.stanford.edu

M. Steinert
Department of Engineering Design and Materials (IPM), Norwegian University of Science and Technology (NTNU), Richard Birkelandsvei 2B, NO - 7491, Trondheim, Norway
e-mail: martin.steinert@ntnu.no

M. Jung
Information Science, Cornell University, 206 Gates Hall, Ithaca, NY 14850, USA
e-mail: mfj28@cornell.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_4, © Springer International Publishing Switzerland 2015

1 Introduction

A core premise of the "design thinking" approach is that innovative capacity is less determined dispositionally than it is situationally. In other words, the approach claims that innovative capacity is not to be found in a person's traits (such as a person's preference for divergent or convergent thinking) or demographic attribute, but rather in the characteristics of the situation a person engages in (such as specific design thinking practices) (Brown 2008).

Contrary to this premise is the widely held belief that innovativeness or creativity is a matter of personal disposition rather than engagement in innovative practice. For example people think of themselves and others as creative or non-creative, and they believe that this is the defining criterion in determining the likelihood of producing innovative outcomes (Plucker et al. 2004; Treffinger et al. 1994). Accordingly, a major portion of past research has sought determinants of innovativeness in people's disposition, and much work has focused on finding dispositional characteristics that are predictive of innovative performance. The position that personal characteristics are more important than situational characteristics is also reflected in industry practice as companies choose strategies to select "innovative" employees rather than restructuring their company practices in order to increase their overall innovative capacity.

To disentangle these conflicting views, we have begun developing and testing comparative experiments investigating the effects of dispositional and situational characteristics on innovative performance. Given the lack of research on this kind of comparative approach, we have begun our investigations focusing on individuals rather than teams.

In this chapter, we describe some of the methods we have been developing and discuss preliminary results. We also discuss future plans and experimental designs to further this research.

2 Background

Prior research has shown successfully that disposition can be a critical determinant of innovative performance. For example in a survey study with 172 R&D employees (Scott and Bruce 1994) found that a systematic rather than intuitive problem solving styles was correlated with decreased innovative behavior. Additionally, Goldsmith (1986), using the Kirton Adaption-Innovation Index (Kirton 1976), found that problem-solving styles are highly correlated with personality. The idea of disposition as an important determinant of innovative performance has also been extended to teams. For example Kress and Schar (2012) explored how the variation of cognitive styles in a team predicts its performance.

In addition to studying dispositional determinants of innovative performance, researchers have also found many situational determinants of innovative performance. For example Dow et al. showed that engaging in prototyping can increase design performance (Dow et al. 2010, 2011). Highlighting the importance of the environment, studying the situational influence of color, Mehta and Zhu (2009) have found that exposure to the color blue, as opposed to red increases creativity in a product design task.

One of the major contributions of social psychology has been the insight that behavior can be predicted far more reliably by the characteristics of the situation a person is engaged in than by the characteristics of his or her personality (disposition) (Ross and Nisbett 1991). The power of situational influence has been demonstrated most prominently in Milgram's (1974) studies of obedience to authority, Asch's (1956) studies of group conformity, and Latane and Darley's (1968) studies on bystander inhibition. Comparing situational and dispositional determinants directly, Darley and Batson (1973) showed that people's personality characteristics could not predict whether they would help a person in need. However the situational characteristics (being in a hurry vs. not) were highly predictive of helping behavior. Despite these findings, people systematically underestimate the influence of situational factors in favor of dispositional ones, exhibiting a tendency that has been termed the Fundamental Attribution Error (Lee Ross 1977). This tendency partially explains the predominant focus of current research on dispositional determinants of innovative performance. In line with the social science literature on situational influence we, however, assume that engagement in design thinking practices will outweigh the influence of personality characteristics on innovative performance by far. To our understanding, the influence of dispositional and situational characteristics on innovative performance has not been studied comparatively in design.

Given the split among dispositional and situational factors as predictive determinants of innovative performance, we are interested in seeing how each are related to each other. Our intent is to study and understand how dispositional and situational factors interact with each other to influence design outcomes. We propose the following research questions:

RQ1: Do dispositional factors predict design outcomes during a design task?
RQ2: Do situational factors predict design outcomes during a design task?
RQ3: Do either dispositional or situational factors have more impact on design outcomes?
RQ4: Do dispositional and situational factors interact and influence design outcomes more so than each factor alone?

To begin answering these questions we have begun identifying various dispositional and situational factors that can be controlled for during quasi-controlled experiments. In addition, we have developed a prototype of a study to explore the influence and interactions of both factors.

2.1 Design Thinking Contexts

There are a variety of situational factors that are unique to design that may be interesting avenues for exploration. In this section, we outline various situational contexts which have important characteristics associated to design thinking.

2.1.1 Environment

One of the most striking and context dependent factors associated with design thinking is the environment that we as designers work. The workspace of designers is often shown covered in multi-color Post-It notes, random artifacts from past projects, and whiteboards. The environment evokes a sense of chaos, excitement, and playfulness. But more so, theses environments invite “mindful modification.” In their book on designing the d.school “Make Space,” Scott Doorley and Scott Witthoft (2012) describe their process as one of continuous prototyping, modification, and iteration. These design thinking spaces in turn embody these ideals by both supplying tools and materials to help designers during their process and by the rooms themselves being open to change and modification based on the designers’ needs. By allowing users of a space to modify and alter the space gives one the power to redesign their own environment. This quality may allow users to open up their thinking, giving them a sense that anything can be changed (Fig. 1).

2.1.2 Materials

In addition to the rooms themselves, the materials often associated with design thinking evoke a sense of modification. Post-its and whiteboards allow for ideas and concepts to be quickly created, modified, and thrown away without significant feelings of loss. The low fidelity prototyping supplies such as pipe cleaners and aluminum foil allow for quick physical creation of physical ideas. Aside from allowing quick action and realization of an idea in physical space, the materials invite themselves to be modified and altered quickly. Unlike pre-production prototypes made of plastic or metal, these low fidelity prototypes are more easily modified allowing the designers to create, reflect, and modify in real time (Fig. 2).

2.1.3 Design Teams

One of the most important factors influencing design thinking is the design team. Design teams are often put together to provide a number of skills and different viewpoints. This variety in teams allows for teams members to share ideas and challenge each other’s ways of thinking. In turn, this may cause design team members to question their assumptions and their own ways of thinking. This

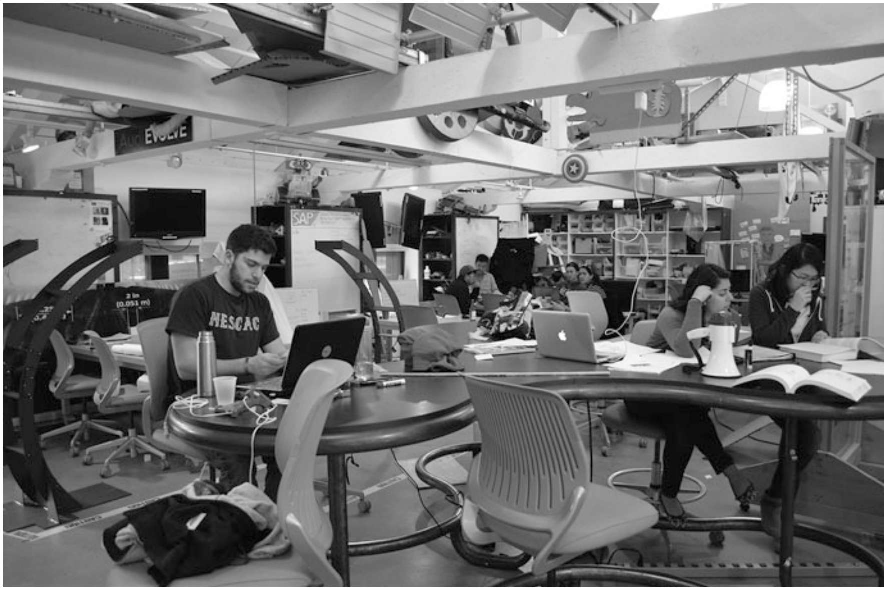


Fig. 1 Mutable design work environment

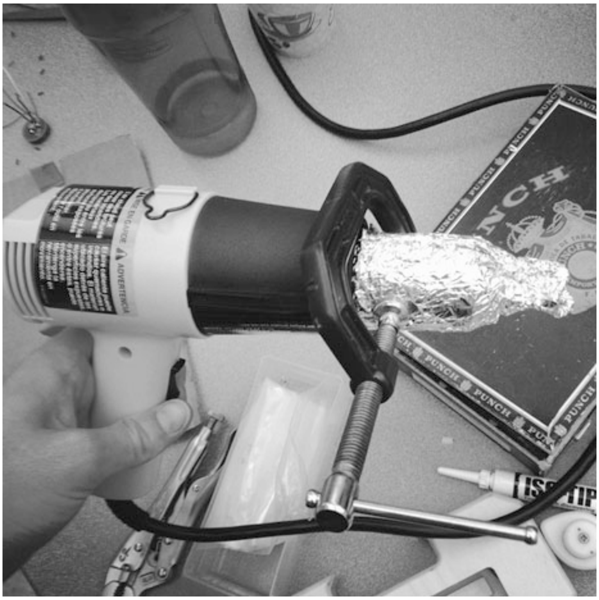


Fig. 2 Low fidelity prototype of an improved heat gun design

questioning spurs design team members to change and alter their ways of thinking throughout a project and overtime even after a project. Thus, the design team may also be another means for allowing mutability in the cognitive process of a designer.

2.1.4 Design Questions

The design team provides the stimulation to the individual that allows them alter their process or thinking, on both a short and long time scale. One of the primary ways that individual designers and design teams can challenge cognitive processes is through the questions they ask. Specifically, two types of questions have been found to correlate with improved design team performance, Deep Reasoning Questions and Generative Design Questions (Eris 2004). Deep reasoning questions seek to converge on a fact based answer, for example, "how much weight can this bean support?" The questioner expects the answer to be true. Alternatively, Generative Design Questions aim to create many possible answerers without the need for fact or truth. For example, one may ask, "How might we support this weight?" This question allows for the generation of many alternatives, such as a beam, a rope and pulley system, a hovercraft, a magnetic levitation system, ect. This question allows the designer or team to alter and change their thinking process. With each new question comes a new opportunity to alter and change the path of the design being worked on.

2.1.5 Contexts of Mutability

Looking at each of these different contexts, a common element to all of them is their affordance of mutability. Design thinking tools may work by allowing individuals and teams to challenge and alter their process and cognitive style. It may well be that this mutable process is what we consider to be the disposition or personality trait of a designer. Seeing a designer in one context and working on one project may give the illusion of a fixed design personality, however, we hypothesize that (expert) designers in different contexts will alter their own personal processes to best suit the project needs. In addition, we hypothesize that certain contexts that challenge and allow the designer to change will provide more opportunity for the designer to alter their process.

3 Development of an Experimental Approach to Examine Dispositional and Situational Determinants of Design Performance

The aim of our approach is to conduct a series of experiments that compare typical design thinking characteristics from a dispositional and situational perspective. In other words, we want to select participants that either have or don't have a specific design thinking trait (such as exploratory thinking styles) and then expose them to a situation with or without design thinking characteristics.

3.1 Selecting Problem and Solution Focus as a Key Dispositional Factor

During the course of our development, we have considered various candidates for design thinking characteristics that could act as demarcations for dividing study participants into "design thinking" and "non-design thinking." Candidate design thinking characteristics that could be explored included: divergent vs. convergent thinking, innovative vs. adaptive problem solving styles, problem vs. solution oriented problem solving, or mindful vs. mindless design approaches. Of these areas, problem vs. solution focused problem solving strategy is a rich area to begin our inquiry. Problem-solution focus affords itself well to our study as it can be both an aspect of one's personality and is manifested in one's behavior. Problem-solution focus has also been well studied in the past and has been identified as a central construct distinguishing design approaches from scientific/engineering approaches to problem solving. Lastly, Lawson (1979) has developed an experimental task that can characterize and distinguish problem vs. solution-focused behavior in the lab.

Lawson originally made this distinction between cognitive style when running a controlled study between final year architecture and design students. Lawson found that while performing a constraint-based problem solving activity, the architecture students focused on finding a good solution, while the science students focused on understanding the constraints and optimizing a solution. While this finding suggests there are dispositional aspects to one's cognitive style, a follow up study by Lawson found that there were no differences in cognitive styles in first year students. These students had not been through many years of formal training in their respective fields and thus Lawson proposed that ones education has significant impacts on a designer's cognitive style.

This finding gives some indication that what we may think of as design thinking traits are not some inborn parts of our personality, but rather are learned and absorbed over time through the context of one's environment and education.

These findings through are biased in that they examined students rather than design professionals. In another study examining cognitive styles and design outcomes, professional industrial designers were shown to have different cognitive styles (Kruger and Cross 2006). In a protocol study, nine professional designers were asked to design solutions for a train litter disposal system. The designers then completed the design activity using a speak aloud protocol which was later analyzed and used to categorize with a certain cognitive style. To analyze the designer's problem solving style, Kruger and Cross developed an expertise model with the following activities:

1. *Gather* data
2. *Asses* value and validity of data

3. *Identify* constraints and requirements
4. *Model* behavior and environment
5. *Define* problems and possibilities
6. *Generate* partial solutions
7. *Evaluate* solutions
8. *Assemble* a coherent solution

Through counting the number of statements each designer made in each category of the expertise model four cognitive style were identified: (1) Problem driven, (2) Solution driven, (3) Information driven, and (4) Knowledge driven. Information and knowledge driven designers were categorized as subsets of problem and solution drive design, respectively. Of the nine designers, four were either problem or information driven and five were solution or knowledge driven. Although a small sample size, this split in designer cognitive styles suggests that Lawson's notion of designers as solution focused may not be as concrete at the professional level. However, we know very little about the professional designers chosen. While we do know that they were all industrial designers it would not be so far off to assume that their momentary cognitive styles may be influenced by where they were trained, where they work, or even the current project they were working on. Thus, to gain a better understanding of designers during future experiments it may be useful to have a more complete contextual profile of their current projects and positions.

While there was a divide in the problem solving style used between designers, there was a common trend that seven of the nine designers spent most of their time *generating solutions*. This would suggest that solution generation might be a valid design thinking trait to test for. In addition, solution oriented designers were defined to have had a higher ratio of generation vs. gather and identify activities. During a design task, this too could be a method for assessing problem vs. solution orientation on task.

In addition to characterizing the designer by cognition style, Kruger and Cross also had each designer's work rated across a variety of categories including aesthetics, ergonomics, creativity, technical aspects, business aspects, and an overall judgment.[1] The scores for each designer's design (scored by a team of professional designers) were found to have significant individual variation. The only partial trends that seemed to exist were that problem focused designers had slightly higher overall scores and solution oriented designers were rated as slightly more creative. These results however were only suggestive and overall both cognitive styles yielded good results, depending on the designer.

The results presented by Kruger and Cross may seem to suggest that different designers may have an innate ability, and that the process that one takes is not as relevant to the actual design outcome. However, a designer's process is often a personally *crafted* entity. To take away a designer's designers process may be equivalent to taking away their personality. In our own personal experiences as

[1] This overall judgment was not a cumulative or mean score of the other categories, but rather a separate, holistic measure.

designers, we have found that each project causes us to alter and adapt our process given the systems we are working within. Additionally, through reflection of our process throughout and after a design project (sometime even years after) we learn and develop new methods that we can use and incorporate into our process. Thus, the notion of the designer's process as their "design personality" may account for what many call traits (Schön 1983). However, if the personality of a designer were directly linked to a process, then it would be possible for others to become more designerly through crafting and reflecting upon their own process. This is not to say that processes handed down and forced upon people would lead to better designs, but that people must craft their own design processes. Ultimately, what may be the benefit of design thinking is that the methods and tools used are often so different from the rational, linear cognitive styles that are taught that they allow people the freedom to challenge and alter their own processes. It may be through this new found power to alter and adapt one's process depending on the context of a situation that allows one to be more designerly. Situational context that allows for flexibility and reflection of process may be more important that the actual methods and tools of design thinking. What design thinking may do is give one the opportunity to challenge and redefine their own process while being exposed to new tools.

Thus, in an attempt to simply find "design thinkers," we recruited participants from highly analytical engineering programs (i.e. Fluid Mechanics, Thermodynamics) and highly designerly programs (i.e. Product Design, Mechanical Engineering Design[2]) within the university. We then simply labeled participants as either "Analytical" or "Designerly" based on their university program.

Although this is a simple approach, it begs the question as to whether we are overlooking and subverting the very intent of our research. By simply choosing students within analytical or design based programs, one will be quite challenged to figure out if seemingly dispositional traits are not products of the educational environment and the context. While this may be true, what this simplification does allow us to do is to begin exploring the differences in perceived design thinkers and perceived non-design thinkers. This is an important distinction because the identification of an individual as a design thinker does not seem to have clear indicators. Of many potential indicators, the one indicator that has been shown to have some impact on design performance is an individual's ability to empathize with others (Kress and Schar 2012). However, while there are few if any well-defined characteristics, many people often identify themselves and others as design thinkers.

[2] While Mechanical Engineering Design may seem like a highly analytical field, the program at Stanford is highly enveloped in "design thinking."

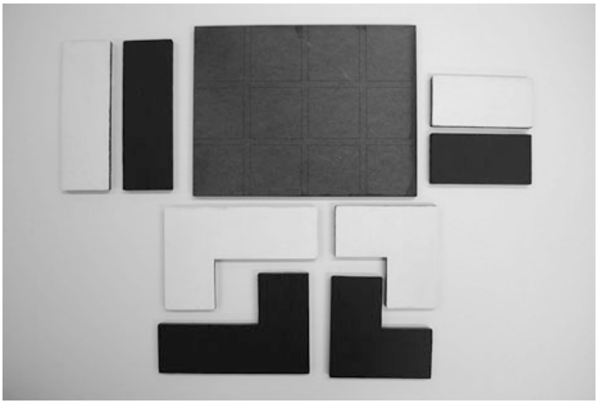


Fig. 3 Lawson block task game board

3.2 Tasks to Characterize Design Process and Design Outcomes

Lawson's original study consisted of participants arranging a single story of blocks with vertical edges of either red or blue. The participant was asked to create a layout with all the outer edges being the same color. The task had "hidden" rules unknown to the subject, making some configurations wrong. As participants worked through the task, they were told if their solutions worked or not and then received a score of the number of correct color vertical sides for appropriate solutions. While this task is not a true design task, it is a task where participants were required to create a physical solution with a clear goal, but with initially ill-defined constraints and relationships (Fig. 3).

Lawson conducted this test with a group of final year architecture and science students and found that the science students quickly tried many different block combinations in order to gather information and discover the hidden rule. Once they discovered the rule, they would optimize their solution. On the other hand, the architecture students focused on achieving the best correct color perimeter and would alter their solution if the combination were not acceptable. Lawson described the scientists approach as "problem focused" and the architects approach as "solution focused." From this, Lawson concluded that designers were more likely to have a "solution focused" cognitive style, however, even he was unsure as to whether this was a personal trait of designers or a learned attribute. In a follow up study, the same color block task was conducted with first year architecture students and students just entering university. The problem solving styles of these two groups showed no consistency within each group, suggesting that the different education styles of the scientists and architects was a determining factor in shaping their cognitive style.

While Lawson's task is quite good at teasing out cognitive style from a participant, it is far from a true design problem. In order to explore and assess the impacts of any of our interventions on design outcome we need a task that is easily controlled in a lab setting but is also less constrained and not easily optimized. In studying team process and outcomes, Wooley (2009) developed an architectural design and build task wherein participants would create a house, car garage, and pool using Lego™ bricks. Scoring of the designs was based on structure, aesthetics, durability and cost. The scoring was designed to force trade-off to be made such as cost over durability, that made the problem more like real, open-ended design problems. Additionally, it prevented any clear optimizations and invited many different solutions.

Using each of these tasks we can create a "petri dish" to explore both cognitive process and design outcomes given the influence of dispositional and situational factors.

3.3 Testing Our Approach Through an Experiment: Experiment on Problem Oriented vs. Solution Oriented Process Priming

To begin to understand these mechanisms may require various types of inquiry on our part. We have begun our inquiry by developing prototype, controlled experiments to see the effects of various design thinking tools on individual's design processes and design outcomes. We have developed prototype study exploring problem oriented vs. solution oriented process priming along with problem vs. solution dispositions on design process and outcomes. For this experiment we expected to find differences between the process and design outcomes of problem and solution oriented participants. We also expected to see difference between participants that were primed with a solution-oriented process vs. participants primed with a problem-oriented process. Lastly, we anticipated seeing a varying degree of effect of the priming dependent on the disposition of the participant. For example, problem-oriented participants primed with a solution-oriented process would perform differently from problem-oriented participants with a problem-oriented process.

3.3.1 Participants

Taking example from Lawson, we decided to use the Problem-Solution spectrum of cognitive styles to separate solution-oriented and problem-oriented thinkers based on their degrees of study. For example, Product Design students were considered to have solution-oriented dispositions and Fluid Mechanics students were considered to have problem-oriented dispositions. Students were recruited from Master's levels programs around campus and thus we attributed solution-orientated programs as more designerly and problem-oriented programs as more scientific/

engineering focused. We recruited 35 students, 21 science/engineering students and 14 design students.

3.3.2 Procedure

Before beginning the experiment, participants completed a short demographic and pre-task survey. After the pre-questionnaire, we primed each participant with a specific problem or solution-oriented process by having him or her follow a set of steps through a short redesign task. Specifically, the participants redesigned a soup can. The soup can redesign task processes altered the activities and times spent on activities that the participants took in generating a new soup can design.

Problem-oriented process	Solution-oriented process
5 min: Problem analysis	5 min: Ideate
3 min: Synthesis	3 min: Iterate
2 min: Evaluation	2 min: Select and argue

The problem-oriented process gave the majority of the time to problem analysis and then short sections on synthesizing a solution and evaluating that solution. The solution-oriented process gave the majority of the time to ideation and then short sections on iterating on the ideas created during the ideation and selecting and arguing for one idea from the iteration. The intent of this priming activity was to set a context for the following activities.

Directly following the priming activity, we had participants complete the Lawson block activity. After completing the Lawson block activity we then had participants complete a short questionnaire asking about affect and cognitive load. Next, participants completed the more open ended a design and building activity. The activity was modified from the architectural design activity used by Wooley (2009). Participants were given the task of building a model house based on a variety of design goals and constraints. Participants were given 20 min to plan, design, and build their model homes. After completing the architectural design task, participants were given another affect and cognitive load questionnaire.

3.3.3 Measures

Before receiving any priming or completing any of the study activities we administered a questionnaire to measure various aspects of the participant's personalities. We used the Triarchic Theory of Intelligences instrument (Sternberg 1985) to see if any differences existed between analytical participants and design participants and their preferences for analytical, creative, or practical thinking. We also asked on a scale of 1–10 how much each participant felt they were a designer and how much they felt they were an engineer.

During the Lawson activity, we measured the number of iterations, number of valid solutions, number of invalid solutions, final score, and score per iteration. In addition to these quantitative measures, we observed the general problem strategy participants took. For the Wooley architectural task we measured the participant's time spent planning and their design outcome scores in the following categories (1) Structure, (2) Quality, (3) Aesthetics, (4) Constraints Met, (5) Cumulative Score.

In addition to these outcome behavioral and outcome measures, we also measured the participant's affect between tasks using a 10-point Self-Assessment-Manikin Scale (Lang 1980). We also measured perceived cognitive load using the NASA-TLX scale (Hart and Staveland 1988) after each task.

3.3.4 Results

Quantitative Measures

Analysis of the Triarchic Intelligence Inventory showed no significant patters of intelligence preference between either analytical or design participants. There were no significant differences for engineering self perception ("How much of engineer are you"). There was a significant difference ($p < 0.05$) for designer self perception ("How much of a designer are you?") between analytical and designer backgrounds with those with designer background rating themselves higher as designers ($M = 7.21$, $SD = 2.05$) over participants with analytical backgrounds ($M = 5.86$, $SD = 1.77$).

ANOVA testing of number of iterations during the Lawson task showed a significant main effect due to treatment ($F = 6.148$, $p = 0.019$) and a marginally significant main effect due to Disposition ($F = 3.233$, $p = 0.082$). The main effect shows that both designers and analytical participants made more solution attempts after the problem-oriented priming treatment and fewer solution attempts after solution-oriented priming treatment. In addition, analytical participants tried more solutions overall than designers through this was only marginally significant ($p < 0.10$).

There was a significant main effect due to treatment ($F = 6.280$, $p = 0.018$) *for rule finding*. More participants found the rule during problem-oriented treatment. There is also a marginally significant difference based on participant disposition ($F = 3.254$, $p = 0.081$) with designers finding the rule more often than analytical participants. All other measures showed no significant results.

Results of the architectural design task showed a marginally significant main effect for aesthetic score based on treatment ($F = 2.959$, $p = 0.096$). The aesthetic scores were higher during solution-oriented treatment for both analytical and designer participants. Scores or quality, structure, constraints met, and cumulative score were insignificant. Difference in planning time between groups and treatments was also found to be insignificant.

No significant results were found for either affect or cognitive load measures.

Qualitative Observations

During the Lawson block test, we observed similar behaviors to Lawson's observations *with* analytical participants taking a more problem-oriented approach and the design participants taking a more solution-oriented approach. Specifically, analytical participants tried significantly more solution iterations than the design participants. Problem-oriented participants would try out many solutions, changing just one block at a time in an attempt to discover the hidden rule and optimize their score. The designer participants took a more *solution-oriented* approach, changing many blocks at once to find a global optimization. Ultimately, there were no significant differences between the two groups final scores. As with Lawson's results, we attribute these differences in process to the educational programs that each participant was from.

3.3.5 Discussion

The insignificant results for the Triarchic Intelligence Inventory suggest that intelligence preference may not be deterministic of one's design thinking abilities. Insignificant results for affect and cognitive load also suggest that neither population nor treatment method induced serious alteration in participant's mood or how hard they had to work during the task.

While the Lawson block task allowed us to observe process in a controlled manner, the task itself is quite far from the real-world design tasks involving various competing aspects of design. We attribute the higher number of solution iterations for analytical participants and problem-oriented process participants to the optimization nature of the task. The fact that both groups were able to achieve high scores suggests that both methods of problem solving can be effective. Wooley's architectural design task is a more realistic design and build activity and has outcome metrics similar to real products, such as aesthetics, durability, function, and interaction. None of the metrics showed significant differences among the participants. However, participants who were given the solution oriented priming activity had marginally better aesthetic scores. This may suggest that aesthetics may be more important when considering the entire solution of a design problem and may be one characteristic that does differentiate between analytical and design thinking styles.

Qualitatively, our results from the Lawson block task agree with Lawson's results, showing that students with different backgrounds (analytical vs. designer) utilize different problem solving mechanisms. Our results from the Wooley activity agree with Kruger and Cross, showing no real differences between problem-oriented and solution-oriented problem solving styles. It appears that our priming of process does not have a large enough impact to alter the problem solving style of the participants. Overall, problem vs. solution orientation may not be the best delineation between designerly and non-designerly cognitive styles. There may be too much overlap with process and background to truly separate the two for use in a laboratory study.

Moving forward, we feel that there are a variety of areas where we can improve our study design. The use of the Lawson task may be a better way to separate participants based on cognitive style rather then be used as a measurable design task. The Wooley architectural task appears to be a solid foundation upon which to examine problem-solving processes as it does have direct design outcomes and is a much more realistic design problem. As for manipulations to context, more in-situ interventions may have a stronger effect than the priming intervention we have tried. This aligns with the ideals of design thinking tools and methods as they are often introduced and used while working directly on a project rather than before working on a project. For example, a future study may alter the environment that people work in, or the questions that the participant is asked during the design task. Additionally, we argue that team composition would have a drastic effect on the outcomes of the study. One area of inquiry would be to see how mixed teams (designers & analytical participants) would compare to one-sided teams.

Of all the areas to improve however, the largest area would be to understand how to characterize individuals who are more aligned with design thinking methodologies. We found that simply binning our participants into analytical or designer to be not be as clean as originally intended. New methods for characterizing participants may be based off of traits such as empathy or openness to change.

4 Conclusion

Disposition and context each impact the behaviors of designers and ultimately the design outcomes of their work. However, it is still unknown to what degree each plays in the final outcome of design projects and the success of design teams. Through this research, we have highlighted the tension in our field between situational and dispositional design performance determinants. We have also laid a foundation for studying the impact of context on designers through providing an experimental platform to test various context manipulations against dispositional factors. Lastly, we have provided a set of situational factors that are likely to contribute to design performance. Results of our prototype study suggest that short priming activities may not be the best intervention for aligning people with a more design-oriented process. Rather, more integrated interventions and contexts may have more impact on design process than something like a weekend design thinking introduction. Through creating integrated contexts that engage participants in reflection and push them to alter their own processes, we hope to show that context does play an important role during design and that we can use the tools and methods of design thinking to alter and improve the design outcomes of both individuals and teams.

References

Asch SE (1956) Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychol Monogr Gen Appl 70(9):1–70

Brown T (2008) Design thinking. Harv Bus Rev 86(6):84

Darley JM, Batson CD (1973) "From Jerusalem to Jericho": a study of situational and dispositional variables in helping behavior. J Pers Soc Psychol 27(1):100–108. doi:10.1037/h0034449

Doorley S, Witthoft S (2012) Make space: how to set the stage for creative collaboration. Wiley, Hoboken, NJ

Dow SP, Glassco A, Kass J, Schwarz M, Schwartz DL, Klemmer SR (2010) Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans Comput Hum Interact 17(4):1–24

Dow S, Fortuna J, Schwartz D, Altringer B, Schwartz D, Klemmer S (2011) Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In: Paper presented at the proceedings of the 2011 annual conference on human factors in computing systems

Eris Ö (2004) Effective inquiry for innovative engineering design: from basic principles to applications. Kluwer Academic, Norwell, MA

Goldsmith RE (1986) Personality and adaptive-innovative problem solving. J Soc Behav Pers 1 (1):95–106

Kirton M (1976) Adaptors and innovators: a description and measure. J Appl Psychol 61(5):622–629. doi:10.1037/0021-9010.61.5.622

Kress GL, Schar M (2012) Teamology, the art and science of design team formation. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Berlin, pp 189–209

Kruger C, Cross N (2006) Solution driven versus problem driven design: strategies and outcomes. Des Stud 27(5):527–548

Lang PJ (1980) Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski JB, Johnson JH, Williams TA (eds) Technology in mental health care delivery systems. Ablex, Norwood, NJ, pp 119–137

Latane B, Darley JM (1968) Group inhibition of bystander intervention in emergencies. J Pers Soc Psychol 10(3):215–221. doi:10.1037/h0026570

Lawson BR (1979) Cognitive strategies in architectural design. Ergonomics 22(1):59–68

Mehta R, Zhu R (2009) Blue or red? Exploring the effect of color on cognitive task performances. Science 323(5918):1226–1229

Milgram S (1974) Obedience to authority. Harper & Row, New York

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. North Holland Press, Amsterdam

Plucker JA, Beghetto RA, Dow GT (2004) Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. Educ Psychol 39 (2):83–96. doi:10.1207/s15326985ep3902_1

Ross L (1977) The intuitive psychologist and his shortcomings: distortions in the attribution process. In: Leonard B (ed) Advances in experimental social psychology, vol 10. Academic, New York, pp 173–220

Ross L, Nisbett RE (1991) The person and the situation. Temple, Philadelphia, PA

Schön DA (1983) The reflective practitioner: how professionals think in action. Basic Books, New York

Scott SG, Bruce RA (1994) Determinants of innovative behavior: a path model of individual innovation in the workplace. Acad Manag J 37(3):580–607

Sternberg RJ (1985) Beyond IQ: a triarchic theory of human intelligence. Cambridge University Press, New York

Treffinger DJ, Isaksen SG, Dorval KB (1994) Creative problem solving: an overview. In: Runco MA (ed) Problem finding, problem solving, and creativity. Ablex, Hillsdale, NJ, pp 223–236

Woolley AW (2009) Means vs. ends: implications of process and outcome focus for team adaptation and performance. Organ Sci 20(3):500–515

Theaters of Alternative Industry: Hobbyist Repair Collectives and the Legacy of the 1960s American Counterculture

Daniela K. Rosner and Fred Turner

Abstract This chapter describes initial results from an ethnographic study of design and engineering engagements in community-operated sites at which hobbyists mend and repair mass-produced goods. We conducted participant observation at seven repair events and two collectives in the San Francisco Bay area where consumer electronics are reassembled, and spoke with approximately eighty repair practitioners. Here we describe surprising connections between repair and social movements that, in turn, reveal deep ties between contemporary hobbyist repair and countercultural design practices of the 1960s. These links, we argue, open new and important areas for design research.

1 Introduction

Errors, omissions, and failures underlie almost everything we do. Our cell phones inevitably break, our software becomes outdated, and our appliances wear out. In response, we fix and maintain what we already have; we upgrade our software and replace broken parts, often in highly creative ways. For example, bookbinders have both restored and transformed books for centuries (Rosner 2012). Likewise, hobbyists have used broken artifacts to spur design innovation (Tanenbaum et al. 2013). One has turned over-wound alarm clocks into a guitar amp (Repplon 2008); another has converted a broken desk lamp into a sleek iPhone stand (Ikeahackers 2012). In each case, the breakdown of one technology created an occasion for making something entirely new.

D.K. Rosner (✉) • F. Turner
Program in Science, Technology, and Society (STS), Stanford University, Stanford, CA, USA
e-mail: danielar@stanford.edu; fturner@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_5, © Springer International Publishing Switzerland 2015

Still, breakage and repair tend to be overlooked as important sources of technology design and innovation. We conceptualize repair as the process of sustaining, managing, and repurposing technology in order to cope with attrition and regressive change. Building on our prior investigations of countercultural and hobbyist design movements (Turner 2006, 2009a, b; Rosner and Bean 2009; Rosner 2013, 2014) and a growing body of scholarship on repair (Henke 1999; Jackson et al. 2012; Jackson 2013; Orr 1996; Rosner and Taylor 2011; Suchman 1987), we have conducted a detailed ethnographic study of repair collectives in the San Francisco Bay area. This study has revealed unexpected and surprisingly extensive ties between the repair and redesign of industrial technologies and the ideological legacy of the counterculture. By exploring those legacies here, we hope to show two things: first, that repair, like innovation, is an integral part of the process of technological design and development, and second, that the ideals of the counterculture continue to shape design practices in the San Francisco Bay area, and potentially, far beyond it.

1.1 Why Study Repair? And Why Study Hobbyists?

The study of repair cultures grows out of a body of research in science and technology studies focused on the social contexts of innovation and technology use, particularly in the case of information technology. A small but vibrant ethnographic tradition has emerged around the study of everyday maintenance. For instance, Lucy Suchman, Julian Orr and colleagues have turned to the lives of photocopy machine repair workers to illuminate the limitations of codifying maintenance techniques (Suchman 1987; Orr 1996). Orr's influential accounts of individual diagnoses of machine malfunctions have exposed skilled service work as "necessarily improvised, at least in diagnosis, and centered on the creation and maintenance of control and understanding" (Orr 1996, p. 161). Orr has shown how repair workers not only use manuals and codified organizational knowledge, but also rely on the retelling of "war stories"—personal accounts from the field often shared over lunch or informal meetings. As Orr's work suggests, every repair activity involves situated actions whose intent, in Suchman's terms, "must be contingent on the circumstantial and interactional particulars of actual situations" (Suchman 1987, p. 186).

Beyond IT development, analysts have focused on maintenance work to reconsider features of building reconstruction (Brand 1994), vehicle repair (Crawford 2010; Dant 2010; Harper 1987; Van Maanen 1990), electricity procurement (Graham and Thrift 2007), craft practice (Sennett 2008; Rosner 2012), routine workplace activities (Henke 1999), and shared infrastructures (Star and Strauss 1999). Other studies have considered mending conversational breakdowns as a critical form of repair, as in Garfinkel's (1967) experiments designed to break social norms in order to study how people respond and restore common understandings. Others have studied the arcana of free software through the continuously rewritten fabric of the Internet (Kelty 2008). Most recently, Jackson et al. (2012) has traveled to Namibia to explore

IT repair cultures where programmatic interventions create policy barriers and problems of control that complicate local repair efforts.

Together this scholarship has introduced two views of repair. On the one hand, it has demonstrated a largely unacknowledged connection between repair work and creativity. It has also illustrated how repair leads to different ways of understanding technological change, particularly when reuse and maintenance become necessary (Burrell 2012; Jackson et al. 2012). On the other hand, this work has pointed to a broader blurring of boundaries between leisure and professional labor of which repair is an integral part (Crawford 2010; Sennett 2008). The cases presented in this chapter begin to broaden these perspectives by illustrating what happens when the forms of creativity and labor that arise from repair become entangled with 1960s countercultural ideologies, especially when such ideologies get embedded in contemporary hobbyist design movements and high-technology industries.

Given our emphasis on repair as it relates to design innovation, it might seem more sensible to study professional repair workers rather than hobbyists. Yet, we've found that in many cases, it is hobbyists doing the innovating. Just outside the institutional walls of design consultancies and corporations, a growing number of makers are extending and defying conventional notions of creative production. Whether we call them "geeks," "makers," or "hackers," a new generation of amateur technologists and designers has emerged (Kelty 2008: 35). Moreover, while we often think of repair work as organized by professionals in factories, fabrication labs, and other sites of material experimentation, in these settings we see repair organized by particular interest groups and communication media. Repair activities coalesce around mailing lists and Twitter feeds, hacker spaces and fair grounds, often inspired by a do-it-yourself ethos. Their interests are well represented in the mass media too, especially in *Make* magazine. As Faith Levin and Cortney Heimeri have shown in the film and book *Handmade Nation* (2008), this "new generation" of amateur makers celebrates different facets of everyday creative work. From building circuitry and upgrading software to fashioning shoes and screen-prints, Levin and Heimeri show that makers "are reshaping how people consume and interpret the handmade" (Levin and Heimerl 2008: xi).

1.2 Research Methods

Several overarching questions have guided our study:

1. What are the range of practices, technologies and programs that support or subvert specific repair activities? How do these practices evolve over time?
2. What role does background knowledge of design practice play in makers' repair work? Conversely, how does repair work shape makers' other design practices?
3. What resources do fixers rely on to produce or police the social and technical resources necessary for repair? What adjustments do fixers make in different repair situations?

In order to investigate these questions we took a qualitative, ethnographic approach.[1] We began the study by observing fixers' practices in their own environments and documenting them through a combination of video, audio, photos, and field notes. We participated in and observed a range of repair and maker collectives in the San Francisco Bay area, including an annual convention of Macworld, the East Bay Mini Maker Faire, the San Mateo Maker Faire, and meetings of the Dorkbot collective, a loosely affiliated group of artists, inventors, designers, and engineers. We engaged in informal conversations at these events with roughly 60 participants.

We complemented our ethnographic work with extensive formal interviews with 20 participants whose repair activities have critically informed the development and maintenance of contemporary repair movements. Our interviewees included leaders of pop-up repair groups such as the Fixit Clinic and the Repair Café, participants in public repair workshops and nonprofit collectives with strong links to community-operated workspaces for electronics tinkering, and organizers of related technology development endeavors such as Partimus and the Flaming Lotus Girls. Lastly, we conducted in-depth research in the Fixit Clinic and Repair Café's online archives and in individual participants' collections of artifacts and writings.

2 What We've Learned So Far

Our initial research has revealed a surprising connection between repair work and social movements associated with environmentalism and sustainability. We began our work focusing on the interactions of hobbyists with particular devices, with the assumption that design innovations would emerge out of interactions between the makers and the technologies with which they worked. But we soon saw that our subjects had taken up the practice of repair within a rich conceptual and even political framework. Participants believe that their acts of repair constitute interventions in large-scale social processes and that they can have effects far beyond their local setting.

[1] Qualitative methods characterize causal processes, recognize new phenomena, present auxiliary evidence for existing hypotheses, and identify counterexamples (Burrell and Toyama 2009). Unlike statistical methods, qualitative methods are good at pinpointing what about people's lived experiences of repair is important and why (Bauer and Gaskell 2000). Through long-term observation and interviews we can examine why people choose to repair some possessions and discard others, and how certain artifacts achieve heirloom status. We cannot make representative claims, test hypotheses, reveal trends, or answer questions of how often and how much — aims that qualitative methods are ill-suited to address. Instead, we seek to produce "observable-reportable" (Garfinkel and Sacks 1970: 342) understandings of the practical (and practiced) work of repair.

This ideological framework represents a blending of the legacy of the counterculture of the 1960s (Turner 2006, 2009b) and of the practices traditionally found in craft communities (Rosner 2012, 2014). More specifically, it echoes a design ideology that permeated the New Communalist wing of the American counterculture: Buckminster Fuller's "comprehensive design" (Turner 2009b). First articulated in a 1949 essay that was reprinted and widely circulated in Fuller's 1963 volume *Ideas and Integrities*, the doctrine of comprehensive design solved a problem for the young adults of the 1960s. To the post-war generation, technology presented two very different faces. On the one hand, large-scale military technologies such as fighter planes and aircraft carriers and above all, the atomic bomb, threatened to destroy the planet. On the other hand, consumer technologies produced by the same military industrial complex such as transistor radios and automobiles and even LSD, provided extraordinary individual freedom and personal satisfaction. To the young longhairs of the counterculture, a question hung in the air: How could a person embrace small-scale technologies and at the same time, turn away from mass industrial processes and the threat of war?

Buckminster Fuller offered an answer. Technology itself was not the problem, he explained. On the contrary, the problem was one of design and resource allocation. Too many of the world's natural and technological resources were concentrated in military hands, he said. Yet, independent individuals could act to reshape the world system by taking the technologies developed in the industrial sphere and putting them to work in their own lives, on behalf of a more egalitarian way of living. In short, they could become "comprehensive designers" of their own lives, and of a better world (Fuller 1963: 173). Between 1966 and 1973, thousands of young counterculturalists took up Fuller's vision. They built geodesic domes on the plains of Colorado out of old car tops and transformed industrial plastic sheeting into windows on everything from houses to cribs. They saw their work as simultaneously material and symbolic. By repurposing the products of industry, they would remake their own lives and show others how to change the world.

2.1 Comprehensive Design and Repair

In many ways, today's repair practitioners are following in the New Communalists' footsteps. To see how, consider the case of artist and activist Miriam Dym. In December of 2011, she founded *Dym Products*, an eccentric enterprise dedicated to celebrating (and questioning) re-use and repair. The business came to life in a series of unusual and largely unviable product design initiatives bearing such names as the *Suboptimal Object Project* (a collection of abject, incomplete works), the *Logo Removal Service* (a service for replacing logos on tee-shirts, hats and bags with colorful textile shapes and contrasting stitching), and the *Infinite Stripes Project* (an upcoming performance of continually painted stripes on fabric).

Dym developed each project to explore the relation between meditative, considered craftsmanship and strained manual labor, and did so in several ways.

By including half-spun baskets and an incomplete set of lamps in the *Suboptimal Object Project*, she drew links between the well-made and the unprofessional. While preparing to dye endless stripes on old upholstery and other used fabrics for the upcoming *Infinite Stripes Project*, she troubled notions of domesticity, manual labor, and convenience ("a bit of a joke on buying a painting to match your couch," she explained). Dym described the *Suboptimal Object* and *Infinite Stripes* projects as both art (e.g., painting) and utilitarian (e.g., textiles), a framing she used to unsettle longstanding distinctions between the two production processes and raise questions around the visibility of manual labor.

The *Logo Removal Service*, on the other hand, served to challenge the aesthetics of branding, a slightly different political project. Low on clothing, Dym was delighted when a friend gave her an extra tee shirt from the launch of a local start-up. She wore the shirt and visually appealing logo until reactions to the shirt began to change. The company took off, and the logo became instantly recognizable, leaving Dym feeling rather uncomfortable: "I didn't have a strong enough opinion to back up the claim I had across my chest" ("*Logo Removal Service*," 2013). To preserve the utility of the shirt, but remove the corporate affiliation, Dym cut out a shape around the logo, and replaced it with a scrap of colorful fabric. "That new shape held something in a way than an abstract shape can," she explained (Dym 2013). It held a critique, both of the aesthetics of branding and the process that would someday lead the fabric to the landfill and pave the way for obsolescence.

Dym believed that mass produced goods could help people imagine a more human manufacturing scale. While stitching her son's tattered jeans at an exhibition of her repair work, Dym commented on the irony of being a middle-class woman with three Ivy League degrees willing to spend hours mending her son's cheaply produced H&M trousers. She described the resulting mend as of higher quality than the original manufacturing job: "It's a kind of statement about expensive labor provisionally fixing something made cheaply" (Dym 2013). In manipulating a mass-produced object and highlighting her intervention in brightly colored thread, she slowed down the production process to draw attention to the artistry and manual labor with which it was made. She felt that by reducing the production volume in favor of what is produced, people could become accustomed to repairing or repurposing what they have.

In this regard, Dym's repair work echoed the practices of the 1960s New Communalists. Like them, she worked to transform the products of a mass-industrial production system into tools for personal and collective transformation. Unlike them, however, she also blended concerns for visual aesthetics with the idiosyncrasy of "expensive labor." She even posited repair as an entrepreneurial interest as well as a conceptual framework for her artistic practice. She explained, "I feel like if I'm going to be in business I need to acknowledge the mass production. And if I'm going to be an artist I need to acknowledge the mass production *and* I need to try to compete with machines in the way that chess players compete with an IBM machine… So it's completely quixotic" (Dym 2013). To compete with machines meant trying to "become the factory," a project without end and without direct practical impact. Dym's material interventions produced a paradox of time and material investment that transformed the work of repair into something commercially less-than-effective but symbolically powerful.

Though it may seem odd that a woman would want to challenge the global economic system by stitching her child's pants, Dym's ideas were not new to California, nor even the surrounding art world. In fact, it was during the 1998 Los Angeles MOCA exhibition "Out of Action: Between Performance and the Object, 1949–1979," that Dym discovered the elusive power of public-facing performance art, work that integrated object production with a political agenda. Struck by how effectively a performance could convey a political message through subtle, often indirect means, Dym began shifting her art practice toward the performative—and in the late 1990s she decided to stop throwing things away. Following process artists of the 1960s and 1970s, she celebrated the beauty of waste by composting orange peels and stitching old shoes.

Yet, this philosophy of activism was not identical to what had come before. Dym described herself as the descendant of those who took to the communes 40 years ago and as what she called a "proto-hippie": "Someone who's a hippie now, and not a hippie like it was in the 1970s. They know about marketing and have a website . . . availing themselves with the latest technologies, they weren't trying to go back to the farm to change the world" (Dym 2013). Dym saw her efforts to interact and engage with the public as an entrepreneurial and environmental act. In building a business around dying and stitching, she critiqued industrial processes of planned obsolescence and made these arguments known to the world at large. As we will see in other pop-up sites for repair, it is in this semiotic, ritualized display that practitioners orient repair toward a countercultural conceptual framework for social change.

2.2 Beyond the Individual: Repair as Conceptual Framework

As a practitioner invested in the meeting of art and engineering through repair, Dym embodied a philosophy shared by many actively participating in what we call public sites of facilitated repair. These sites include "pop-up" events like the Fixit Clinic and Repair Café in which repair-savvy volunteers help local residents disassemble and fix their broken things: toasters that no longer heat, iPhones with shattered screens, and electronic games that cease to play. Since 2009, the events have occurred at museums, libraries, community centers, and the like, roughly once a month in the San Francisco Bay area. They engage people in repair at no cost, though visitors can sometimes offer a donation.

We first saw links between repair and the politics of sustainability in the East Bay Fixit Clinic and neighboring hackerspaces such as Noisebridge, a community-operated workspace in the San Francisco Mission District, where activities focused on motivating reuse through electronics tinkering. Members raised questions of electronic waste ("e-waste") in particular. They wondered how devices should persist as they became no longer usable, serviceable, trendy or desirable. Their questions framed and sometimes motivated volunteers in their repair efforts. At a meeting of the Post-Waste Nexus, a collective launched at Noisebridge, members discussed their project as "techno-activism," circumvention through consensus decision-making to promote the re-use of broken and abandoned hard drives, cell phones and the like. For Chris Witt, a Fixit Clinic volunteer, participation at the Fixit Clinic was part of "being nice to the world that give us life." It makes more sense, he explained, "to fix or alter or somehow reengineer an existing resource than it does to chop down a whole new resource and mine it and create all the toxic—in all the senses of the word—aftereffects or side effects that come with new construction. It makes more sense to me to use what we've got instead of throwing it away and creating a new one" (Witt 2012). For his part, Witt saw the work of repair as advancing environmental stewardship in addition to fostering an alternative relationship with the factory floor.

Yet, the Fixit Clinic organizers were initially skeptical of their interventions. As Peter Mui, the founder of the Fixit Clinic explained, "the first time we had one [a Fixit Clinic] I thought we'd have a big pile of e-waste in the corner" (Mui 2012). Yet, no such pile emerged. Instead, volunteers helped participants replace fused and bonded batteries in electronic toothbrushes and oil sewing machine gears. As trained and amateur engineers, they saw their work to repair and tinker with electronics as par for the course—or as Mui explained, "I personally don't know anybody who became a maker who wasn't a fixer first." ("Open Make @ The Hall: Cities 1/19/2013," Google+ video, 2013). The Fixit Clinic provided a means for members of the public to unearth how designers and engineers have contributed to the world by making the products they use on an everyday basis and prompting them to figure out how engineers achieved what they set out to make. The volunteers viewed repair, in this sense, as an integral part of industrial design and engineering.

Yet, for the volunteers, returning functionality to devices also did something more. It saved the devices from the landfill and minimized motivations for further consumption, which could eventually lead to more waste. To do this, they used their own tools and supplies as well as digital resources: online hobby shops such as iFixit.com that distribute tools, parts, and video instructions for fixing consumer electronics from the web. Using these physical materials and online resources, the volunteers searched for spare parts, identified the requisite instruction manuals, and dove into repairs. Their work made new purchases less necessary by offloading some of the purchasing (or "conspicuous" consumption) on the hunt for replacement parts.

As the repair efforts of the Fixit Clinic and the Repair Café sent people home with working devices, they received new attention from an international community concerned with ecological waste. Traces of success circulating on dedicated websites and social media outlets like Facebook and Twitter enabled pundits and media outlets to follow fixing events on the ground. As Peter Skinner, the founder of the Palo Alto Repair Café, noted, "it was more about being part of this global network. I got contacted from New Zealand asking about starting one of these, and someone up in Calgary. And other people locally about how to kick off something like this. I don't know what they found on our website... but it's nice to be part of this larger [movement]" (Skinner 2013).

In addition to connecting engineering and art practice, Mui saw his Clinic as a call for social change:

> I really want to demystify science and technology. And my alternate surreptitious goal is that I'm hoping at some point we'll be able to make better policy choices as a society. And so the classic example I give is, and it may be apocryphal: In Japan right now, if you buy and you make an appliance, the manufacturer of the appliance you're getting rid of has to come to your house and remove it and recycle it properly. So they truly have cradle-to-grave ownership of the device. It certainly changes their incentives about how they manufacture something. They don't want to get back [the device] prematurely (Mui 2012).

Mui first became interested in repair while doing "goofy things" with his father's train set (Mui 2012). Now however, he believed that tinkering and disassembly could challenge the cultural apparatus of electronic waste and reveal the mechanisms underpinning technical progress. His curiosity had become political and he hoped that his repair work would serve as an example for others.

3 Repair as a Social Movement: Insights for Design Researchers

Beyond device-level design, we found the extent to which the Fixit Clinic and the Repair Café participants connected their repair practices back to social movements rather striking. For many, repair was not only appealing as a manual process of manipulating wires and screws, but also as a mode of political action. In that sense, we believe that the amateur repair communities offer a powerful reminder that design is shaped by historical forces that swirl far beyond the interactions of designers and their materials. In this case, we saw repair workers such as Miriam Dym turning the products of global industry into displays of potential alternatives to that industry. Like the New Communalists of the 1960s, Dym and her cohort are actively seeking to redesign not only goods, but their lives. In the process, they too hope to rebalance political and ecological forces they believe have gone out of whack. The work itself matters only in small part for the goods it produces. It matters much more as a performance of an alternative mode of industry and a more person-centered way of life.

At the same time, unlike the New Communalists, today's repair workers are not heading back to the land. On the contrary, they are creating temporary arenas in which to gather and work together. Like the communes, these clinics are in some sense cities on a hill. They are meant to demonstrate the power of creative re-manufacturing to change the world—here and now for the moment, but over time perhaps, everywhere. They are in fact theaters of alternative industry.

What then is likely to become of their performances? In the 1960s, the New Communalists failed to transform the American political landscape. Yet, they went a long way toward helping Americans re-imagine design as a simultaneously material and political practice. Today’s makers and fixers are once again asking critical social questions: How can devices become the centers not only of individual creativity, but egalitarian community? How can designers help make not only things but whole societies work better? What role should aesthetics play in shaping collective action? And what roles should our collective ideals play in shaping our designs?

It’s too early to tell if the citizens of the Fixit Clinics and repair collectives will succeed in answering these questions. For now however, we are confident that participants have gained a new awareness of the political potential of small-scale design by tinkering with industrial devices. They have also begun, however quietly, to integrate the contemporary work of design and engineering into the San Francisco’s Bay area’s longstanding pursuit of social change.

Acknowledgements The authors would like to thank the Hasso Plattner Design Thinking Research Program HPDTRP for funding this project. We would also like to thank the participants in the repair collectives we observed and our research subjects who made this work possible.

References

Bauer M, Gaskell G (2000) Qualitative researching with text, image and sound. Sage, London, pp 336–350

Brand S (1994) How buildings learn: what happens after they’re built. Penguin, New York

Burrell J (2012) Invisible users: youth in the internet cafés of urban Ghana. MIT Press, Cambridge, MA

Burrell J, Toyama K (2009) What constitutes good ICTD research. Inf Technol Int Dev 5(3):82–94

Crawford MB (2010) Shop class as soulcraft: an inquiry into the value of work. Penguin, New York

Dant T (2010) The work of repair: gesture, emotion and sensual knowledge. Sociol Res Online 15 (3):7

Dym M (2013) Interview with Daniela Rosner, 8 August

Fuller RB (1963) Ideas and integrities, a spontaneous autobiographical disclosure. Prentice-Hall, Englewood Cliffs, NJ

Garfinkel H (1967) Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ

Garfinkel H, Sacks H (1970) On formal structures of practical actions. In: Mckinney JC, Tiryakian EA (eds) Theoretical sociology: perspectives and developments. Appleton, New York, pp 337–366

Graham S, Thrift N (2007) Out of order understanding repair and maintenance. Theory Cult Soc 24 (3):1–25

Harper DA (1987) Working knowledge: skill and community in a small shop. University of Chicago Press, Chicago, IL

Henke CR (1999) The mechanics of workplace order: toward a sociology of repair. Berkeley J Sociol 44:55–81

Ikeahackers (2012) “Forså Camera Stand,” Ikeahackers.Blog. 21 June. http://www.ikeahackers.net/2012/06/forsa-camera-stand.html

Jackson SJ (2013) Rethinking repair: breakdown, maintenance and repair in media and technology studies today. In: Boczkowski P, Foot K, Gillespie T (eds) Media meets technology, MIT Press, forthcoming

Jackson SJ, Pompe A, Krieshok G (2012) Repair worlds: maintenance, repair, and ICT for development in rural Namibia. In: Proceedings of the 2012 computer supported cooperative work conference

Kelty CM (2008) Two bits: the cultural significance of free software and the internet. Duke University Press, Durham

Levine F, Heimerl C (2008) Handmade nation: the rise of DIY, art, craft, and design. Princeton Architectural Press, New York

Logo Removal Service. Dym products, accessed 9 August 2013. http://www.logoremovalservice.com/lrs-story/

Mui P (2012) Interview with Daniela Rosner, 5 November

"Open Make @ The Hall: Cities 1/19/2013," Google+ video, posted by "The Lawrence Hall of Science," 19 January 2013. https://plus.google.com/100725769399791437356/posts

Orr JE (1996) Talking about machines: an ethnography of a modern job. Cornell University Press, Ithaca, NY

Repplon J (2008) April 28 (2:40 am). "Clockamps—portable practice amps made from alarm clocks," The Steampunk Forum at Brass Goggles Blog. http://brassgoggles.co.uk/forum/index.php?action=printpage;topic=8540.0

Rosner DK (2012) Material practices of collaboration. In: Proceedings of the 2012 computer supported cooperative work conference, pp 1155–1164

Rosner DK (2013) Mediated craft: digital practices around creative handwork. In Buechley L, Peppler K, Kafai Y, Eisenberg M (eds) Textile messages: dispatches from the world of education and e-textiles, Peter Lang

Rosner DK (2014) Making citizens, reassembling devices: on gender and the development of contemporary public sites of repair in Northern California. Public Cult 26(1 72):51–77

Rosner DK, Bean J (2009) Learning from IKEA hacking: i'm not one to decoupage a tabletop and call it a day. In: Proceedings of the 27th international conference on Human factors in computing systems, pp 419–422

Rosner DK, Taylor AS (2011) Antiquarian answers: book restoration as a resource for design. In: Proceedings of the 2011 annual conference on Human factors in computing systems, pp 2665–2668

Sennett R (2008) The craftsman. Yale University Press, New Haven

Skinner P (2013) Interview with Daniela Rosner, 15 January

Star SL, Strauss A (1999) Layers of silence, arenas of voice: the ecology of visible and invisible work. Comput Support Coop Work 8(1):9–30

Suchman LA (1987) Plans and situated actions. Cambridge University Press, New York

Tanenbaum JG, Williams AM, Desjardins A, Tanenbaum K (2013) Democratizing technology: pleasure, utility and expressiveness in DIY and maker practice. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 2603–2612

Turner F (2006) From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism. University of Chicago Press, Chicago, IL

Turner F (2009a) Burning man at Google: a cultural infrastructure for new media production. New Media Soc 11(1–2):145–66

Turner F (2009b) Chapter 9: R. Buckminster Fuller: a technocrat for the counterculture. In: Chu H-Y, Trujillo R (eds) New views on R. Buckminster Fuller. Stanford University Press, Stanford, CA, pp 146–159

Van Maanen J (1990) Escape from modernity: on the ethnography of repair and the repair of ethnography. Hum Stud 13(3):275–284

Witt C (2012) Interview with Daniela Rosner, 3 December

Part II
Empowering Team Collaboration

Assessing the Development of Design Thinking: From Training to Organizational Application

Adam Royalty, Karen Ladenheim, and Bernard Roth

Abstract Increasingly organizations are turning to off-site design thinking professional development programs as a way to grow design competencies in their workforce. This paper has two main goals (1) to develop an initial assessment tool that helps identify how well organizations support employees' continued learning and application of design thinking. (2) To describe a process for constructing design thinking assessment tools. The assessment created is informed by an exploration of existing design thinking Executive Education programs and tested in a large organization committed to using design thinking.

1 Introduction

The focus of this work is to develop and test an assessment that captures the extent to which organizations support design thinking professional development. Little research has been done on this issue. The hope is that the assessment generated, and the process used to design it, will lead to more work in this area.

Companies around the world are turning to design as a driver of innovation. SAP, P&G, Intuit, and JetBlue have all expanded the role of design in their culture (Korn and Silverman 2012). They are leveraging design in order to tap into consumers in a more authentic way. The hope is that a human centered way of working will lead to breakthrough innovations. To boast their capacity for design, these organizations are not necessarily hiring more designers, instead they are training existing employees, many of which have no background in design thinking

A. Royalty (✉) • K. Ladenheim • B. Roth
Hasso Plattner Institute of Design, Stanford University, Stanford, USA

Stanford University, Stanford, CA, USA
e-mail: aroyalty@stanford.edu; karenl2@stanford.edu; broth@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_6, © Springer International Publishing Switzerland 2015

(Cross 2007). The aim is to have design as a competency that runs throughout the organization (Courage 2013). It is important to note that design as a competency goes beyond design skills like interviewing and prototyping. The goal is to instill design dispositions so that employees have the ability to behave like designers. This is why these companies are turning to design thinking.

This paper will look at one of the predominate design thinking training programs, the Executive Education Initiative at the Hasso Plattner Institute of Design at Stanford (d.school). The d.school Executive Education program began in 2007 with the goal of training business leaders in design thinking so that they can bring this paradigm back to their organizations. The initiative has developed a variety of offerings, but the most common is a 3 day "bootcamp" that introduces design thinking concepts and goes over how they can be applied in organizations. The participants are businessmen and women who take time off from their jobs to attend the program at Stanford. The expectation is that they will return and spread the design dispositions learned at the d.school throughout their organization.

Generally speaking, design thinking professional development can be split into two categories: the initial off-site training and the continued on the job application. Ultimately this work is about assessing an organization's ability to support the growth of design thinking within individuals and teams. It might seem that only the on the job application is worth studying. However, many of the companies that send participants through the Executive Education program have developed in house design thinking trainings based on the d.school's model of teaching (Courage 2013). This means that it is important to understand the initial training experience in order to provide context and a framework for assessing on the job application. Furthermore, there is evidence to suggest that the trainings are successful in instilling design dispositions (Royalty et al 2014). This means that we can begin to look for post program design thinking development by comparing it to the development we see during the program. However, there clearly is a shift in context from the classroom to the real world that we can expect to alter the way people learn design thinking. To mitigate this tension, the first step is to create a pilot assessment that captures how well an organization supports on the job application of design thinking based on the initial training experience. The next step is to then work with an organization to modify the assessment to better fit the context. To do this we must answer two main research questions:

1. How does the d.school Executive Education program support design thinking development?
2. How can we assess an organization ability to support design thinking development?

To answer these questions this paper presents two studies. The first study is an investigation of the Executive Education program practices culminating is an instructional model. The second study describes the construction and testing of an assessment tool that aims to capture how well a large service organization supports an interdisciplinary design team that is learning to apply design thinking.

2 Study 1: Constructing a d.School Executive Education Instructional Model

As was mentioned above, the Executive Education Initiative offers numerous professional development programs. This study focuses on the single most taught program, the design thinking bootcamp. This program has the most participants annually and is the most refined. It is taught three times a year and has between 50 and 60 participants each session. The duration of the program is 3 days at Stanford. The first two of which are spent working on a design challenge in teams of five executives with one d.school coach. The teams immerse themselves in the methods and mindsets of design thinking to solve a challenge. The third day is spent working on how to bring design thinking back into their organization. Typically each participant devises a game plan for how they are going to apply what they learned the first week they return to work.

The format of instruction follows a basic pattern. The teams go through a five-step design process. At the beginning of each step they receive a short lecture introducing the basic principles and techniques need to complete that step. From there they split into teams and work on their challenge until the beginning of the next step when they receive another short introductory lecture. Each lecture is approximately 10 min while each working session lasts between 1 and 2 h. Skills such as brainstorming and user testing are taught via the lecture, but the coaches facilitate the actual design thinking practice during the teamwork time. Given that most of the support takes place in the small teams, understanding how the coaches drive design thinking should be at the heart of this model.

Although small team facilitation has been studied (Hackman and Wageman 2005), there has been little work on coaching a design thinking process, especially in the professional development context. This is an exploratory, qualitative study that seeks to uncover the strategies d.school coaches use to successfully teach participants design thinking over the course of a 3-day workshop. One of the main assumptions is that there are similarities between coaches in terms of style and approach. This assumption is based on the fact that coaches regularly share practices with one another during the course of the workshop. Furthermore, because all teams receive the same global introductory lectures, no team can stray too far from the program.

2.1 Method

2.1.1 Participants

For this study the subjects were 20 d.school coaches who were working as a part of the March 2012 Executive Education bootcamp. As is customary, the coaches were a mix of d.school employees, d.school alumni, and former bootcamp participants who excelled at applying design thinking and returned to coach. Each coach had experience teaching design thinking workshops. Half of the coaches had 3 or more years of experience teaching design thinking.

2.1.2 Procedure

The study was conducted in a 1-h session during the preparation day immediately preceding the workshop. A researcher orally led the coaches through a guided reflection where they were asked to think about specific experiences teaching design thinking. There were a total of 19 prompts given to the subjects. They had approximately 2 min to respond to each prompt. The subjects captured their thoughts on a “journey map” (see Appendix).

2.1.3 Materials

The journey map used was a six-page paper packet. Five of the pages corresponded to the five steps of the process taught at the program. The sixth page had space for overall reflections. The three types of prompts (Table 1) allowed coaches to reflect on strategies for both cognitive and emotional outcomes of this program. Because the form of instruction is so experiential, cognitive outcomes alone would not be enough to describe this way of teaching. The journey map was piloted three times prior to the study with coaches who did not participate in the March 2012 program.

2.2 Results

Two researchers open-coded each journey map independently. Statements taking the form of individual sentences or sentence fragments were coded. Each of the nineteen prompts elicited between one and five statements per coach. After the first session the researchers came together and approximately 30 categories emerged.

In order to narrow down, the researchers decided to explore categories that fit into one of two overarching groups: coaching tactics and participant responses. This meant putting aside many of the categories focusing on single design thinking skills. Although interesting in of themselves, these extra categories did not seem to shed any light on the overall strategy of teaching design disposition.

The researchers individually combed through the 30 categories to find ones fitting into the overarching groups. They came together and settled on eight codes: five teacher tactics and three student responses. The coaching tactics are the most common actions coaches took to support their teams. These statements primarily emerged from the instructional strategies and goal prompts. Participant responses represent how coaches observe students responding to their facilitation.

Table 1 Selected journey map prompts

Question type	Example
Goals: what learning and emotional goals do coaches have for their students?	What are you trying to teach your students?
Learner perceptions: based on past experience, how do students respond during different steps of the design process?	What are participants' conceptions of design during prototyping?
Instructional strategies: how do coaches move students through the process?	What strategies do you have for teaching prototyping?

These primarily came from the learner perception prompts. These codes constitute what this paper suggests are the **instructional elements** of this program.

Lists of key terms appearing in the data were created as markers of each element. What follows is a list of the eight elements organized by group. Each element has a title, some of the key terms, an explanation of what it means based on the responses, and an example from the data.

2.2.1 Coaching Tactics

Title Discomfort

Key Terms Uncomfortable, intimidated, irritated

Explanation This is commonly referred to as pushing people beyond their comfort zone. Design thinking is a new way of working for most of the participants. They are not typically allowed to work in a familiar way during the program, and this can create a lot of anxiety. Beyond that, there is not a great deal of time for the participants to prepare for this as they begin using design thinking right away. Coaches utilize discomfort as a marker that indicates their participants are trying something new.

Example "Introduce them to a stranger and step away"

Title Constraints as Scaffolds

Key Terms Team roles, material limits, time limits

Explanation This happens when coaches assign certain boundaries or limits to the team or individuals to help them further engage in the design process. Working in a new way can be overwhelming. Limiting materials that the team uses for prototyping or assigning roles to different team members helps simplify the work they are doing. In extreme cases it also prevents participants from reverting back to traditional ways of working. If a participant starts overanalyzing ideas in a brainstorm, the coach can make that person facilitate the

session which does not give them time to overanalyze anything. All of this encourages concentration on the most essential parts of the design process.
Example “Give the team members clear roles [during testing]”

Title Safety
Key Terms Secure, permission to fail, trust
Explanation Creating a space were participants feel emotionally supported to take risks and try new things makes up this element. This is important because many participants are relying on their creativity more than they have in a very long time. It is a side of them that they may not be used to showing. If they feel unsafe, they are likely to revert back to traditional, and more innately safe ways of working. Part of this is making failure acceptable. Failure is an important part of design (Petroski 1992) though it likely is not an acceptable part of their normal working style.
Example “[Creating a] ‘yes and’ attitude is key”

Title Momentum
Key Terms Trust the process, move forward, lean into it
Explanation Momentum refers to the concept of keeping the process moving. It is important for a few reasons. Logistically, there is not much time in the program so the teams have to accomplish a lot in a short period and there is no way to catch up if they fall behind. In terms of working style, design moves quite quickly. They need to repress any instinct to stop and over think the situation. Finally, time constraints can actually increase participants’ creativity (Hawthorne et al 2014).
Example “[I] inject energy when momentum in low”

Title Engagement
Key Terms Excited, enjoy, passion
Explanation Keeping participants interested in the project and process are essential. Beyond providing motivation, an essential part of any learning experience, high engagement allows participants to personally connect with design thinking. This is important because often times learning design thinking is a personally transformative experience (Royalty et al 2012).
Example “[They feel] passionate about the work they did”

2.2.2 Participant Responses

Title	Optimism
Key Terms	optimistic, hopeful, belief
Explanation	This is faith in themselves and their creative abilities. It manifests itself in two main ways. First, that they believe they have the capacity to work using design thinking. Second, that this process can lead to a novel and interesting solution. It is important to have this sense of optimism because it is the first time many of them have used design thinking and they have no reference for what a successful process or outcome look like.
Example	"Trust in themselves as experts to identify needs"

Title	Confusion
Key Terms	Unsure, lost, unclear
Explanation	This is simply not understanding aspects of design thinking. It is a normal part of any learning process. It is up to the coaches to resolve their team's confusion.
Example	"Still not sure how [interviewing] relates [to design]"

Title	Surprise
Key Terms	Amazed, didn't expect this, wow this was unexpected
Explanation	Surprise in this case is being surprised at ones' own ability to succeed using design thinking. This, fittingly, was the most unexpected element. The previous two participant response elements are a normal part of learning. This encapsulates the experience participants have at the end of a project when they see that their creativity led to a novel and interesting outcome. To be clear, it is less about what they actually made and more about the creative capacity in themselves that came out during the process.
Example	"Whoa, I didn't think I could be out of the box!"

2.3 Discussion

Given these key instructional elements, the next step is to use them to construct a possible instructional model for design thinking. Looking at the coaches strategies elements reveals an interesting progression. Coaches challenge participants by

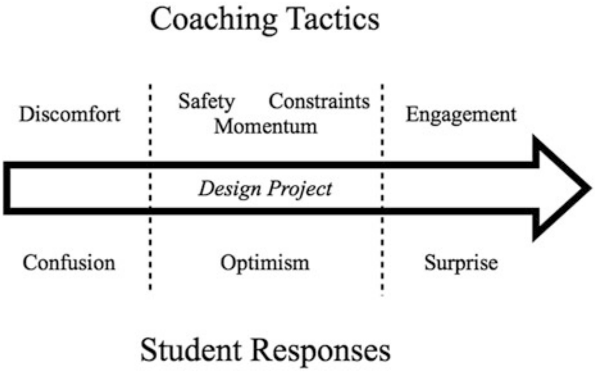


Fig. 1 A model of Executive Education design thinking instruction

making them uncomfortable. They create a safe space where it is all right to be uncomfortable, perhaps because everyone on a team is more or less equally uncomfortable. From there coaches help participants move forward in two ways. The first is by offering constraints as a scaffold to work in a design thinking manner. The second is to build and maintain momentum that ensures they stay on track. Finally the coach keeps the level of engagement high throughout the program as a way of connecting the participants personally to the design process.

Shifting to the participant response elements another progression emerges. There is a lot of initial confusion but the participants remain optimistic and see the process through. Ultimately, they are surprised with what they are able to accomplish.

Combining the two progressions forms a model illustrated in Fig. 1. The discomfort created from coaches immediately forcing participants to jump in to the new design thinking process leads to confusion. The safety, constraints, and momentum that carry teams through the process help instill a sense of optimism. In the end, being personally engaged opens the participants up to be surprised in their own creative abilities.

There are of course some major questions that this model brings up. One is how accurate are the participant responses? They come from the experience of seasoned d.school coaches, but ultimately they must be validated studying participants directly. Another is, are these elements necessary and sufficient to effective design thinking Executive Education programs? A way to test would be to remove one or more of these elements in a subsequent training and measure the effectiveness.

For the purposes of developing a tool that assess participants continued design thinking growth in organizations, the initial assumption is that the elements are all necessary. However, that may be adjusted as more is learned about the context to which participant return.

3 Study 2: Developing and Testing a Design Thinking Organizational Support Assessment Tool

How do we know if participants who attend training at the d.school receive continued support when they return to their companies? This study describes the development and implementation of an assessment tool in an organization that has sent multiple employees through d.school Executive Education bootcamps with the hope of developing design thinking as a core competency. The organization, which we will refer to as Bishop Industries was chosen, in part, because it is relatively large, over 30,000 people, and it has made a commitment to design. This means that employees who receive training at the d.school head back into a workplace with a fairly entrenched working culture. Another factor was that this is a service company that never utilized design. Most Executive Education participants work in organizations that have similar traits.

The assessment tool was loosely based on the journey map in study one. It sought to capture what employees learned through design thinking and how it felt to go through this process. The researchers worked with a Design Catalyst at Bishop Industries on the wording of the tool to align it with the language the company used around design thinking. The Design Catalysts at Bishop Industries are responsible for driving a design thinking process in existing project teams. The Design Catalyst whom we will refer to as Ted helped identify a team to test the assessment with.

It is worth recalling that the ultimate point of the assessment tools is not to determine if the team is good or bad at using design thinking. The point is to discover if the team is supported in developing design thinking competencies. The assumption is that if the organization is being supportive, then the team's design thinking will improve. So in reality, the organization is being assessed as much as the team is.

3.1 Method

3.1.1 Participants

Ted is a Design Catalyst at Bishop Industries. He has been trained at the d.school and is tasked with spreading design thinking within the company. His main job is to work with existing teams to teach and drive a design thinking process that is used in tandem with existing work processes.

The project team was an interdisciplinary group working on a mobile application. The team was comprised of a Project Manager, a Designer, a Software Engineer, and a Business Lead. Only the Design Catalyst had previous design thinking training. However, the Designer had extensive design education.

3.1.2 Procedure

Two unstructured interviews were conducted with Ted in order to hone the assessment tool and choose which team to test it with. The tool was then given to the team at the beginning and end of a 5-week project. After the project was over, a final semi-structured interview with Ted was conducted to review the outcomes of the assessment.

3.1.3 Materials

The assessment tool is a single form with three sections. There is a paper version and a digital version compatible with most smartphones. They were filled out individually. Table 2 shows the types of questions asked.

It is important to note that like Study 1, employees were prompted to share cognitive and emotional responses to the process.

3.2 Results

There are two primary results from the assessment. The first comes from the process scale items. When asked at the end of the project if the team is using a strong design thinking process, the respondents with no design background each indicated that they did with (+1 average). This was an improvement from their responses at the beginning of the project (+.33 average). By contrast, the Designer and Design Catalysts both indicated that they believed the team was not using a strong design thinking process (−1 average). However, each respondent felt like using the process more (+1.2 average).

The second interesting result came from the question, "What three words describe how you felt while [going through design thinking]?" The responses where coded using the same eight codes generated in study one. Only the participant response codes were applicable because Ted was the only person in an instructional capacity and none of his responses fell into the coaching strategy codes. All five respondents reported feelings that were coded as optimistic. Four of the five reported feelings that were coded as confusion. However, no one reported any feelings that could be coded as surprise.

The interview with Ted following the test revealed some additional information about the team. Ted moved onto help other groups after the 5-week project ended. He checked in with the team members individually a few weeks later. After the project ended the team did not feel like they could use design thinking in subsequent projects.

Table 2 DT organizational support assessment items

Item	Type	Example
Events: Important project events (four questions)	Short answer	What did you do or make?
Learning: What the employee learned from the event (two questions)	Short answer	How did that learning affect your next steps
Process: Feelings the employee has about using design thinking (four items)	Likert Scale from −2 to +2	I feel like using this process less/more

3.3 Discussion

Judging from the opinion of the design experts, the team was not applying design thinking very well, despite the other members believing that it was. Furthermore, the team was not able to apply it after Ted left. That said, the team was excited about the prospect of using design thinking more.

One possible explanation for these results is that the employees experienced a "healthy" mix of confusion and optimism but never managed to surprise themselves. The high optimism could account for why non-experts felt the team was successfully using a design process. It could also explain why everyone wanted to use design thinking more. The lack of surprise might be the reason why they were not able to apply this process on their own. Perhaps people need to be amazed by their own creative capacity before they are confident enough to apply design thinking on their own. This would suggest that Bishop Industries could benefit from reflections or some other techniques that allow employees to surprise themselves.

4 Conclusion

The first test of the assessment tool in an organizational context resulted in some interesting outcomes. Specifically a few poignant successes and limitations arose. From these outcomes it is possible to delineate guidelines for future iterations of this tool.

The tool was successful in capturing emotional responses in addition to cognitive responses. Asking participants to share how using design thinking made them feel elicited highly affective terms like anxious, excited, and disenchanted. This is important because emotion is an important aspect of learning this process and something that needs to be recorded. Also, by assessing each individual in the group, it was possible to identify contradictions, which can be a valuable way to understand the inner workings of a team (Goldman et al 2014).

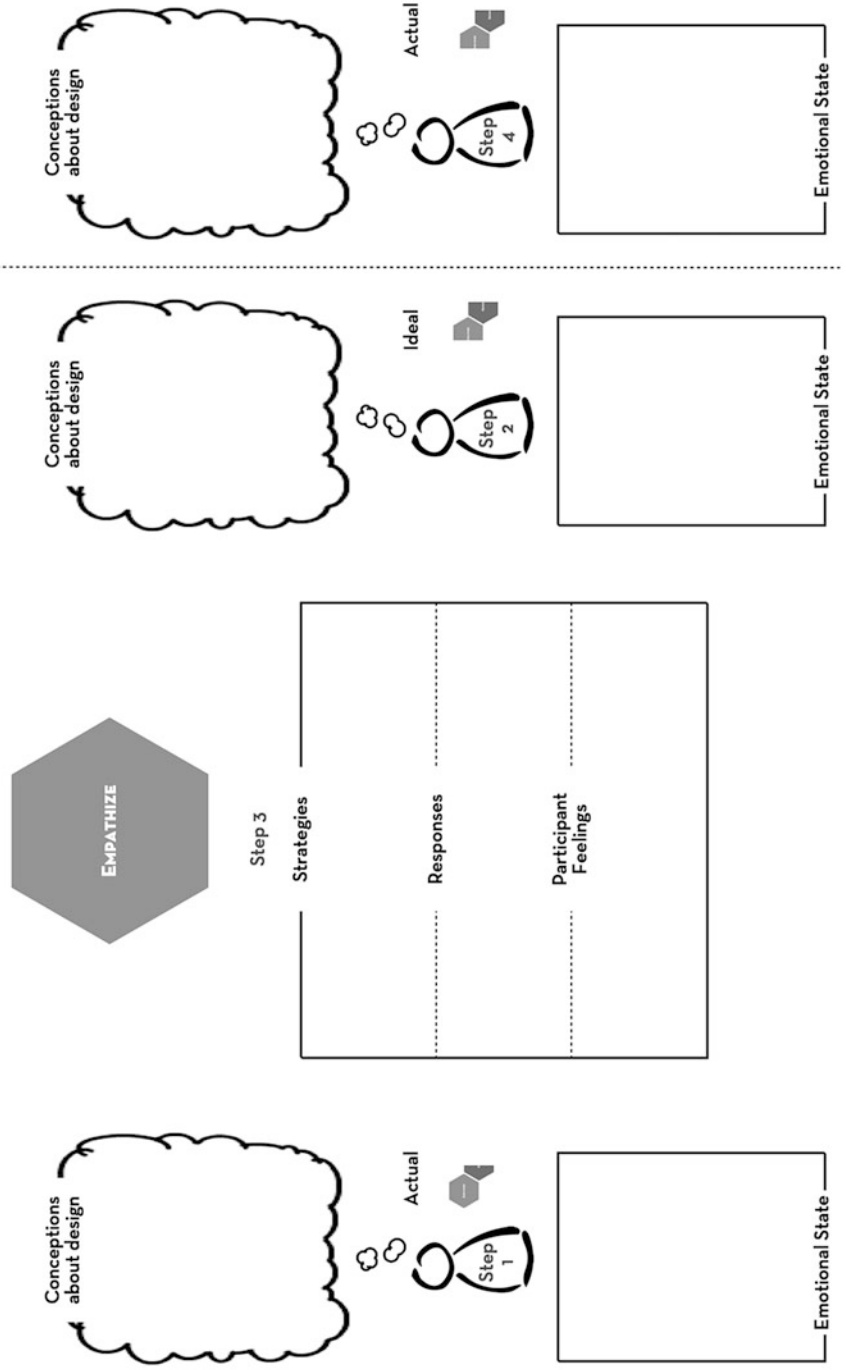
There were, however, some limitations. One is that this tool is primarily self-report. An early prototype asked employees to take a picture of something they made as part of the process. This was meant to show the tangible output of the process. Bishop Industries confidentiality policy made taking photos virtually impossible. Another limitation is that this tool focused only on what people created.

It was unable to capture more subtle environmental factors like how the company's incentive structure helps or hinders design thinking.

Given these outcomes, there are a few characteristics that should be incorporated in further iterations. One is that the tool must be fast and useful for employees. Most people do not have time to fill out a form unless they are forced to. A better strategy is to have the assessment serve an additional purpose. One example is an assessment tool that captures a team's workflow such that they can recount their working process when they present their project to corporate leaders. This makes it useful for researchers as well at participants. A second characteristic is that the tool must be calibrated to the organization based on the company's intended application of design thinking. Although most organizations are excited about design as a core competency, some do so in order to become more empathetic with customers while others are more focused on increasing the level of experimentation their employees engage in, for example. Knowing the intended application can direct an assessment tool towards specific behaviors the company wants to encourage. Finally, it is useful to have a design expert connected to the team that can serve to triangulate the data. The interview with Ted following the testing period put many of the assessment responses in context. All three of these characteristics show how important the organization itself is in the design of the assessment tool.

This study has uncovered factors that exist in successful design thinking professional development. Additionally, it has presented an example of how to assess the organization's support structure. Hopefully this will serve to inform similar work by encouraging researchers and organizations to co-design assessment tools that measure how well employees bring design thinking from the classroom to the workplace.

Appendix: Journey Map



References

Courage C (2013) Reweaving corporate DNA: building a culture of design thinking at Citrix. Resource document. Management Innovation eXchange. http://www.managementexchange.com/story/reweaving-corporate-dna-building-culture-design-thinking-citrix. Accessed 12 Dec 2013

Cross N (2007) Designerly ways of knowing. Birkhauser Verlag AG, Boston, MA

Goldman S, Kabayadondo Z, Royalty A, Carroll MP, Roth B (2014) Student teams in search of design thinking. In: Leifer L, Plattner H, Meinel C (eds) Design thinking research. Springer, Heidelberg, pp 11–34

Hackman JR, Wageman R (2005) A theory of team coaching. Acad Manag Rev 30(2):269–287

Hawthorne G, Quintin EM, Saggar M, Bott N, Keinitz E, Liu N et al (2014) Impact and sustainability of creative capacity building: the cognitive, behavioral, and neural correlates of increasing creative capacity. In: Leifer L, Plattner H, Meinel C (eds) Design thinking research. Springer, Heidelberg, pp 65–77

Korn M, Silverman R (2012) Forget B-School, D-School is hot. Resource document. http://online.wsj.com/news/articles/SB10001424052702303506404577446832178537716. Accessed 12 Dec 2013

Petroski H (1992) To engineer is human: the role of failure in successful design. Vintage books, New York

Royalty A, Oishi L, Roth B (2012) “I use it every day”: pathways to adaptive innovation after graduate study in design thinking. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Heidelberg, pp 95–105

Royalty A, Oishi LN, Roth B (2014) Acting with creative confidence: developing a creative agency assessment tool. In: Leifer L, Plattner H, Meinel C (eds) Design thinking research. Springer, Heidelberg, pp 79–96

TeamSense: Prototyping Modular Electronics Sensor Systems for Team Biometrics

Joel Sadler and Larry Leifer

Abstract Electronic sensors systems can be used to unobtrusively gather real-time measurements of human interaction and biometrics. However, developing custom sensor systems can be costly, time intensive and often requires high technical expertise in embedded mechatronic systems. We present a prototyping case study of a real world system, TeamSense, with the scenario of a manager who wishes to use embedded sensors to develop data-driven insights on team performance. Team Biometrics is a term used here to refer to a sensor system that measures some physical characteristic of a group of individuals. We explore how existing novice electronics toolkits, such as Arduino, can be used to develop a custom wireless biometric sensing network, without requiring deep technical experience, time investment, or cost. A series of functional data collection prototypes are presented, and we present lessons learned from initial testing with live deployment in a team setting. The need for more (1) modular and (2) mutable electronics and software components were discovered to be a limiting factor in allowing more experimentation in the early stages of sensor system prototyping. Modularity enables fixed functional blocks to be swapped in and out of a system (enabling combinations), and mutability allows modification of blocks to change their function (enabling mutation). We propose a future sensor platform that explores how modularity and mutability affects electronics prototyping with sensors. This work has broad implications for Designing Thinking, and importance of toolkits in reducing the barriers to entry for rapid prototyping with sensors.

J. Sadler (✉) • L. Leifer
Stanford University Center for Design Research, 424 Panama Mall, Stanford, CA 94305, USA
e-mail: jsadler@stanford.edu; leifer@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_7, © Springer International Publishing Switzerland 2015

1 Introduction

Electronic sensors allow the inquisitive to probe the world and take snapshots of the physical phenomenon that surround us. As Design Thinkers, the act of physically *observing* the human environment, and rapidly *prototyping* solutions, are core activities in developing *understanding* of a design problem. As sensors and electronics become more accessible to the technically inexperienced, there is an increasing opportunity to *enhance the designer's toolkit with technology*. However, incorporating sensors into a design process can be challenging, time-consuming, costly, and often requires some existing technical experience (Hartmann et al. 2006). How might we reduce the barriers to creative prototyping with sensors? What kinds of new questions can designers ask with sensor-enabled systems? Can we develop a better understanding of human systems through real-time instrumentation of body and space?

In order to motivate this work with a concrete example we explore a specific scenario, *TeamSense*, in which we attempt to measure human activity within a team. Here we present a prototyping case study through the hypothetical lens of a manager, who wishes to use embedded sensors to develop data-driven insights on team performance. We explore how existing novice electronics toolkits, such as Arduino (Mellis et al. 2007), can be used to develop custom wireless biometric sensing networks, without requiring deep technical experience, time investment, or cost. Here we consider TeamSense as a general example of biometrics, through the measurement of the unique physical characteristics of a team. Finally, we present a series of functional prototypes, and discuss the lessons learned from initial testing with live deployment in a team setting.

2 The TeamSense Scenario

2.1 Motivation: A Manager Monitoring Team Dynamics

The measurement of human activity within a team serves as a useful prototyping exercise with (1) a real world problem and (2) the potential use of electronic sensors to measure human activity. Here we imagine that a scenario of a manager wishes to better understand the dynamics of a team. As a broader motivation to this problem, consider that an increasing amount of work takes place in a team context, and there is a existing need to understand what factors can effectively diagnose, facilitate and predict team performance (Skogstad et al. 2009). Prior research has shown that certain team dynamics indicators are strong correlates with long-term innovative performance (Jung et al. 2012; Kress and Schar 2011). However, current team observation techniques are generally time-intensive (e.g. direct observation in the field), substantially asynchronous (e.g. offline video analysis) or otherwise

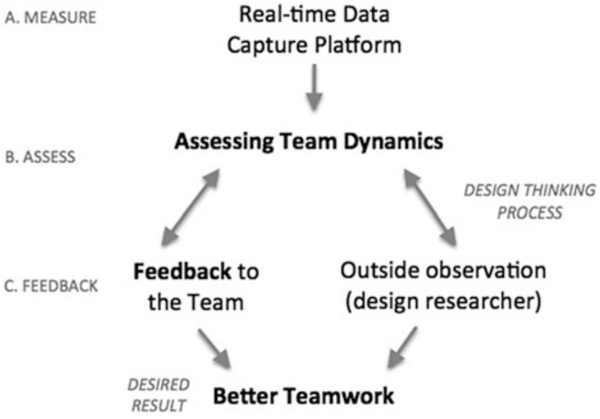


Fig. 1 TeamSense model

obtrusive to team function (e.g. bringing teams into the lab setting) (Kress et al. 2012; Tang and Leifer 1991).

Managers and teachers in project-based courses rarely have the time or opportunity to observe teamwork in progress, and so may miss critical dynamics cues. Additionally, they may simply be unaware of what those cues are or how to address them. Real-time, in situ team dynamics monitoring could have substantial benefit to team performance and learning, both through direct feedback to the team and through mediated feedback. To answer these questions in this scenario we propose the ***TeamSense*** System—a modular platform for precision real-time data capture, analysis and feedback. We envision that a combination of (1) unobtrusive sensing hardware in the collaboration environment, (2) software analysis tools for detecting patterns of team activity and, (iii) dynamic feedback mechanisms for behavioral intervention, may give team members and managers more actionable insight on how to improve the team performance (Fig. 1).

2.2 Desirable Features of Sensor System for Team Measurement

In order to achieve the goal of creating a system capable of capturing insightful data on team activity we consider the following requirements:

1. **Unobtrusive**: The system should not significantly alter or interfere with the activities of the team. The sensors need to be placed in a way that is minimally unobtrusive to the individual or groups' normal activity. Ideally the presence of the system provides quiet "ambient" capture of data that merges into the background, out of any noticeable attention.
2. **Continuous Logging**: The logged sensor data has to occur over a meaningful continuous time frame with minimal interruption to the data stream. Ideally the

system can operate without physical intervention over a number of weeks, without requiring for example frequent recharging of a battery, or manual intervention. We consider 6 weeks to be a reasonable time over which a team can complete a sub-task.

3. **Modular (Reconfigurable)**: As the exact types of sensors that will be effective are initially unknown, it is necessary to have an architecture that allows many different types of sensors to be plugged in, interchanged, and tested. A modular system that allows low-effort reconfiguration of sensors allows a greater variety of sensors to be experimented with in a given time. Using pre-exiting sensor modules has the advantage of increased system robustness and leverages known working components that are physically consistent,
4. **Mutable (Modifiable Functionality)**: By using only off-the-shelf sensors, reliability and consistency may be high, but modules may not be easily modified in function beyond a fixed design. In the case of creating a novel sensing technique, optimizing the system, or modifying the way that data is transformed, having the ability to modify the function of the system is highly desirable. The ability to change, or mutate, the function of the system is especially important for this initial prototyping stage—when the design is still in a state of flux.

3 Challenges to Rapid Prototyping with Sensors

In the *TeamSense* senario we assume that there is a *technically inexperienced user* that wishes to create an electronic sensor platform for continuous data logging of team's physical activity. We are particularly interested in exploring the current barriers that this user would encounter from an initial idea to a working sensor logging prototype. Creating a TeamSense system, as described in the previous section, requires integration of many diverse parts including mechanical, electrical and software components. For the technically inexperienced, creating such a sensor system can be time intensive, costly and require a significant amount of prior knowledge. Over time, electronics toolkits have emerged to address the need for low barrier to entry electronics creation. In the context of TeamSense, what are the key considerations and continuing challenges in using such toolkits for sensor prototyping?

3.1 Biometric Sensing: Instrumenting the Body vs. Instrumenting Space

In biometrics we wish to identify measurable physical characteristics of an individual or groups of individuals. Physical phenomenon includes measurable *internal* characteristics, e.g. heart rate, respiration, galvanic skin response, and *external* phenomenon such as movement, acoustic output and proximity. Measuring internal

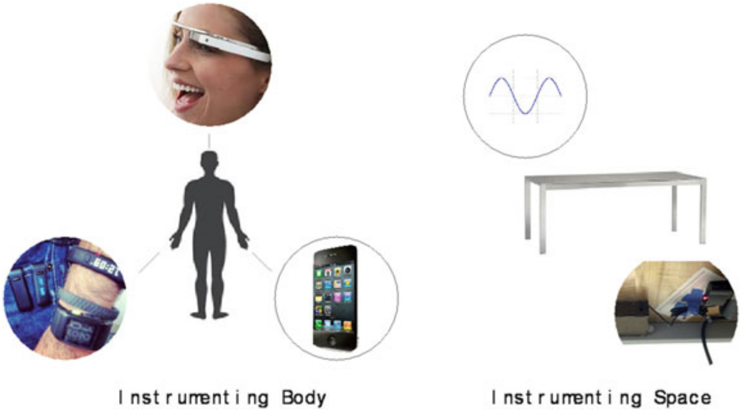


Fig. 2 Two alternative instrumentation approaches: body and physical space

vs. external biometrics leads to two distinct instrumentation approaches, one in which (1) the body is instrumented with sensors and (2) another in which we focus on instrument the physical space. Creating on-body sensing systems has the advantage of persistent data capture that follows the individual as they move through different spaces, but is potentially more obtrusive and is highly sensitive to form-factor. Instrumenting the physical space represents a reasonable starting point for unobtrusive ambient sensing, free from form-factor constraints, at the cost of a more limited window of data. With TeamSense we focus our initial explorations on the physical space (Fig. 2).

To illustrate examples here we can see a variety of possible sensors that may be used to instrument the physical characteristics of teammates (Fig. 3, Table 1).

3.2 Common Barriers Novices Face while Prototyping with Sensors

Electronic sensors transform physical phenomenon (e.g. pressure, movement, light, etc.) into electrical signals. The raw electrical signals can then be read by interfacing device capable of detecting these electrical changes in a self-contained embedded system, such as a programmable microcontroller. Working with these embedded systems can be especially challenging due to the following factors:

1. **Knowledge gap in both hardware and software**: Interfacing with microcontrollers often requires both software and electrical hardware familiarity. The user may have to author and debug code to specific the behavior of the system, as well as creating electrical circuits. Specifically in working with sensors, often supporting circuitry may need to be included in order to condition the signal, such as amplification and noise filtering.

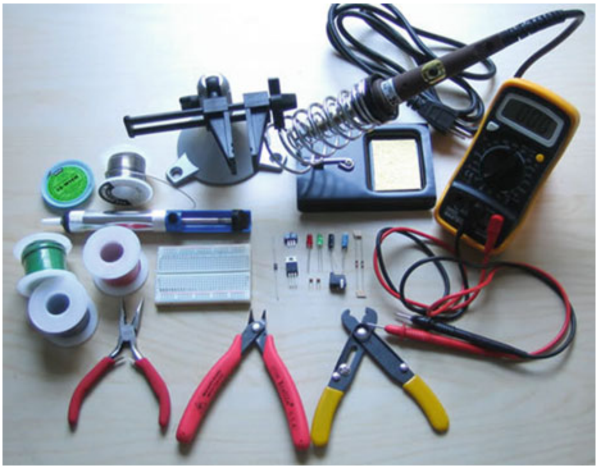


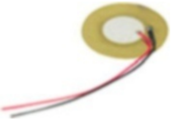
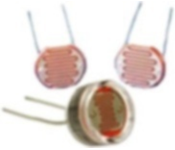
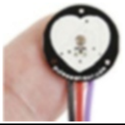
Fig. 3 Typical tools for prototyping with sensors, including soldering

2. **Constraints of tiny computers**: Microcontrollers are resource constrained—in that they are essentially small computers with limited amounts of memory (on the order of kilobytes) and limited computational power. Programming under such constraints typically requires the use of a lower level language such as C or C++, and more attention needs to be paid to the limits of the system, such as memory management. Higher level programming languages, such as Python, may be more novice friendly, but may also come at an overhead cost to the system.
3. **Special tools requirements**: In working with electrical hardware, physical, digital and electrical tools may be necessary to create even a most basic sensor system. For example the use of a soldering iron, breadboards, and wiring cutting are often necessary. Without the necessary tools, progress may be stalled or halted.
4. **Component availability**: Without physically having the components in hand there is often some time lag in identifying and acquiring the necessary parts. The lack of having "parts on hand" is more likely for a less technically experienced user. Delays of hours, or days can be significant if components need to be shipped to a location. In the context of rough and *rapid prototyping*, where a prototyping session might be on the order of an hour, these delays are significant factor.

3.3 Prototyping Toolkits for Electronics and Sensors

To address some of the challenges to prototyping with electronics, novice toolkits, such as Arduino (Mellis et al. 2007), attempt to provide programmable microcontroller platforms that are more accessible to technically inexperienced users (e.g. children, artists, musicians and tinkerers). The use of such toolkits have

Table 1 Sample biometrics and sensors for measuring team activity

Sample metric	Sensor	Physical example
Physical movement	3-axis accelerometer (ADXL345)	
Vibration	Piezoelectric transducer (VELLEMAN TV4)	
Motion	PIR motion sensor (VUPN5943)	
Proximity	Infrared distance sensor (GP2Y0A21YK0F)	
Ambient light	LDR light dependent resistor (TrueOpto 58-0128)	
Acoustic output	MEMs microphone (ADMP401)	
Pulse	Heart rate monitor (SEN-11574)	

grown in popularity since their introduction, with recent estimates of over one million Arduino boards sold to date, and tens of thousands of registered to their online community (Arduino.cc 2013). Toolkits like Arduino provide an integrated experience to creating with electronics, in that they provide *both hardware and a software development environment* (IDE) tailored to the hardware. Arduino based systems are convenient for interfacing with sensors with since they abstract away much of the lower level electrical details, but still provide room for mutability (modification) and manual building of circuits. Stackable add-on boards, referred to as “shields”, allow of more specialized functions to be added modularly, such as SD card logging or wireless capability (Fig. 4).

For a detailed review of the history and design of microcontroller-based toolkits, the authors recommend Blikstein’s 2013 review of microcontrollers and their use with technical novices in education (Blikstein 2013).

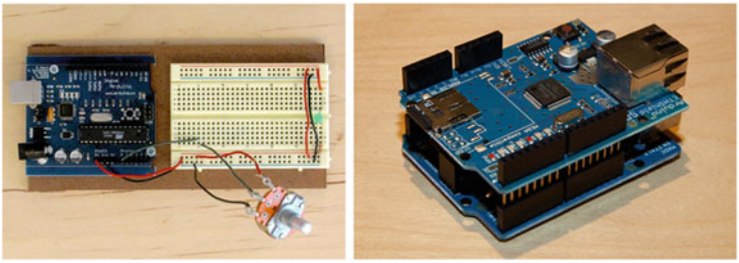


Fig. 4 An Arduino board with a breadboard. Stackable shields (*right*)

3.4 Modular vs. DIY Sensor Toolkits

Toolkits like Arduino fall in the category of DIY "*do it yourself*" systems, where flexibility of modification (mutability) is highly valued. Alternative electronics toolkits such as Phidgets (Greenberg and Fitchett 2001), and d.Tools (Hartmann et al. (2006)) take a different approach to interfacing with sensors, and strive for a higher level of *modularity* in their components. In contrast with Arduino, no circuitry is required with these systems, and interfacing with a sensor is achieved by connecting a plug-and-play "smart module" over a standardized connection. These systems automatically recognize when a particular component is plugged-in, identify the type of component, and react accordingly. The main advantage of using these modular systems is in the convenience of a true plug-and-play interface. However, this added convenience comes at the cost of:

1. **Added component cost and complexity**: Smart modules often add an additional lightweight micro-controller to uniquely identify a component such as a "motion sensor" and communicate its data over an electrical bus.
2. **Reduced mutability** to modify the system: modules tend to be less flexible in their affordance for modification. Within a platform choices may be limited to modules specifically designed by the third party (The Arduino DIY style overcomes these limitations by exposing the raw circuitry to users) (Fig. 5).

In the context of TeamSense, an idealized sensor system would combine the mutability benefits of DIY style toolkits like Arduino, with the usability benefits of plug-and-play modules.

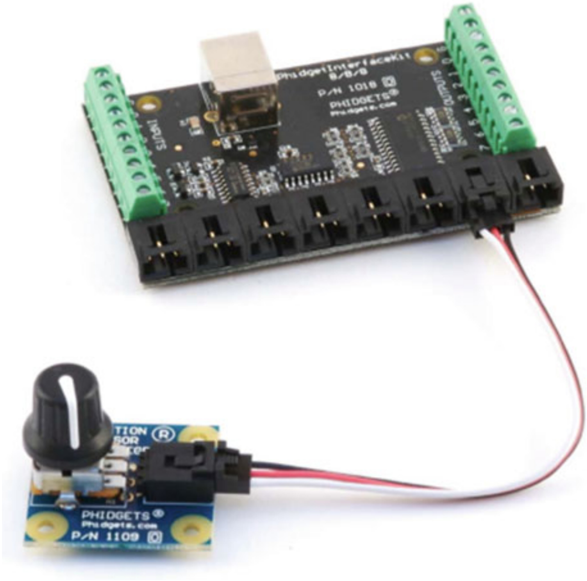


Fig. 5 The Phidgets toolkit encourages modular plug-and-play components

4 TeamSense System Prototypes

4.1 Acoustic Vibration Monitoring Prototype

In order to demonstrate the feasibility of DIY sensor logging system—we constrained ourselves to use only tools typical of technically inexperienced users (within the Arduino ecosystem). We selected acoustic vibration as an initial proof-of-concept metric to instrument a physical tem space. We constructed and tested two sample prototype TeamSense units at Stanford with the aim of capturing workspace physical activity through this vibration metric. Arduino-based programmable microcontroller boards and a Xbee wireless shield provide the base of a minimal sensor logging unit capable of reading attached analog sensors and transmitting data to a near by computer. Each unit is designed to be mounted to the underside of a team's table in the shared workspace. Physical activity is registered in real time by means of a piezoelectric contact microphone, and transmits this information wirelessly to a base station PC along with the Team ID. Using a Xbee wireless network, multiple sensors can be deployed in a workspace and all transmit data to a single point that can be accessed remotely. This allows for real-time monitoring of one channel of team activity as well as data-logging for asynchronous team observation. By attaching various analog sensors, this setup allows modular expansion different types of sensing techniques (Figs. 6 and 7).

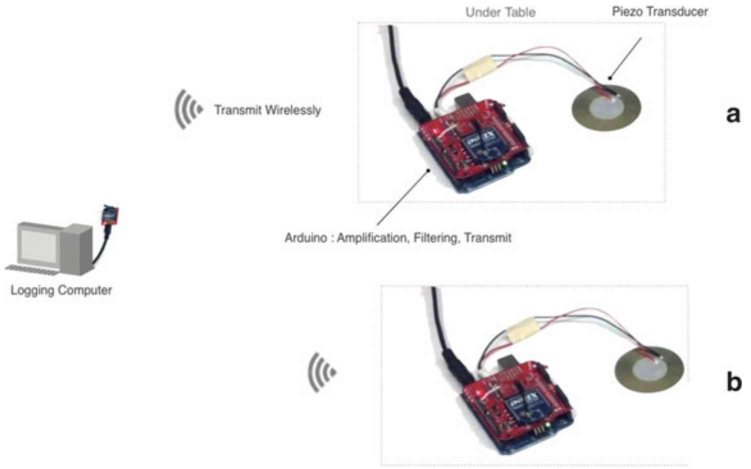


Fig. 6 TeamSense acoustic vibration overview

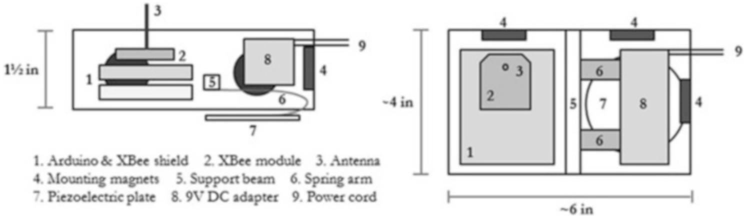

1½ in
1
2
3
4
5
6
7
8
9
~4 in
~6 in
1. Arduino & XBee shield 2. XBee module 3. Antenna
4. Mounting magnets 5. Support beam 6. Spring arm
7. Piezoelectric plate 8. 9V DC adapter 9. Power cord

Fig. 7 Schematic diagram of TeamSense unit

4.2 Pilot Deployment Results

The two prototype units were tested with ME310 engineering design teams at Stanford over a course of 6 weeks. Units were mounted to the underside of each team's table, and team gave consent to have their activity logged as a part of the pilot. The figure below shows the physical setup of the pilot units. Each unit required a connection to a power outlet and was designed so that data capture resumed automatically after an accidental power-outage (with an indication in the log that this has occurred). A Java based client app (written in Processing) running on a nearby computer, received data events and logged them to text file, with unique time stamps and team IDs (Fig. 8).

In reviewing the sample data we found that data transmission and logging methods were quite robust; we recorded a 6-week-longitudinal data stream from both sensors without serious interruption or the need to intervene. Teams indicated that the device was unobtrusive to their work and that they quickly forgot about it as it was "out of sight, out of mind." A sample of acoustic data between two teams is shown below. The data shows distinct qualitative differences in the pattern of work

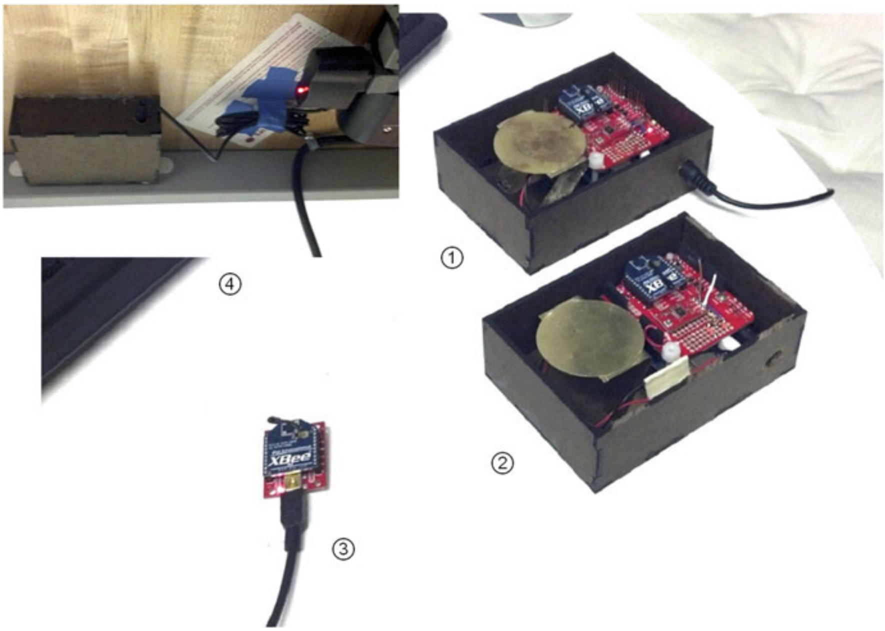


Fig. 8 Mountable sensing units (1 and 2) Communicate wirelessly to a Xbee receiver (3). (4) Shows mounting to the underside of a table

between the two teams. Both the absolute event frequency (number of acoustic events) and relative timing of activities varied. The data clearly shows one team as constantly more "physically active" in their space (Fig. 9).

The pilot serves as a proof of concept of a DIY sensor logging system, leading to qualitatively comparable data in team activity. However discovered several ongoing challenges:

1. **Limited Channel**: Drawing robust conclusions about team performance requires more than a single channel of sensor data. Acoustic events serves as a interesting proof of concept for a logging system, but a more detailed picture of team activity would be needed in order to draw detailed conclusions about the link between sensor data and team dynamics.
2. **Minimizing Variations between Sensor Systems**: With this DIY method, the manual creation of supporting circuitry to interface with sensors can result in misleading variations in sensor data. The electrical and software calibration of the system has a significant effect on the sensor signal variation. Variations in wiring, components, and fabrication appear to significantly change the quality of the sensor output. For example, during prototyping a poorly soldered wire was initially found to be source of drop in data. Finding more ways to increase the consistency, and robustness between units is essential to draw more actionable conclusions from the data.

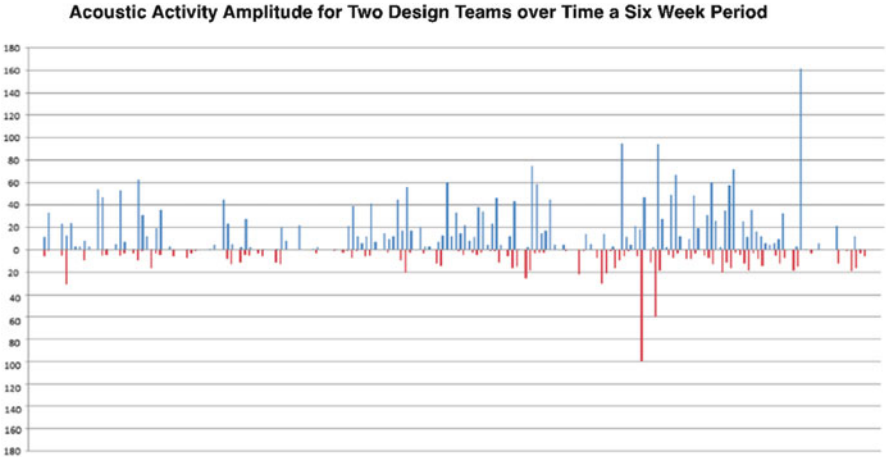


Fig. 9 Comparing acoustic events collected over a 6 week period

3. **The Need for Increased Modularity**: As mentioned above, it is especially important to minimize the variations between sensor systems. By using off-the-self sensor modules, with known working configurations some of the variability can be reduced. Using modules minimizes the supporting circuitry that needs to be created by the designer, and the chance of introducing errors and variations in fabrication is reduced. By using an approach similar to other plug-and-play module systems, we can promote a greater ability to swap different types of sensors in and out with low effort. Ideally such systems maintain the benefit of being directly mutable if modifications are needed.

4.3 Method for a More Modular TeamSense System

The pilot uncovered a number of challenges in prototyping sensor systems for team measurement. One alternative approach to a DIY sensor system is to create a more modular system of plug-and-play components that reduces the errors in manual circuit creation, and promotes experimentation of different sensors. The research system d.Tools (Hartmann et al. (2006)) and commercially available systems like Phidgets (Greenberg and Fitchett 2001) are good examples of this approach. He we attempt to create such a system tailored to the TeamSense data logging challenge (Fig. 10).

The figure above shows an iteration on the TeamSense units that uses:

1. **A smart module converter**—that connects an off-the-shelf analog sensor to a small additional microcontroller (ATTiny45). This additional microcontroller add "smart" identification of what sensor is plugged in, over a common electrical bus (I2C protocol). This module uses an approach similar to Hartmann

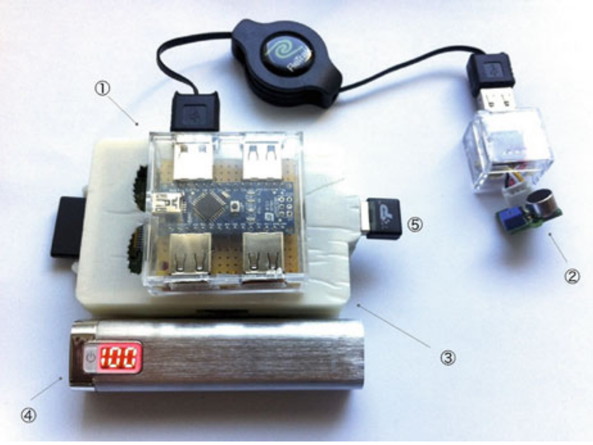


Fig. 10 Arduino (1) with plug-and-play smart sensors (2). A Raspberry Pi (3) serves as a logging computer with battery (4) and wireless (5)

et al. (2006), however the programming of the module is achieved using only novice accessible Arduino tools. This way the components achieve both high mutability and modifiability.

2. **A lighter weight computer**: A microcomputer, such as a Raspberry Pi, can replace the computer shown in the previous setup. For data logging the use of a Raspberry Pi ($35), or similar platform, represents a cost effective way to continuously log and store data in a TeamSense setup.

Using the method describe above we may create a team sensing system that combines the benefits easily swappable modular components, with highly modifiable (mutable) customization. We expect that future platforms can explore the tradeoffs between DIY and Modular sensing methods discussed here for measuring team activity.

5 Conclusions

In this article we describe a prototyping case study of TeamSense, a modular electronics platform for measuring team activity with electronic sensors. By focusing on creating a DIY system that a technically inexperienced user could create, we uncover some of the advantages and limitations of using existing electronics toolkits for sensor prototyping. Through prototyping and testing of two different systems we successfully demonstrate the feasibility of creating a DIY sensor data-logging platform for instrumenting a team space. Sample data of acoustic vibration events with live teams, highlighted the need for more robust methods of data capture and system fabrication.

Through this exercise we uncover the inherit limitations in creating electronic sensor systems—and future works points to the need for more modular plug-and-play systems to increase system robustness, and to promote rapid experimentation with additional sensors. We propose a future sensor platform that explores how modularity and mutability affects electronics prototyping with sensors. This work has broad implications for Designing Thinking, and importance of toolkits in reducing the barriers to entry for rapid prototyping with sensors.

References

Blikstein P (2013) Gears of our childhood: constructionist toolkits, robotics, and physical computing, past and future. In: Proceedings of the 12th international conference on interaction design and children (IDC '13). ACM, New York, NY, pp 173–182

Greenberg S, Fitchett C (2001) Phidgets: easy development of physical interfaces through physical widgets. In: Proceedings of the 14th annual ACM symposium on user interface software and technology (UIST '01), ACM, New York, NY, pp 209–218

Hartmann B, Klemmer SR, Bernstein M, Abdulla L, Burr B, Robinson-Mosher A, Gee J (2006) Reflective physical prototyping through integrated design, test, and analysis. In: Proceedings of UIST 2006, October

Jung M, Chong J, Leifer L (2012) Group hedonic balance and pair programming performance: affective interaction dynamics as indicators of performance. In: CHI '12 proceedings of the 2012 ACM annual conference on human factors in computing systems, ACM, New York

Kress G, Schar M (2011) Initial conditions: the structure and composition of effective design teams. In: Proceedings of the international conference on engineering design (ICED), Copenhagen, Denmark

Kress G, Schar M, Steinert M (2012) A standardized measurement tool for evaluating and comparing team reframing capabilities. In: Proceedings of the international design conference (DESIGN). Dubrovnik, Croatia

Mellis DA, Banzi M, Cuartielles D, Igoe T (2007) Arduino: an open electronics prototyping platform. In: Proceedings of the conference on human factors in computing (alt.chi) (CHI'07), ACM, New York

Romano Z (2013) Designboom Visits Officene Arduino in Torino. Retrieved 1 Dec 2013. http://blog.arduino.cc/2013/09/06/designboom-visits-officine-arduino-in-torino/

Skogstad P, Steinert M, Gumerlock K, Leifer L (2009) Why a universal design project outcome performance measurement metric is needed—a discussion based on empirical research. In: Norell Bergendahl M, Grimheden M, Leifer L, Skogstad P, Seering W (eds) Proceedings of ICED '09, vol 6. Design methods and tools, The Design Society, USA, pp 473–484

Tang JC, Leifer LJ (1991) An observational methodology for studying group design activity. Res Eng Des 2(4):209–219. doi:10.1007/BF01579218

Tele-Board MED: Supporting Twenty-First Century Medicine for Mutual Benefit

Julia von Thienen, Anja Perlich, and Christoph Meinel

Abstract Tele-Board MED is a medical documentation system designed to support patient-doctor cooperation at eye level. In particular, it tackles the challenge of turning medical documentation from a necessity, which disturbs the treatment flow, into a curative process by itself. With its focus on cooperative documentation, Tele-Board MED embraces a call uttered by many scientists and politicians nowadays for twenty-first century medicine and patient empowerment. At the same time, the project is deeply rooted in the culture of design thinking. Accordingly, the benefit for patients should not be at the expense of doctors. Rather, the needs of all stakeholders shall be discerned and served. Behaviour psychotherapy has been chosen as a first field of application for Tele-Board MED. Using quantitative and qualitative methods, an initial feedback study was launched with 34 behaviour psychotherapists. It showed that many therapists are skeptical towards digital documentation and record transparency in general. Nonetheless, Tele-Board MED is considered helpful and promising. In particular, therapists estimate to save one third of their normal working time when assembling case reports with the system. The vast majority of therapists can well imagine using Tele-Board MED with patients. Apart from that, quantitative methodological strategies—though seldom used in the design thinking community—proved to be potent tools for carving out needs and insights that will inspire the next generation of Tele-Board MED.

J. von Thienen (✉)
Hasso Plattner Institute for Software Systems Engineering, Prof.-Dr.-Helmert-Street 2-3, 14482 Potsdam, Germany
e-mail: julia.vonthienen@hpi.uni-potsdam.de

A. Perlich • C. Meinel
Internet-Technologies and Systems, Hasso Plattner Institute for Software Systems Engineering, Prof.-Dr.-Helmert-Street 2-3, 14482 Potsdam, Germany
e-mail: anja.perlich@hpi.uni-potsdam.de; christoph.meinel@hpi.uni-potsdam.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_8, © Springer International Publishing Switzerland 2015

1 Starting Point: No Superhuman Doctors with Instant and Infallible Documentation

When you see a doctor, he will create a patient file regarding your case. It will include some initial data like age, preliminary medication or intolerances and all the data that accumulates in the process of your treatment.

Consider how many patients your doctor sees! Think of all the intricacies each case brings along. That is a lot of information your doctor needs to have in mind—or in his files.

Yet, to remember your case in all its details the doctor will hardly go through your record at length when seeing you anew. Typically, there are only a couple of seconds between patient appointments. Can the doctor truly grasp all the important issues of your case this quickly?

In addition, the patient file is typically visible for the doctor alone. Any gaps or mistakes in previous documentation are likely to remain in your file. Your chances of detecting an incorrect compilation of symptoms, for instance, are rather mediocre.

If your doctor was a perfect documenter, his files would be complete and infallible. Before starting your treatment, he would read your file carefully and consider every important detail. Yet, reality departs from this ideal sometimes a little and sometimes a lot. Mostly, it is the patient who pays the toll.

Yet, to be sure, the problem is not only that unreliable doctors create insufficient documentation. What might worry us even more is a trade-off: At some point, which comes rather sooner than later, it seems that doctors can provide only one, either good treatments or good documentation.

Imagine a diligent doctor who really wanted to be a perfect documenter. He might indeed come up with quite comprehensive files if he wrote detailed protocols of every session, making you wait whenever something potentially important has been said or done, or occurred. How often would that be, every 10 s? Your treatment would drag on and on. Soon you might feel the bureaucracy of documentation was more important than your concerns, which the doctor would have to postpone again and again. Every short look at you would be followed by long glances at the monitor where his elaborate documentation filled page after page.

2 Making Medical Documentation Helpful from the First Moment On

Typically, it is accepted that documentation disturbs the current treatment to some degree since it will hopefully help when meeting the next time. We envision a form of documentation that makes treatments better—from the first moment on. According to our perspective, the redesign of medical documentation should address three key issues. (1) Creating medical files: Here, it is crucial that the activity of documenting does not hinder healing. Ideally, the act of documenting would itself be curative. (2) Medical file quality: Records should be as correct and complete as possible. (3) Profiting from medical files: Finding relevant information in the records and working with it should be as easy and fruitful as possible.

To answer these challenges we created Tele-Board MED. The basic idea is to fade out documentation as an extra step apart from the treatment. Instead, whatever doctors and patients do anyway to advance a treatment shall be supported to become means of documentation as well.

A likely starting point is to help patient and doctor exchange information, which is an integral part of most treatments. E.g., the patient reports symptoms or complaints. The doctor generates diagnostic findings and explains them. He may name different treatment options. Jointly, doctor and patient consider the advantages and disadvantages of the options in deciding on one. . .

How can Tele-Board MED support this kind of information exchange? Figure 1 shows a first prototype of the user interface. Headlines like “diagnosis,” “therapy options,” “intolerances” or “present medication” provide orientation. The doctor can use such keywords to structure information in a flexible way.

Flexibility is crucial since patient-doctor conversations cannot be predicted. E.g., it may well happen that the patient does not recall all his intolerances at once when asked for them. He may recall an important intolerance only at the end of the session and it should then easily become part of the file. Tele-Board MED does not predefine question-and-answer sequences or anything concrete that must be entered in the system at some point. It solely highlights issues that typically figure in patient records and helps to cluster information.

The first Tele-Board MED prototype depicts all information on a large screen visible to both patient and doctor. In this way, the patient also has the chance to detect errors or notice missing pieces of information.

Pictures may be used to speed up understanding. The doctor does not have to read long complicated sentences in the patient record, searching for crucial bits of information, instead, he can catch important points in just seconds. Besides that, pictures may also help patients understand. After all, “doctor’s talk” can sometimes sound like gobbledygook.

Whether it is truly possible to attain medical documentation without spending extra time on the task by using Tele-Board MED does not, of course, only depend on the general idea but also on matters of implementation. According to the vision of Tele-Board MED, users should be able to enter and move around keywords or pictures so quickly and intuitively that it barely takes any time and almost goes unnoticed.

The concept of Tele-Board MED has been realized as a software system. It can be used on a broad range of devices, such as a desktop computer, laptop, digital whiteboard, tablet or even a mobile phone. The user can choose whatever works best for him.

From our point of view, a big touch-screen is ideal. This allows both doctor and patient to see the relevant information and point at issues of current concern. The

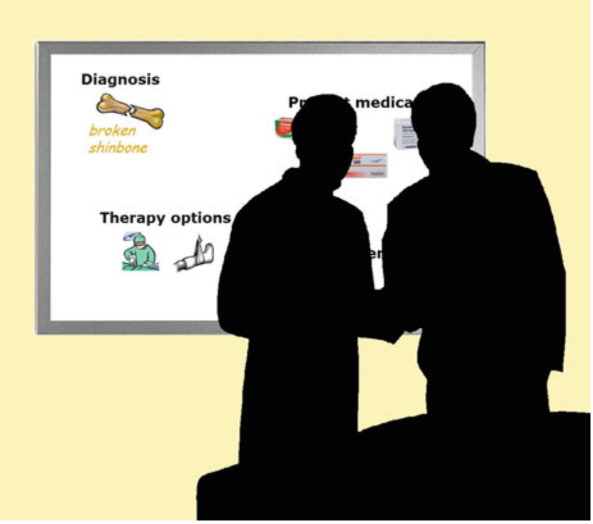


Fig. 1 The general idea of interacting with Tele-Board MED

touch-function makes it possible to sort crucial pieces of information with just the swipe of a finger. Whenever some new piece of information enters the discussion, such as a major symptom or some diagnostic finding, a keyword may be written down quickly on a keyboard or with a digital pen (where automatic character recognition is already available). The keyword appears on a sticky note at the screen. You can put your finger on it and move it naturally and quickly to the place of the file where it belongs. In addition, the patient may be encouraged to enter information too, e.g., by providing an extra keyboard for him.

3 Empowering Patients: Twenty-First Century Medicine

Quite obviously, Tele-Board MED is not a traditional documentation tool that supports a doctor in his classical role as an authoritative figure who pulls all the strings alone (and who carries all responsibility alone). While in the past doctors typically kept patient records to themselves, Tele-Board MED makes it easy to share a file with the person concerned—the patient. More generally, it is designed to help patient and doctor **cooperate for mutual benefit**.

The advantages of moving towards cooperation are quite comprehensible. Doctors may thus attain better knowledge bases. Patients can help ensure the correctness and completeness of files by detecting errors (such as a false list of current medication) or adding new data (e.g., an additional intolerance they are aware of). Patients, on the other hand, attain more control in a matter that is crucial for them: their own health. Indeed, Tele-Board MED might reach its ideal of **turning documentation into a curative process** if only the system could help patients acknowledge and take responsibility for their own well-being. After all, patients ought to choose their doctors wisely, decide on treatment options that work well for them and engage in personal health behaviour like sports or diet.

The insight that cooperation among responsible patients and approachable doctors promotes health much better than the traditional model of doctors as superordinate authorities has inspired a lot of work lately, both in science (e.g., Kalra 2011; Koch 2012; Koch and Vimarlund 2012; Perlich 2012) and in politics (e.g., Bahr 2013; Bundesgesetzblatt 2013). Common keywords are "patient empowerment" and "self-management" in English or "Patientenbeteiligung" and "mündiger Patient" in German.

For the first European Conference on Patient Empowerment, held in Denmark in 2012, the *European Network on Patient Empowerment* (ENOPE) compiled a series of case studies, which they introduced with some general remarks.

> Health systems have often been organized with the needs of the clinician and the system taking priority in the delivery of care to patients. In such a model the professional is at the center of the system – he or she has exclusive access to knowledge and the patient is expected to comply with the instructions given by health professionals.
>
> In many countries this is now changing: health care is considered a process of co-production in which professionals and patients jointly work on solving health problems [. . .].
>
> (ENOPE 2012, p. 7)

Far-reaching changes are called for. ENOPE (2012) invites both doctors and patients to "change a mindset which is based on hierarchical expectations towards one based on dialogue and co production [. . .]. Information needs to be much more easily available and understandable. Patients need to [. . .] ask questions, express needs and expectations and implement jointly agreed treatment programmes" (p. 7). Furthermore, ENOPE calls for a "transformation of the doctor/patient relationship away from a traditional paternalistic arrangement to one of partnership, where the patients experience and expertise is fully utilized" (p. 10).

In terms of scientific approaches, the *Chronic Disease Self-Management Program* (CDSMP) developed at the Stanford Patient Education Research Center has not only been quite successful (Lorig et al. 1999, 2001), but also quite influential. It seeks to support self-management through workshops and networking, connecting patients and health care professionals. The program inspired large-scale health programs in England,[1] Switzerland and the German-speaking countries[2] as well as in Denmark.[3]

Scientists also write about how to empower patients who suffer from particular illnesses. That such publications treat illnesses like diabetes (i.e., Funnell and Anderson 2004) may come as little surprise given that patient involvement has been common in this field for a long time. Yet, scientists also start calling for an empowerment of patients who suffer from illnesses like cancer (McCorkle

[1] http://www.expertpatients.co.uk

[2] http://www.evivo.ch/evivo-partnerschaften

[3] http://www.patientuddannelse.info/aod/om-projektet.aspx

et al. 2011), where in former times doctors were seen as the exclusive party to bring about health.

In 2004, the Swiss Academy of Medical Sciences (SAMS) and the Swiss Medical Association (FMH) published the results of a brain trust that was commissioned to clarify "the objectives and tasks of medicine at the beginning of the twenty-first century". Right at the beginning the panel states:

> *Empowerment*: More than ever, patients should be involved cooperatively [. . .] in decision-making processes. Access to high-quality information ("knowledge") regarding health and disease must be ensured. At the same time, the self-responsibility of citizens for their own health and treatment of diseases needs to play an increased role.
>
> (Swiss Academy of Medical Sciences et al. 2004, p. 12, our translation)

In Germany, a new law was even passed in 2013 that regulates medical documentation anew to promote **patient-doctor cooperation at eye level** and help patients **practice responsibility** (Bundesgesetzblatt 2013, part I, no. 9). Federal Health Minister Daniel Bahr (2013) says: "With the patients' rights law, we strengthen the rights of patients. Our model is the informed, responsible patient who can confront physicians at eye-level" (our translation).[4]

The law states clearly that doctors are obliged to document the whole process of treatment promptly and comprehensively (§ 630f). Apart from strictly regulated exceptions, there may be no treatment without explicit order by the patient. That is, doctors must communicate to the patient in an understandable manner the diagnosis and what treatment options there are (§ 630e). It is then the patient who decides how to continue (§ 630d). In addition, patients have the right to see their own patient record any time and in a complete form (§ 630g). Patients also have the right to receive an electronic copy of their file if they choose (§ 630g), so that they can take it home or to other doctors.

Tele-Board MED supports just this kind of twenty-first century medicine where doctors and patients are invited to cooperate at eye level for mutual benefit.

4 Starting with One Medical Domain: Behaviour Psychotherapy

When it comes to developing Tele-Board MED, it is clear that documentation needs may vary from one medical domain to the next. An orthopaedic specialist, who wants to convey information about illnesses in an easily understandable manner, might need graphics for different malpositions of the spine. A dentist, however, would be likely to use diverse icons for teeth. Presumably, Tele-Board MED will unfold its full potential only when tailored to suit the particular needs of each specific domain where it is used. Therefore, we have decided to adapt the system for one medical domain to first test the whole approach thoroughly there. The first

[4] http://www.bmg.bund.de/praevention/patientenrechte/patientenrechtegesetz.html

domain of usage we chose is behaviour psychotherapy. Why this field? On the one hand, purely paper-based documentation is still very common in behaviour psychotherapy despite all the disadvantages it entails. On the other hand, resistance towards fully transparent and digital patient files might also be strong. Therefore, behaviour psychotherapy seems to be a well-suited domain for studying Tele-Board MED both in its positive potential as well as in the doubts among clinicians, which the approach might cause.

4.1 Disadvantages of Paper-Based Documentation

Starting with the needs of clinicians, many psychotherapists still write major parts of the patient files by hand, instead of using digital tools. Writing by hand is considered a quick and easy form of documentation, which is also assumed to be least disruptive to the flow of conversation. Yet, at the same time it has many disadvantages.

Redundancy in Writing Clinicians need to write detailed case reports regularly (so that insurance companies pay for the treatment or to inform other doctors). Thus, the same information which a clinician already wrote down by hand needs to be sorted and (re-)typed into a computer. After all, it goes without saying that official documents are sent in a machine-written format today and not written by hand.

No Search Function When writing case reports the clinician needs to compile a lot of information. But where in handwritten notes can a specific piece of information be found? The lack of an automatic search function is particularly unfortunate for behaviour psychotherapists given that they typically accumulate huge piles of handwritten documentation.

Readability Issues Another problem is to integrate information provided by patients. In behaviour psychotherapy, patients are often asked to fill in several pages of anamnesis questionnaires to report their symptoms, clinical history etc. Once again, the answers are typically handwritten. Deciphering a patient's handwriting can be quite intricate for a therapist who tries to compile case reports.

Exchanging information the other way around is difficult too. Patients may want to read their files at some point (in Germany, as stated above, they even have a legal right to do so.) It may well be the case that a patient cannot decipher the therapist's handwriting. For this reason, again, electronic documentation would be advantageous.

Considering all these issues together, marked deficits in current documentation strategies suggest that new approaches—as with Tele-Board MED—may be beneficial for clinicians and patients.

4.2 Scepticism Towards Electronic Documentation

Even though new technologies promise great advantages, many clinicians still adopt them rather hesitantly. Among behaviour psychotherapists, scepticism may be particularly pronounced—for the following reasons.

Transparency Is Delicate Clinicians in general and behaviour psychotherapists in particular may be reluctant to share their files with patients. E.g., as a therapist, how will you document what you consider to be a delusion of the patient if he has insight to your notes all the time?

Technology Disturbs Another challenge is that handwritten documentation seems to have important advantages. In particular, many clinicians assume that handwriting is less disturbing for the treatment than typewriting. In psychotherapy, this issue is particularly important since a high quality patient-doctor-relation is considered one of the major curative factors (Grawe 2005). Therapists certainly do not want to introduce new technologies if these make therapeutic interactions more bureaucratic.

In summary, both the needs and the obstacles of modern documentation, which figure in many medical domains, are particularly pronounced in behaviour psychotherapy. On the one hand, digital documentation could be of great advantage. It eases the task of writing case reports and helps to make patient records accessible for the people concerned: patients. On the other hand, it is difficult to find formats of documentation that allow doctors to feel good about sharing their notes with patients. In addition, it is challenging to find a way of documenting which allows doctors to both document comprehensively and treat patients soundly at the same time.

5 New Functions in Tele-Board MED to Support Behaviour Psychotherapy

Preparing Tele-Board MED for use in behaviour psychotherapy is certainly an iterative process where you start at some point and—hopefully—make it increasingly valuable.

We started with Tele-Board, a digital whiteboard system that was developed at the *Hasso-Platter-Institute* at the *University of Potsdam* (Gumienny et al. 2011). Originally, it was built to help teams of inventors analyse problems and solve them creatively. The users (design thinkers) typically gather a large amount of information regarding some challenge they are working on. Practically any hardware of choice may then be used to enter the information into Tele-Board: a keyboard, tablet and electronic pen, mobile phone etc. The information appears on a digital whiteboard where the users can work with it (see Fig. 2).

Fig. 2 Tele-Board in action

To create a version of Tele-Board particularly suited for behaviour psychotherapy, we collaborated with the *BFA* (Berliner Fortbildungsakademie) as well as with a classic psychotherapeutic group practice. The *BFA* is a major ambulant psychotherapeutic clinic in Berlin where about 100 therapists work and where over 200 are being trained to become approbated behaviour psychotherapists. Yet, once trained most behaviour psychotherapists work in relatively small single or group practices, such as the second institution we worked with.

We started by observing the entry phase of over one hundred patients, looking at the information exchange with therapists, including bureaucratic affordances. We collected and analysed anamnesis questionnaires handed out to patients, some internationally used, some individual adaptations by the practitioners themselves. We found strong congruencies among all of them. We also considered the information demanded by insurance companies to make pay/no-pay decisions (“case reports”). Finally, we took into account thousands of pages of training material for behaviour psychotherapists, paying special attention to models which should help gather and analyse information in the course of a treatment.

Interestingly, case reports turned out to be condensed roundups of the information worked on in therapy itself (see Fig. 3).

At least in Germany, case reports sent to insurance companies cover three main subjects (1) a description of the patient’s case, (2) an analysis of his problems, (3) an outline and justification of the treatment.

Both anamnesis sheets, which patients fill out alone, and most of the questions asked by therapists in early sessions accumulate the kind of information that is needed for a case description (1). Then, models of analysis are applied which help therapist and patient jointly concretize the problems (2). In particular, there is one model which is used basically with every patient in behaviour psychotherapy. That is the SORC model, which stands for Stimulus-Organism-Reaction-Consequence. Finally, measures of treatment are scheduled (3). Based on our observations, measures are chosen dependent to a large extent on empirical evidence of efficacy and preferences of the therapist and only to a minor extent on preferences of the patient.

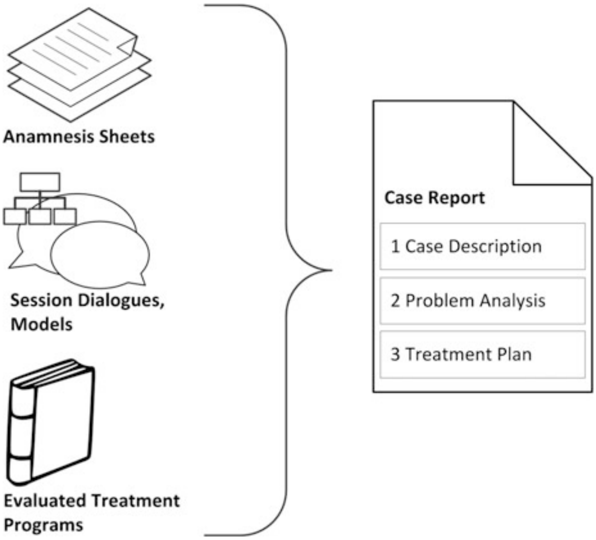


Fig. 3 Major sources of information for case reports

As a result, we began our design of Tele-Board MED for behaviour psychotherapy (see Fig. 4) by adding three features to Tele-Board.

Firstly, we created an anamnesis template with headlines and pictures to support the gathering of information for case descriptions. This template addresses the most prominent issues that we had found both in anamnesis sheets and early therapy conversations. The template was revised and improved in an iterative process as several experienced behaviour psychotherapists tried out paper prototypes in real anamnesis sessions. Secondly, we created a template with the SORC model in a way that would be easily understandable for lay people. Thirdly, we added a "Word-export function" so that the cooperatively collected information can automatically be casted into text blocks, providing facts for case reports.

Apart from that, a whole new challenge appeared with Tele-Board being used in a medical context. For good reasons, the standards of data security are rigorous in medicine. Nothing may go wrong, not even in the first trials of a new technology.

Therefore, we considered in detail both legal requirements (Bundesdatenschutzgesetz[5]), and recommendations of the *German Medical Association* (Bundesärztekammer 2008a, b). Each demand we translated into a concrete implementation plan, which was then verified by four data security experts who considered the intended measures from judicial, medical and technical perspectives.

Implementing these security measures is, of course, a labour-intensive task. In the meantime, for reasons of seriousness and responsibility, tests can only be launched with therapists while the go-ahead for use in therapy sessions with real patients must wait until the full range of security measures is in place.

[5] http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf

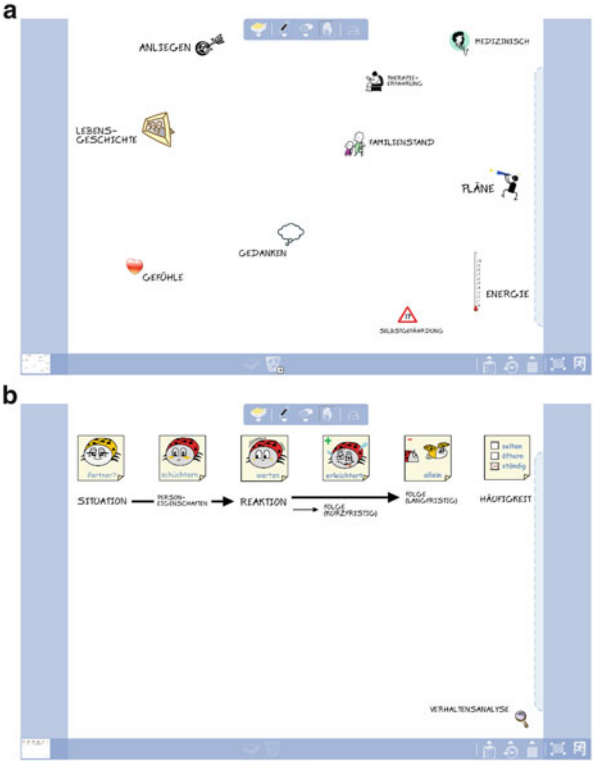


Fig. 4 (**a**) The anamnesis template (**b**) the SORC template

6 The Design Thinking Outlook: Investigating Needs

The vision of empowering patients is typically discussed as though it would favour patients more than doctors. In some discussions, the term “right” is constantly associated with “patient”, while the word “duty” goes along with “doctor”. This is certainly not our vision for Tele-Board MED. For us it is crucial to design a system that makes patient empowerment, or patient-doctor cooperation at eye level, useful and satisfying for both sides.

In line with our general outlook as members of the design thinking community, we try to understand in detail the needs and worries of all those involved to arrive at a tool which suits the users in a way that seems tailor made. Especially since the discussion of patient empowerment sometimes focuses on the needs of patients (which is good) by taking attention away from the needs of clinicians (which is bad), we actively want to serve both sides. Therefore, it makes a lot of sense to explore carefully the perspective of doctors—i.e. behaviour psychotherapists in our case. (Exploring the perspective of patients is, of course, just as important. This we can do thoroughly once Tele-Board MED fulfils all security demands so that patients may try it out safely.)

While it is common to use qualitative approaches in the field of design thinking (as described in Stanford's *bootcamp bootleg* 2010, for instance), we decided on a combined approach, quantitative and qualitative, to utilize the strengths of both strategies. After all, we want to learn as much as possible about the needs and worries of therapists—to make a start.

In the best case, the results will not only help us understand the perspective of the therapist. They might also contribute to making quantitative approaches more fashionable in the design thinking community—since, in our view, these can be truly potent tools to generate important insights regarding the needs of users.

6.1 Setting Up a Quantitative and Qualitative Study to Learn About the Needs of Therapists

To introduce the aims and functionality of Tele-Board MED, we generated a 15 min long video showing the system in action (von Thienen 2013a, b).[6] The prototype of Tele-Board MED that therapists became acquainted with through this video is shown in Fig. 5.

An e-mail was sent to all therapists of the *BFA*, including a link to the introductory video and a nine-page feedback questionnaire. It is clear that some people will be more likely than others to reply to such a call for participation. We wanted to avoid attracting only participants who were interested because they had already considered using a digital tool for documentation. By announcing a 50 Euro participation reward, we hoped to add at least one other motivation, which had nothing to do with documentation preferences.

The mail was sent around one evening during the national summer vacation. The next day before noon we had to withdraw the call for participation and temporarily remove the video from the internet. Overnight more therapists had watched the video and returned questionnaires than our budget foresaw.

Why were people so interested? The financial incentive barely seems to explain the great interest we observed. Given that participation would take about 2 h, the therapists received less than their normal payment. Two participants even abstained from their reward, saying they just caught interest in the project. In addition, many participants kept sending us feedback even after having received their money. So, it seems the subject truly captivated the audience.

[6] Anyone interested can view the introductory video at our project homepage https://med.tele-board.de. Here you will find two short videos named "Need" and "Solution", which had been online as one long video throughout the feedback study.

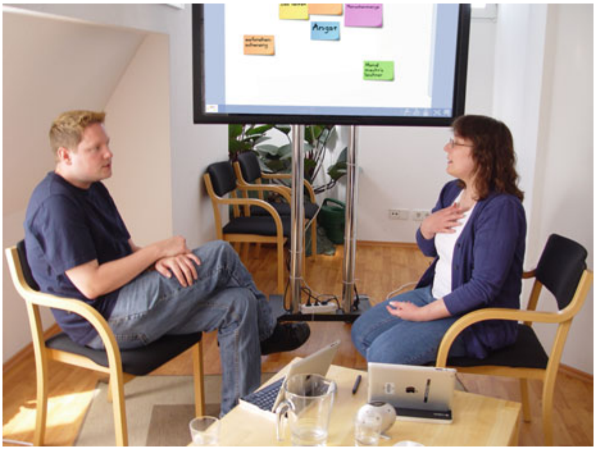


Fig. 5 Scenario of a behaviour psychotherapy session using Tele-Board MED with an interactive whiteboard and tablet computers

6.2 Sample of 34 Therapists Who Provide Feedback on Tele-Board MED

34 therapists participated in the study. The age ranged between 27 and 61 with an average of 35.4. In line with the general distribution of gender at the *BFA*, there were more females than males. Most therapists had 1–2 years of work experience (see Fig. 6).

From earlier studies we knew that affinity towards technology affects the appreciation of digital compared to analogue documentation tools (Gumienny 2013). To assess the technology preferences of participating therapists, we used two different strategies. Firstly, we asked people directly.

How would you describe your attitude towards technology
(computers and mobile phones in particular)?

□	□	□	□	□
technology hostile	technology sceptic	neutral	technology friendly	technology enthusiastic

Secondly, we listed several common devices (PC, Mac, mobile phone, video camera), programs (Word, mail, Skype) and message services (SMS, video), asking people how often they had used these so far. Multiple choice answers ranged from ‘never’ to ‘often’. Thus, we could calculate an average estimate of technology usage.

As one would expect, the attitude people reported towards technology correlated substantially with the behaviour they stated in terms of actual technology usage

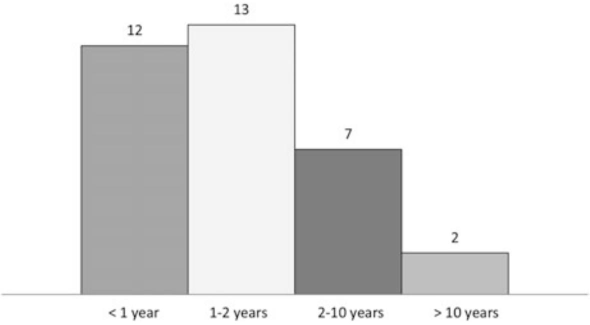


Fig. 6 Work experience of therapists in our sample

(.497**, $p \leq .003$).[7] That is, people who say they like technology also seem to use it more frequently.

In addition, there is a slight negative correlation with regard to age. That is, younger therapists are more technology-friendly than older therapists. Yet, this correlation is not significant (−.165, $p \leq .351$). Somewhat stronger but still statistically below significance is the finding that older therapists use less technology. "Age" correlates with "technology usage" by −.263, $p \leq .132$.

6.3 The Status Quo of Writing Case Reports

One of the major advantages Tele-Board MED may hold in store for therapists is that it eases the writing of case reports. But is this truly so? Maybe therapists themselves don't feel a need to improve their writing experience.

We asked the therapists. . .

How annoying do you find writing case reports?

□ not annoying at all	□ somewhat annoying	□ highly annoying

No therapist selects the first multiple choice option, indicating there would be nothing annoying about writing case reports for him. By contrast, more therapists state they find it highly annoying (18 out of 34) than therapists who find it just somewhat annoying (16 out of 34).

[7] Unless specified otherwise, all correlations are Pearson correlations. P-values describe levels of statistical significance. One asterisk (*) signals that the correlation is significant at a level of $p \leq .05$. Two asterisks (**) indicate that the correlation is statistically significant at a level of $p \leq .01$.

But why is writing case reports so unpopular among therapists? A first explanation might be that the reports address ill-posed questions. Maybe therapists feel the treatment itself does not profit from working out such reports. Maybe they consider the task nothing but a bureaucratic necessity to obtain payment.

We asked the therapists. . .

How sensible do you find it in general to think about
the subjects that case reports address?

□ not sensible at all □ somewhat sensible □ highly sensible

Here, the answers are even more pronounced. Almost all therapists declare they find the issues addressed by case reports highly sensible (30 out of 34). The remaining four therapists still find them "somewhat sensible". So, obviously, a lack of sense does not explain why the task of writing the reports is unpopular.

6.3.1 Case Report Content Makes Sense for Newcomers and Seniors

In our sample, there are a lot of young therapists with only a few years of work experience. Maybe they find it particularly sensible to think about the subjects of case reports (such as spelling out and justifying treatment plans) because it ensures that the job be done carefully. Yet, more experienced therapists might be able to do all of that easily in their minds. Maybe experienced therapists don't have to write everything down. Thus, more experienced therapists might assign lower sense-ratings to case reports than less experienced therapists.

However, this is not the case. There is no negative correlation or, said in another way, a minute positive correlation between the work experience a therapist has and how sensible he or she finds deliberating the issues of case reports. Statistically, the variable "years of practice" correlates with "sense" by .089 ($p \leq .618$). "Number of written case reports" correlates with "sense" by .041 ($p \leq .817$)—all Spearman correlations, taking into account that work experience has been assessed on an ordinal scale level.

6.3.2 A Task That Drags On

One likely reason why therapists consider writing case reports annoying (despite finding its questions sensible!) might be that the process of writing is currently very inefficient. After all, therapists typically take handwritten notes during treatment sessions nowadays, which need to be sorted and typed in a computer to produce machine-written reports. This way of producing case reports may be a task that requires a large amount of time.

To estimate how long writing one case report takes, we did a minute pre-study before sending out questionnaires. Two therapists with strongly diverging levels of experience timed themselves whenever they wrote a report. In that pre-study, the "newcomer therapist" had about 1 year of work experience while the "senior therapist" had about 30 years. Considering 20 reports each, the newcomer needed around 8 h on average for one report while the senior needed only 6 h.[8]

Our feedback study with *BFA*-therapists suggests that these numbers are quite typical. We asked the therapists:

How much time do you typically need to write a case report?

□ up to 5 hours □ 5 to 8 hours □ more than 8 hours

Most therapists reply they need 5–8 h for one report. Among the therapists who don't find themselves in this range, more people say they need even longer than people who indicate they are faster (see Fig. 7).

Given that it is not always easy to spend several hours en bloc especially since therapists typically see one patient after the other throughout a workweek, it might well occur that they fail to finish these reports promptly, e.g., within a week. Accordingly, we asked the therapists:

How often do you need a week or more to finish one case report?

□ never □ seldom □ often □ almost always

Among the 34 participants nobody states that he or she basically always gets done in a week. More than two thirds of the therapists say they often or almost always need more than 1 week (see Fig. 8).

[8] In an earlier lecture we mentioned this comparison between a newcomer and a senior therapist based on timekeeping for ten case reports each. We continued timekeeping for another ten reports each and obtained very similar numbers once again.

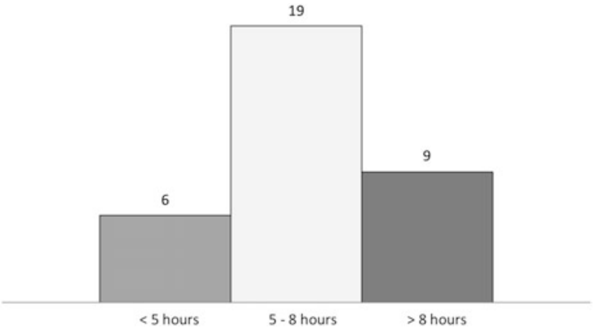


Fig. 7 Therapists report the time they need for one case report

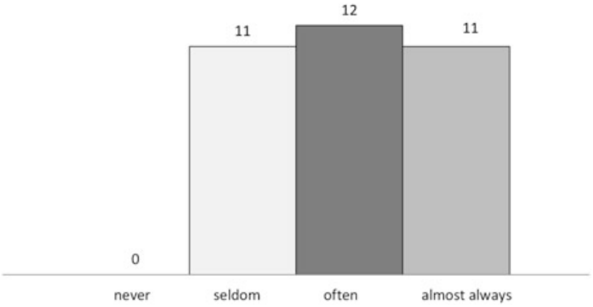


Fig. 8 How often more than a week is needed to finish a case report

6.3.3 Task Irrelevant Obstacles

Another reason why therapists find writing case reports annoying might be that task-irrelevant obstacles delay the process. For instance, therapists typically use anamnesis questionnaires filled out by patients to assemble all the information needed for case descriptions. If it is hard to decipher the patient's handwriting, that might be quite annoying—especially if you bear in mind that patients typically fill out several pages of anamnesis questions, so that therapists have a lot to read. We asked the therapists:

How often does it happen that you cannot read the information provided by your patients?

□	□	□	□
0-20% of patients	20-50% of patients	50-70% of patients	70-100% of patients

Indeed there are strong indicators that the job of having to decipher the patient's handwriting figures quite substantially in the practice of therapists nowadays. Why?

Because therapists get better at it over time! They obviously train a lot due to their job.

In terms of statistics, the correlation between "years of experience as a therapist" and "percentage of patients with unreadable handwriting" is markedly negative (Spearman correlation −.419*, $p \leq .014$). The same result is obtained when work experience is measured by the number of case reports one has already written. Again, the correlation is negative at a statistically significant level (Spearman correlation −.357*, $p \leq .038$). In other words: The more work experience a therapist has, the better he can read his patients' handwriting.

Of course, one may wonder if age could provide an alternative explanation for such a relation. Perhaps everyone trains reading other people's handwriting year after year in life. Then, older therapists should be better at reading their patients' handwriting—regardless of how much work experience they have had. But this is not the case. The correlation between age and the (in-)ability to read the patients' handwriting is impressively low at .004 ($p \leq .982$)—again, a Spearman correlation for reasons of comparability. Obviously, age has no effect. It is not the case that older therapists are more capable of reading their patients' handwriting. Only therapists with greater work experience are better at deciphering.

6.3.4 Pragmatism Reduces Time and Quality

An indicator for the therapists' urgent wish to speed up the writing of case reports may be that they acquire pragmatic strategies, which help them to finish quicker—at the risk of being less precise or even erroneous in their reports. Apart from that, pragmatic strategies might also be invoked to increase the chances of receiving payment from insurance companies.

How often are you pragmatic in the following points to finish a case report as quickly as possible or to increase granting opportunities?

□	□	□	□
never, in no report	seldom, up to 10% of reports	often, 10-50% of reports	frequently, more than 50% of reports

We offered a list of six likely strategies and in a blank field, therapists could also add pragmatic strategies they used but which we had not mentioned.

To discuss just the list provided by us, therapists do in fact admit to use pragmatic strategies even at the cost of attaining less precise or erroneous case reports. Figure 9 gives an overview.

Correlations with "hours needed for each report" suggest that being pragmatic does actually pay off for the therapists in terms of reduced writing hours. The more

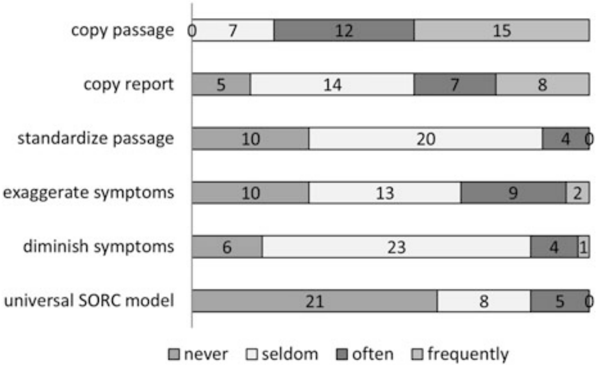
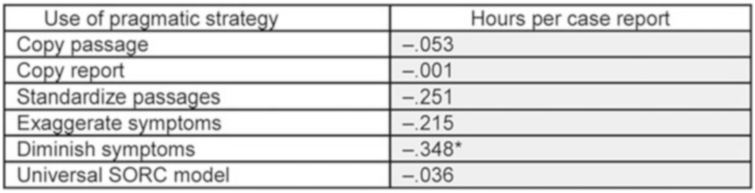


Fig. 9 How often therapists report to use six predefined pragmatic strategies



Use of pragmatic strategy	Hours per case report
Copy passage	–.053
Copy report	–.001
Standardize passages	–.251
Exaggerate symptoms	–.215
Diminish symptoms	–.348*
Universal SORC model	–.036

Fig. 10 The relation between “pragmatism” and “writing time” is consistently negative. *Grey cells* indicate a negative correlation for speed-reading

they use pragmatic strategies, the quicker they finish their reports (see Fig. 10). These correlations hold even when you control for other factors that might figure in the background, such as the work experience of therapists.

This certainly is an interesting finding as it may be translated into an important goal for Tele-Board MED. Pragmatic strategies as described above endanger the quality of case reports. We want to help therapists reduce their writing time while at the same time holding constant or rather increasing the correctness of files!

6.4 *Tele-Board MED Helps to Write Case Reports*

Tele-Board MED offers several features to reduce the time therapists need for writing case reports. Firstly, all information is collected in a digital format right from the start. Secondly, it makes sense to already sort the data in therapy sessions and not afterwards—especially since this can be done so easily and quickly with just the swipe of a finger. Thirdly, searching digital data can be automated. So, even if there was some unsorted information, it could be sifted automatically and efficiently. Finally, we implemented a text export function, so that all the data gathered on Tele-Board MED can automatically be cast into building blocks for case reports in Word format.

Our introductory video showed an example of text export in real time. We then asked the therapists to estimate how much time they personally would save with this feature.

Since a predefined scale of time-saving in terms of minutes or hours might have biased the replies, we left a blank for the therapists to fill in their estimated time savings. Some participants chose to report hours, some reported percentage of labour time. On average, participants of the first group said they would save 2.01 h per case report if they could use the Word-export function of Tele-Board MED. On average, participants of the second group estimated they would save 37.9 % of labour time per report.[9]

Two participants chose yet another scale as they filled out the questionnaire. One person said he or she would save no time at all. The other person said he or she would save a couple of days per report.

All in all, the replies of the therapists are very well in line with one another, despite their differing scales. Given that many therapists need around 6 h for one case report, saving 2 h per report equals saving one third of the time.

In general, the participants expect great time-savings from the Word-export function of Tele-Board MED. In addition, important for us is that, unlike pragmatic strategies, the Word-export function does not generate errors or haziness in case reports. By replicating information digitally from the pertinent patient record into a Word file, every detail should be transferred correctly.

6.5 Tele-Board MED Helps to Provide Electronic File Copies

Another important advantage that Tele-Board MED may hold in store for therapists concerns new demands on record transparency. Likely, more and more patients will want to see their patient file or even take a copy home. In Germany, for instance, since 2013 every patient has a legal right to receive an electronic copy of his own file.

We asked the therapists to consider the following scenario:

[9] The percentage named here differs from the percentage reported in an earlier oral presentation. Here we could include one more reply of a participant whose questionnaire was only partially readable at first.

Imagine one of your patients wants to exercise his right, granted by the new Patients' Rights Act, to obtain a complete electronic copy of his medical record, including a list of each therapy session with the corresponding treatment measures. How would you go about solving this task without Tele-Board MED?

Most therapists write that they are not at all prepared for such a task. They could provide an electronic copy of a comprehensive patient record only with an enormous investment of extra work. As one participant puts it: "That would be complicated and time consuming. I would have to write a comprehensive summary of every treatment session. If the demand was made after several therapy sessions, that would be highly labour-intensive."[10] One therapist frankly admits: "I would generally discuss the sense of the whole thing. Then I would say that I don't have an electronic file copy and discuss with the patient whether he or she truly needs it. I cannot imagine that more than 2 % of the patients would still insist."

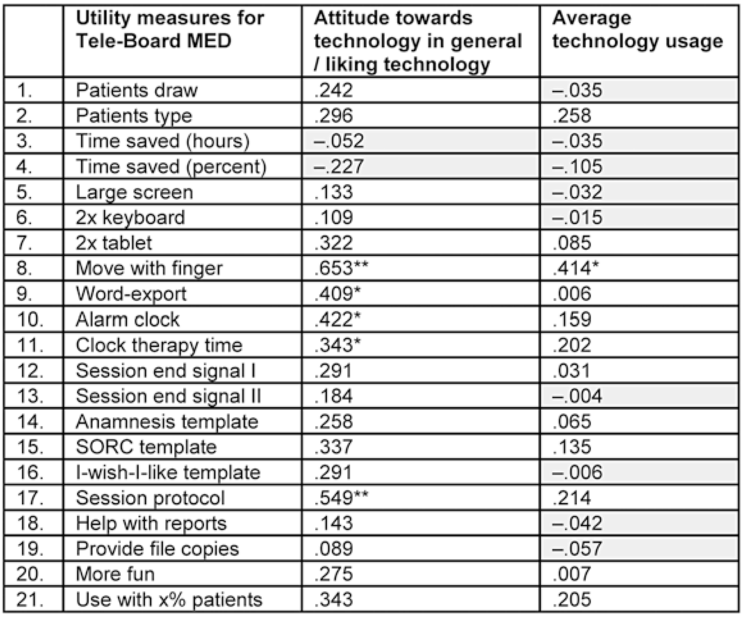
Tele-Board MED helps to provide electronic file copies insofar as all information is stored in a digital form right from the start. The system may also track which templates therapist and patient use throughout a session to create a treatment protocol automatically. For instance, the system may record that in session x therapist and patient worked out a problem analysis, using the SORC template.

After asking the therapists how they would provide electronic file copies apart from Tele-Board MED, we invoked a comparison, asking how electronic file copies could be provided more easily. Was their own strategy the easier one or would it be more easy to use Tele-Board MED? Therapists could pick one of the following answers.

□	□	□	□	□
much easier with my strategy	somewhat easier with my strategy	neither nor	a little easier with Tele-BoardMED	much easier with Tele-Board MED

Mapping these answers on a scale from −5 (much easier with my strategy) to 5 (much easier with Tele-Board MED), the average rating is 3.8. That suggests Tele-Board MED does indeed provide a functionality which cannot be replaced easily by any other strategy the therapists could think of.

[10] This quote and the following from study participants have been translated by us from German to English.



	Utility measures for Tele-Board MED	Attitude towards technology in general / liking technology	Average technology usage
1.	Patients draw	.242	–.035
2.	Patients type	.296	.258
3.	Time saved (hours)	–.052	–.035
4.	Time saved (percent)	–.227	–.105
5.	Large screen	.133	–.032
6.	2x keyboard	.109	–.015
7.	2x tablet	.322	.085
8.	Move with finger	.653**	.414*
9.	Word-export	.409*	.006
10.	Alarm clock	.422*	.159
11.	Clock therapy time	.343*	.202
12.	Session end signal I	.291	.031
13.	Session end signal II	.184	–.004
14.	Anamnesis template	.258	.065
15.	SORC template	.337	.135
16.	I-wish-I-like template	.291	–.006
17.	Session protocol	.549**	.214
18.	Help with reports	.143	–.042
19.	Provide file copies	.089	–.057
20.	More fun	.275	.007
21.	Use with x% patients	.343	.205

Fig. 11 Correlations between utility ratings for Tele-Board MED and measures of attitude towards technology in general or experience with technology; a *grey cell* background indicates negative correlations

6.6 Digging Deeper to Discern Needs and Insights

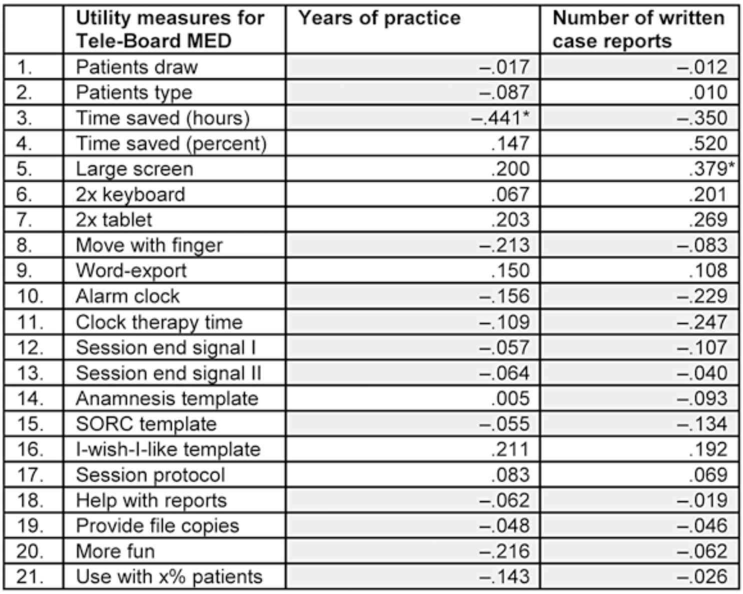
Regarding the compilation of case reports and the ease of handing out electronic file copies to patients, we had had prior hypotheses that were tested—and confirmed—straightforwardly in our feedback study. Yet, we had launched the study also for another reason. We wanted to learn more about the specific situation and needs of therapists. This would, of course, include finding out more about both, the commonalities and important differences among therapists.

One issue that we wanted to look at more closely has already been addressed. We wanted to learn how attitude towards technology or experiences in technology usage would impact the attitude towards Tele-Board MED.

6.6.1 Does Attitude Towards Technology Make a Difference?

To assess how therapists conceive of Tele-Board MED, we asked for 21 distinct utility judgements, some concerning the system in general and some concerning single features of it. Figure 11 gives an overview.

Indeed, therapists rate the utility of Tele-Board MED differently depending on how they conceive of technology in general. While both “attitude towards



	Utility measures for Tele-Board MED	Years of practice	Number of written case reports
1.	Patients draw	–.017	–.012
2.	Patients type	–.087	.010
3.	Time saved (hours)	–.441*	–.350
4.	Time saved (percent)	.147	.520
5.	Large screen	.200	.379*
6.	2x keyboard	.067	.201
7.	2x tablet	.203	.269
8.	Move with finger	–.213	–.083
9.	Word-export	.150	.108
10.	Alarm clock	–.156	–.229
11.	Clock therapy time	–.109	–.247
12.	Session end signal I	–.057	–.107
13.	Session end signal II	–.064	–.040
14.	Anamnesis template	.005	–.093
15.	SORC template	–.055	–.134
16.	I-wish-I-like template	.211	.192
17.	Session protocol	.083	.069
18.	Help with reports	–.062	–.019
19.	Provide file copies	–.048	–.046
20.	More fun	–.216	–.062
21.	Use with x% patients	–.143	–.026

Fig. 12 Spearman correlations between utility ratings for Tele-Board MED and work experience of therapists

technology" and "average technology usage" help to predict utility ratings for Tele-Board MED, the attitude variable is the more potent predictor. Except for two out of 21 variables, there are positive correlations between "finding Tele-Board MED useful" and "appreciating technology in general".

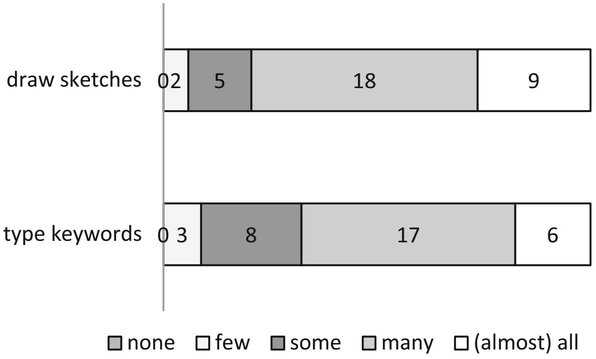
In contrast, there is not such a strong and consistent relationship between "finding Tele-Board MED useful" and "using technology a lot". Here, the correlations are not only lower; some of them are even negative.

Figure 11 shows how "attitude" is a better predictor for utility ratings than "average technology usage". Grey cells indicate negative correlations with utility ratings. The right column of "average technology usage" obviously contains more of these grey cells than the middle column of "liking technology".

6.6.2 Is Tele-Board MED Most Helpful for Newcomers?

Tele-Board MED offers several features which might be particularly attractive for the newcomers among therapists. E.g., when you haven't met too many patients yet, will you truly remember to ask all the anamnesis questions that you should be asking? The anamnesis template of Tele-Board MED might not only provide orientation for patients, but for therapists as well. Therefore, inexperienced therapists might generate higher utility ratings for Tele-Board MED than experienced therapists.

Fig. 13 Therapists answering the question: What do you think, how many patients could. . .?



Yet, this is not the case. To start with the anamnesis template, no statistically relevant relation to work experience shows up. Moreover, considering all 21 utility measures for Tele-Board MED, there is no trend that newcomers appreciate the tool more than senior therapists. In Fig. 12 about as many positive correlations (white cells) as negative correlations (grey cells) can be found.

6.6.3 How Important Is the Involvement of Patients for Therapists?

In our questionnaire, we used three items to assess expected patient involvement from the point of view of the therapists. We asked:

(1) What do you think, how many patients could
draw sketches on a tablet or screen?

□	□	□	□	□
none, 0%	few, up to 10%	some, 10-50%	many, 50-90%	(almost) all, 90-100%

(2) What do you think, how many patients could
type in keywords on a keyboard?

□	□	□	□	□
none, 0%	few, up to 10%	some, 10-50%	many, 50-90%	(almost) all, 90-100%

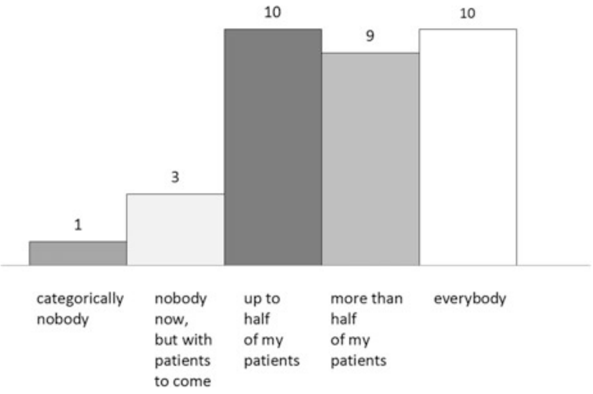


Fig. 14 Therapists answering the question: With what percentage of your patients could you imagine using Tele-Board MED? (The sample size is N = 33 here due to one missing answer.)

The answers diverge strongly (see Fig. 13).

A third measure of expected patient involvement asks for a self-rating of the therapists.

(3) With what percentage of your patients could you imagine using Tele-Board MED?

□	□	□	□	□
categorically nobody, because______	nobody now, but patients to come	up to half of my patients	more than half of my patients	everybody

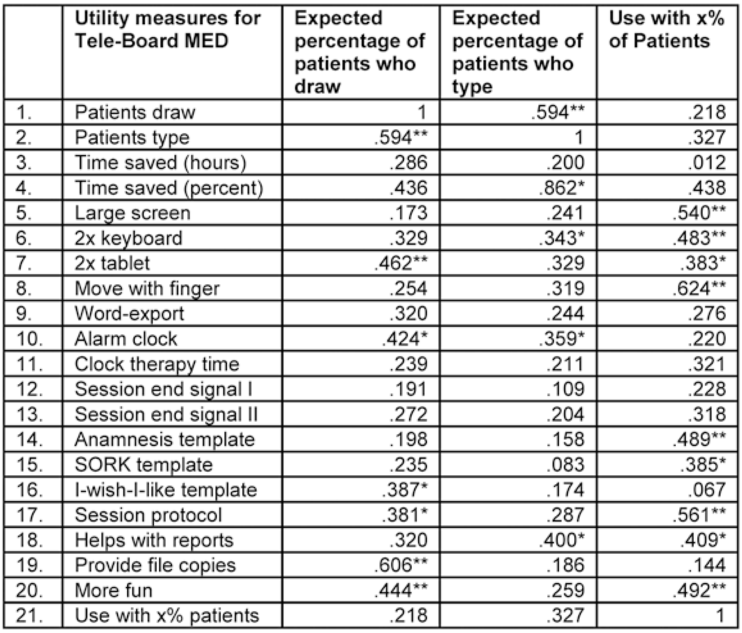
Figure 14 gives an overview of the answers, which diverge strongly once again.

The vast majority of therapists can well imagine using Tele-Board MED (88 %). Only 12 % of the therapists are more hesitant and would not use the system with any of their present patients.

One therapist states that he categorically cannot imagine using the system. In the following free text field he explains: “Since legitimately patients will be afraid that their data might be abused, by health insurance companies for instance, that data might be copied without noticing, that data will get lost etc. I would use the system, but only for a very rough documentation of therapy sessions.”

From the point of view of data analysis, varying answers are typically the richest medium for finding patterns. After all you can explain why some people say “a” while others say “b” only when there is some divergence in the statements.

Interestingly, the expected patient involvement turns out to be a potent predictor of how useful a therapist finds Tele-Board MED in general. Regardless of what measure of patient involvement you pick and regardless of what utility measure for Tele-Board MED you pick—every single correlation is positive (see Fig. 15)!



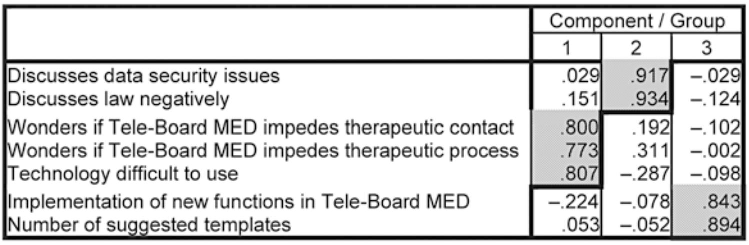
	Utility measures for Tele-Board MED	Expected percentage of patients who draw	Expected percentage of patients who type	Use with x% of Patients
1.	Patients draw	1	.594**	.218
2.	Patients type	.594**	1	.327
3.	Time saved (hours)	.286	.200	.012
4.	Time saved (percent)	.436	.862*	.438
5.	Large screen	.173	.241	.540**
6.	2x keyboard	.329	.343*	.483**
7.	2x tablet	.462**	.329	.383*
8.	Move with finger	.254	.319	.624**
9.	Word-export	.320	.244	.276
10.	Alarm clock	.424*	.359*	.220
11.	Clock therapy time	.239	.211	.321
12.	Session end signal I	.191	.109	.228
13.	Session end signal II	.272	.204	.318
14.	Anamnesis template	.198	.158	.489**
15.	SORK template	.235	.083	.385*
16.	I-wish-I-like template	.387*	.174	.067
17.	Session protocol	.381*	.287	.561**
18.	Helps with reports	.320	.400*	.409*
19.	Provide file copies	.606**	.186	.144
20.	More fun	.444**	.259	.492**
21.	Use with x% patients	.218	.327	1

Fig. 15 Spearman correlations between utility ratings for Tele-Board MED and expected participation of patients/use of system; *white cell* backgrounds indicate positive correlations, *grey cell* backgrounds would indicate negative correlations

Some of the more minute findings are quite expected. Comprehensibly, a therapist who believes many patients can draw is more likely to appreciate the idea of introducing two tablets, one for the therapist and one for the patient to draw on. Correspondingly, therapists who believe many patients can type are more likely to appreciate the idea of introducing two keyboards. Expected correlations like these are always a good sign in terms of data consistency.

An interesting finding that we did not anticipate is that therapists who believe many patients can draw are more likely to think using Tele-Board MED is fun. From the point of view of system design, this is certainly important as one may try to make drawing particularly easy and likely, both for patients and doctors.

A correlation that stands out by sheer number reaches up to .862 ($p \leq .013$). That correlation obtains between “expected percentage of patients who type” and “saved labour time in percent when using the Word-export function of Tele-Board MED”. The outstandingly high correlation seems to suggest these therapists imagine their patients typing in anamnesis data for case reports. Again, this idea is interesting from the design point of view. One might indeed consider replacing traditional paper questionnaires by corresponding questionnaires on Tele-Board MED. For that purpose, a form function would need to be implemented in the system.



	Component / Group		
	1	2	3
Discusses data security issues	.029	.917	–.029
Discusses law negatively	.151	.934	–.124
Wonders if Tele-Board MED impedes therapeutic contact	.800	.192	–.102
Wonders if Tele-Board MED impedes therapeutic process	.773	.311	–.002
Technology difficult to use	.807	–.287	–.098
Implementation of new functions in Tele-Board MED	–.224	–.078	.843
Number of suggested templates	.053	–.052	.894

Fig. 16 A Principal Component Analysis with Varimax Rotation and Kaiser Normalization yields three factors (user groups). Highlighted cells indicate major correlations with qualitative feedback

6.6.4 What Different User Groups Can Be Discerned?

In our study, we invited the therapists on several occasions to add comments in free text fields: thoughts, wishes or criticism. Indeed, the therapists provided a lot of qualitative feedback this way. Some issues caught our special attention since they were discussed by several therapists—and sometimes repeatedly in a single questionnaire.

We counted how many different free text fields each therapist used to (a) discuss issues of data security, (b) criticise the new German law on patients' rights, (c) express concern that Tele-Board MED might impede the therapeutic contact, (d) express concern that Tele-Board MED might impede the therapeutic process, (e) mention that technology may be difficult to use or (f) suggest additional functionalities on Tele-Board MED that would be valuable. In addition, we counted the number of templates that each therapist recommended for implementation in the future.

A factor analysis (Principal Component Analysis with varimax rotation) suggests that basically three different user groups participated in our feedback study (see Fig. 16). All together the identified three factors account for 77.8 % of the overall data variance.

One user group fears that a technology like Tele-Board MED might impede the therapeutic contact and process. For instance, a therapist of this group expresses his fear "that patient and therapist turn towards a screen instead of using eye contact to build up and deepen the therapeutic relationship".

According to this user group, difficulties of handling technology provide an additional reason why Tele-Board MED might disturb treatments. As one therapist remarks, "I don't want to start a therapy by handing over a user manual to the patient. The time he would need to learn to handle the tablet is time I would rather use for therapy."

From our point of view, the fortunate part of this feedback may be that it could be addressed by design. In addition, success may be evaluated empirically. Several setups can be tested until finally a version is found where Tele-Board MED does not disrupt the flow of therapeutic conversation at all. Certainly, it is crucial to attain empirical measures of how the psychotherapeutic contact and process is influenced by Tele-Board MED. Yet we want to stress once more that influencing patient-doctor-interactions favourably is one of the major motivations for developing the system in the first place!

A second user group is sceptical towards the new law on patients' rights. Some wonder whether full transparency of patient files is such a good idea at all. As one therapist puts it: "I find the law somewhat questionable and I think Tele-Board MED would encourage more patients to demand access to their files."

Therapists of this second group also tend to fear that data might not be safe. "Shouldn't our profession strongly oppose this law? It is so critical because of the impossibility to store data safely! Every day we hear of information theft on every possible level!"

Starting with the issue of full record transparency, it remains for our team to suggest ways of documenting where such comprehensive transparency works well for both doctors and patients. That may be a challenge indeed. At the same time, the task is not specific for Tele-Board MED. The whole community is confronted more and more with transparency requests, be it as a result of new laws or due to more self-confident patients.

Regarding the issue of data security it needs to be mentioned that our introductory video did not address data security measures of Tele-Board MED in any concrete form. We only mentioned how crucial the issue is in general. Correspondingly, the therapists don't criticise anything in concrete terms. They rather seem to express a bad feeling.

To address bad feelings regarding data security, empirical evidence will probably not suffice. After all, many practices are common among therapists nowadays, which are not completely fail-safe, but which are used nonetheless and one has gotten used to them. Video or audio recordings of therapy sessions and paper records, for instance, are often kept in places where a committed thief could steal them rather easily. Therefore, users of this second group might need something in addition to solid data security measures—which are nevertheless indispensable. These users may need something that reaches gut feelings. Maybe seeing trustworthy professional colleagues using Tele-Board MED with a positive result might help them acquire a well warranted good feeling too.

A third group of users is not concerned with things that could go wrong. As if part of the developer team, they simply suggest yet other functionalities that might make the system even better. Indeed, we received many pages of ideas that will inspire many new and promising features of Tele-Board MED.

7 Resume and Mission

Everyone is a patient one time or other. From the point of view of a patient, you certainly hope that your medical records, which doctors assemble of your case, are correct. Treatment plans often depend on this information. Against this background, a likely idea for twenty-first century medicine is to invite patients to help check their files and co-decide on their own treatments. Indeed, this is a common call among today's scientists and politicians. However, it happens easily that the focus is shifted towards patients at the expense of clinicians. As much as we embrace the vision of patient-doctor-cooperation at eye level, our perspective as members of the design thinking community decidedly focuses on the needs of all core stakeholders. By creating Tele-Board MED, we hope to improve medical documentation and patient-doctor-interaction in the experience of all protagonists. Starting in the domain of behaviour psychotherapy, we used quantitative and qualitative approaches to learn more about the needs of the stakeholders. A first design thinking driven feedback study with 34 behaviour psychotherapists showed that the means to support cooperation at eye level may indeed have much going for clinicians as well. For instance, Tele-Board MED promises vastly reduced labour hours when therapists write case reports. Apart from these auspicious findings, the consideration of quantitative data as well as qualitative feedback helped gain crucial insight to inform the future development of Tele-Board MED.

References

Bundesärztekammer, Kassenärztliche Bundesvereinigung (2008a) Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt 105(19):A 1026–A 1030

Bundesärztekammer, Kassenärztliche Bundesvereinigung (2008b) Technische Anlage zu den Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt 105(19):1–12

Bundesdatenschutzgesetz (1990) http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf. Accessed 5 Dec 2013

Bundesministerium für Gesundheit (2013) Patientenrechtegesetz—Die Patientenrechte stärken, http://www.bmg.bund.de/praevention/patientenrechte/patientenrechtegesetz.html. Accessed 5 Dec 2013

European Network on Patient Empowerment (ENOPE) (2012) Patient empowerment—living with chronic disease. In: A series of short discussion topics on different aspects of self management and patient empowerment for the 1st European conference on patient empowerment. http://www.careum-congress.ch/documents/10192/13742/ENOPE+Paper+Patient+Empowerment/bc9088a7-6005-4f30-9ef2-de63c07b1e4c. Accessed 5 Dec 2013

Funnell MM, Anderson RM (2004) Empowerment and self-management of diabetes. Clin Diabetes 22(3):123–127

Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten (2013) Bundesgesetzblatt Teil I Nr. 9, ausgegeben zu Bonn am 25. Februar 2013

Grawe K (2005) Empirisch validierte Wirkfaktoren statt Therapiemethoden. In: Report Psychologie 7/8: 311

Gumienny R (2013) Understanding the adoption of digital whiteboard systems for collaborative design work. Unpublished dissertation, HPI at the University of Potsdam, Potsdam, Germany

Gumienny R, Gericke L, Quasthoff M, Willems C, Meinel C (2011) Tele-board: enabling efficient collaboration in digital design spaces. In: Proceedings of the 15th international conference on computer supported cooperative work in design (CSCWD 2011), IEEE Press, Lausanne, Switzerland, pp 47–54

Hasso Plattner Institute of Design at Stanford (2010) Bootcamp bootleg. http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf. Accessed 5 Dec 2013

Kalra D (2011) Health informatics 3.0. Yearb Med Inform 6(1):8–14

Koch S (2012) Improving quality of life through eHealth—the patient perspective. Stud Health Technol Inform 180:25–29

Koch S, Vimarlund V (2012) Critical advances in bridging personal health informatics and clinical informatics. Yearb Med Inform 7(1):48–55

Lorig KR, Sobel DS, Stewart AL, Brown BW Jr, Ritter PL, González VM, Laurent DD, Holman HR (1999) Evidence suggesting that a chronic disease self-management program can improve health status while reducing utilization and costs: a randomized trial. Med Care 37(1):5–14

Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, González VM, Laurent DD, Holman HR (2001) Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 39(11):1217–1223

McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, Wagner EH (2011) Self-management. Enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 61(1):50–62

Perlich A (2012) Designing an e-service for stroke patients. How can visualization support the management of the individual care process? Master thesis, University of Heidelberg, Heidelberg, Germany

Schweizerische Akademie der Medizinischen Wissenschaften (SAMW), Verbindung der Schweizer Ärztinnen und Ärzte (FMH), Medizinische Fakultäten der Universitäten Basel, Bern, Genf, Lausanne und Zürich (2004) Ziele und Aufgaben der Medizin zu Beginn des 21. Jahrhunderts. SAMW/ASSM, Basel

von Thienen JPA (2013) Einführung für Verhaltenstherapeuten. Der Bedarf—Auflagen des neuen Patientenrechtegesetzes und Komplikationen der Antragstellung [video]. https://med.tele-board.de. Accessed 5 Dec 2013

von Thienen JPA (2013) Einführung für Verhaltenstherapeuten. Der Lösungsansatz—Wie Tele-Board MED helfen kann [video]. https://med.tele-board.de. Accessed 5 Dec 2013

Peer and Self Assessment in Massive Online Classes

Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer

Abstract Peer and self assessment offer an opportunity to scale both assessment and learning to global classrooms. This paper reports our experiences with two iterations of the first large online class to use peer and self assessment. In this class, peer grades correlated highly with staff-assigned grades. The second iteration had 42.9 % of students' grades within 5 % of the staff grade, and 65.5 % within 10 %. On average, students assessed their work 7 % higher than staff did. Students also rated peers' work from their own country 3.6 % higher than those from elsewhere. We performed three experiments to improve grading accuracy. We found that giving students feedback about their grading bias increased subsequent accuracy. We introduce short, customizable feedback snippets that cover common issues with assignments, providing students more qualitative peer feedback. Finally, we introduce a data-driven approach that highlights high-variance items for improvement. We find that rubrics that use a parallel sentence structure, unambiguous wording and well-specified dimensions have lower variance. After revising rubrics, median grading error decreased from 12.4 to 9.9 %.

C. Kulkarni (✉) • K. Papadopoulos • J. Cheng
Stanford University, Palo Alto, CA, USA
e-mail: chinmay@stanford.edu; kpapa@stanford.edu; jcccf@stanford.edu

K.P. Wei • D. Chia • D. Koller
Stanford University, Palo Alto, CA, USA

Coursera, Inc., Mountain View, CA, USA
e-mail: pangwei@coursera.org; danchia@coursea.org; koller@cs.stanford.edu

H. Le
Coursera, Inc., Mountain View, CA, USA
e-mail: huy@coursea.org

S.R. Klemmer
Stanford University, Palo Alto, CA, USA

Computer Science and Engineering, University of California, San Diego, CA, USA
e-mail: srk@cs.stanford.edu; srk@ucsd.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_9, © Springer International Publishing Switzerland 2015

1 Introduction

In the past year, hundreds of thousands of students have earned certificates in large online classes—on topics from Databases to Sociology to World Music—and millions have signed up (Lewin 2012a). These classes, often called MOOCs, provide students on-demand video lectures, often along with automated quizzes and homework, and class forums that allow students to interact with each other.

Many such classes use automated assessment [e.g. Widom (2012)], which precludes the open-ended work that is a hallmark of education in creative fields like design (Buxton 2007). Furthermore, viewing and critiquing others' work plays a key pedagogical role in these domains (Schön 1985). Fields like design have also traditionally relied on intimate co-location to enable these activities and to confer values and norms (Schön 1985). However, in a global, online classroom, students lack the shared context co-location provides. How can we scale both evaluation and peer learning in creative domains online?

One approach for scaling assessment and peer learning would be for students to evaluate their peers' work. Peer assessment potentially enables large classes to offer assignments that are impractical to grade automatically. Furthermore, human grading more easily provides context-appropriate responses and better handles ill-specified constraints (Hearst 2000). But, do students have the motivation and expertise to perform peer assessment well? This paper reports on our experiences with the first use of peer assessment in a massive online class. It is the largest use of peer assessment to date. As of June 2013, this technique has since been adopted in many other classes, including 79 MOOCs on the Coursera[1] platform alone.

1.1 The Design Studio as an Inspiration

For over a century, the studio has been a dominant model for architecture and design education, and has expanded into fields including product design (Lawson 2006), HCI (Winograd 1990; Greenberg 2009), and software design (Tomayko 1991). This paper considers the studio as an inspiration for online design education.

The studio model of education was formalized in the École de Beaux-Arts (Drexler et al. 1977). Studios provide an open, shared environment for students to work. This copresence provides social motivation and facilitates peer learning through visibility of work (Reimer and Douglas 2003). Formal and informal studio critique helps students iteratively improve their work (Schön 1985).

Public visibility of self and peer work provides students with a nuanced understanding of design. In particular, seeing their peers' work along with their own work

[1] https://www.coursera.org/

through its evolution allows students to understand decisions and tradeoffs both in their own designs, and in those of their peers (Tinapple et al. 2013).

Formative studio feedback further engages students in reflective practice (Schön 1985). Informal, formative feedback is often through oral critiques or "crits" by teachers or other experts (Uluoglu 2000). Such informal, qualitative feedback is essential, because it encourages iterative practice (Cennamo et al. 2011). Because crits are often delivered in public, students also learn from observing peer work as well as by working on their own (Dannels and Martin 2008).

Expert critiques also serve as summative assessment. Experts often assess design based on trained but tacit criteria (Snodgrass and Coyne 2006). Amabile et al. demonstrate that expert consensus is a reliable measure of the quality of creative work (Amabile 1982). Their Consensual Assessment Technique asks experts to rate artifacts on a scale, and provides no rubrics and does not ask raters to justify their rating.

Other techniques provide an assessment process to observe, interpret and evaluate work (Feldman 1994).

The design studio suggests three requirements for successful design education online. First, it must support open-ended design work with multiple correct solutions. Such work is especially important in design education because successful design often requires generating and reflecting on multiple ideas (Tohidi et al. 2006; Buxton 2007), and on exploration and iteration (Fallman 2003). Second, assessment must allow students to learn the tacit criteria of good design. Criteria for good design are often not explicitly defined (Forlizzi and Battarbee 2004). For instance, interactive interfaces may be subjectively evaluated for whether they are learnable and appropriate (Alben 1996), criteria that require tacit interpretation. Third, assessment must provide students both qualitative formative feedback, and summative feedback.

1.2 The Promise of Peer Assessment

The inherent variability of open-ended solutions, and lack of defined evaluation criteria for design makes automatically assessing open-ended work challenging (Bennett et al. 1997). In addition, automated systems frequently cannot capture the semantic meaning of answers, which limits the feedback that they can provide to help students improve (Bennett 1998; Hearst 2000).

Therefore, open-ended assignments generally rely on human graders. The time-intensive, personalized assessment of grading sketches, designs, and other open-ended assignments requires a small student-to-grader ratio (Hsi and Agogino 1995; Stanley and Porter 2002). This staff effort is prohibitive for large classes: staff grading simply doesn't scale.

Peer and self assessment is a promising alternative, with potential additional benefits. It not only provides grades, it also importantly helps students see work from an assessor's perspective. Peer feedback in design classes also creates an audience that provides honest feedback and multiple perspectives (Tinapple et al. 2013). Evaluating peers' work also exposes students to solutions, strategies, and insights that they otherwise would likely not see (Chinn 2005; Tinapple et al. 2013). Similarly, self assessment helps students reflect on gaps in their understanding, making them more resourceful, confident, and higher achievers (Zimmerman and Schunk 2001; Pintrich 1995; Pintrich and Zusho 2007) and provides learning gains not seen with external evaluation (Dow et al. 2012).

Peer assessment can increase student involvement and maturity, lower the grading burden on staff, and enhance classroom discussion (Boud 1995). Peer assessment has been used in colocated classroom settings for many different kinds of assignments (Topping 1998), including design (De La Harpe et al. 2009; Tinapple et al. 2013). programming (Chinn 2005) and essays (Venables and Summit 2003). How can we make this classroom technique scale to a large online class?

1.3 Scaling Peer Assessment

In-class peers can assess each other well (Falchikov and Goldfinch 2000; Carlson and Berry 2003; Gerdeman et al. 2007). To effectively scale peer assessment, we can learn several lessons from crowdsourcing (Surowiecki 2005). First, crowdworkers perform better when they are intrinsically motivated by the task's importance (Cheshire and Antin 2008). Second, consensus among raters serves as a useful indicator of quality (Huang and Fu 2013). Third, interfaces like FoldIt (Khatib et al. 2011) and NASA Clickworkers (Szpir 2002) demonstrate that short, well-crafted training exercises can enable legions of motivated amateurs to perform work previously thought to require years of training.

Massive online classes provide a valuable living lab (Chi 2009; Carter et al. 2008) for exploring peer-sourcing approaches, and our hope is that peer-sourcing insights from massive classes will contribute techniques that apply more broadly. These peer-sourced systems introduce new challenges and opportunities beyond crowd-sourcing. For example, students using peer assessment both create the work to be assessed and perform the assessment. One theme this paper will explore is the learning benefits that arise from those dual roles.

1.4 Contributions

This paper reports on our experiences with peer assessment over two iterations in the first large-scale class to use it (http://www.hci-class.org). Since our adaptation of peer assessment to MOOCs, variations of the system described here have since been used in dozens of other large online classes, including Mathematical Thinking, Programming Python, Listening to World Music, Fantasy and Science Fiction, and Sociology.

Over both iterations of the class, 5,876 students submitted at least one assignment and participated in peer assessment. Overall, the correlation between peer grades and staff assigned grade was r = 0.73, and the average absolute difference between peer and staff grades was 3 % (positive and negative errors were approximately balanced).

In end-of-course surveys, students reported both receiving peer feedback and performing peer assessment to be valuable learning experiences. On a seven-point Likert scale, the median rating was 6 (7 = very valuable). Surprisingly, 20 % of students voluntarily assessed more submissions than required.

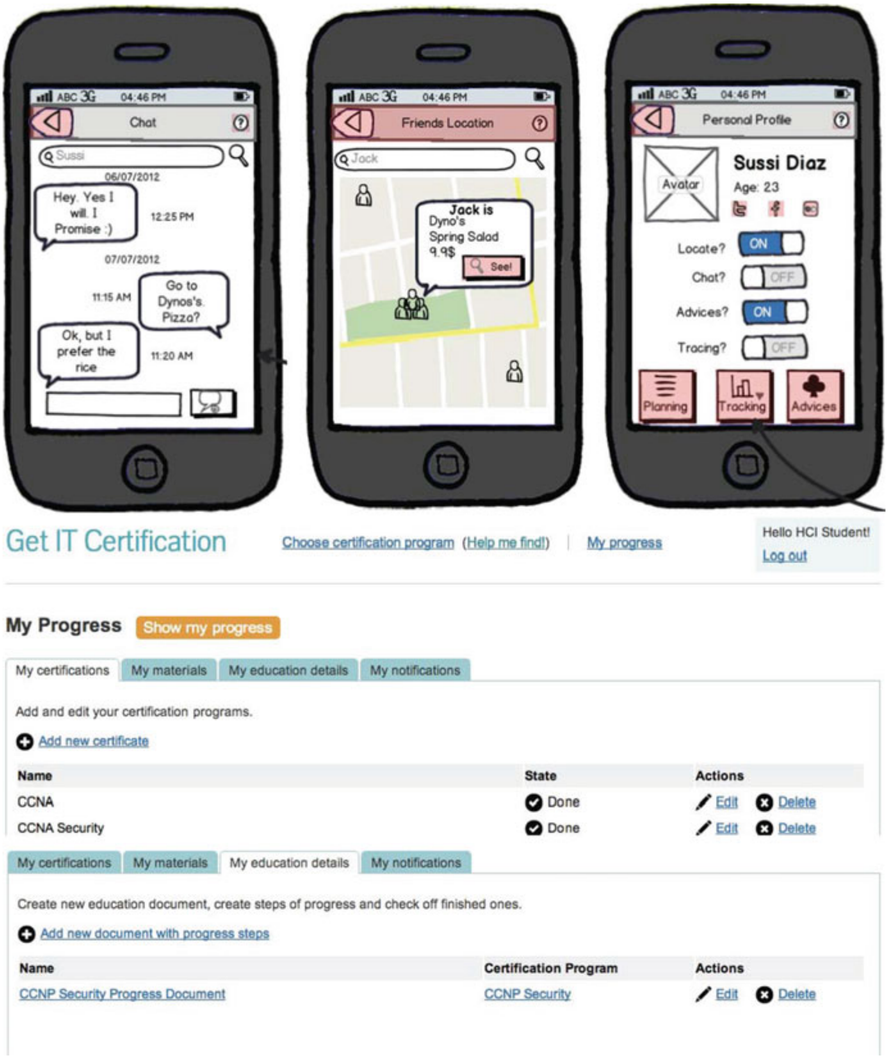
We explored several techniques to improve assessment accuracy and encourage qualitative feedback. First, we found that giving students feedback about whether they scored peers high or low increased their subsequent accuracy. A between-subjects experiment found a 0.97 % decrease in mean error (6.77 % in the experimental group, vs. 7.74 % in the control group). Second, to help students provide peers with high-quality personalized feedback, we introduce short, customizable feedback snippets that address common issues with assignments. 67 % of students obtained open-ended peer feedback using this method. Third, we introduce a data-driven approach for improving rubric descriptions. We distinguish items with high student:staff correlation from those with low correlation, and observed the ways they differ to improve the low-correlation ones. After making these changes, the mean error on grades decreased from 12.4 to 9.9 %.

2 The Anatomy of a Large Scale Online Class

This online class is an introduction to human-centered interaction design. The class is offered free of charge, and is open to any interested student. Material covered in class is based on an introductory HCI course at Stanford University. Over the class duration, students watch lectures, answer short quizzes and complete weekly assignments. In a typical week, students watch four videos of 12–15 min each. Videos total approximately 450 min across the class, and contain embedded multiple choice questions.

Multiple choice quizzes tested students' knowledge of material covered in videos. Most significantly, students completed five design assignments. Each assignment covered a step in a course-long design project where students design a Web site inspired by one of three design briefs (Fig. 1).

Students who complete the course with an average assignment score of 80 % or above earn an electronic "Statement of Achievement" for a Studio track (but no university credit). 501 students earned this statement in the first iteration, and 595 did in the second. 1,573 received a statement of achievement for the Apprentice track comprising watching videos and quiz performance in the first iteration, and 1,923 did in the second.



Get IT Certification

Choose certification program (Help me find!) | My progress

Hello HCI Student!
Log out

My Progress Show my progress

My certifications | My materials | My education details | My notifications

Add and edit your certification programs.

Add new certificate

Name	State	Actions
CCNA	Done	Edit Delete
CCNA Security	Done	Edit Delete

My certifications | My materials | My education details | My notifications

Create new education document, create steps of progress and check off finished ones.

Add new document with progress steps

Name	Certification Program	Actions
CCNP Security Progress Document	CCNP Security	Edit Delete

Fig. 1 Example prototypes from student projects in the online class (*top*: early prototype of a social dining app; *bottom*: a tracker for professional certification at the end of class)

2.1 By the Numbers

Similar to other online classes (Lewin 2013a), the online HCI class attracted numerous and diverse participants. 30,630 students watched videos in the first iteration, and 35,081 did in the second (32.5 % of students in each iteration were female). 55 % of students reported they had full time jobs (in both iterations). The median age range in both iterations was 25–34, with a broad spread (Fig. 2). In both

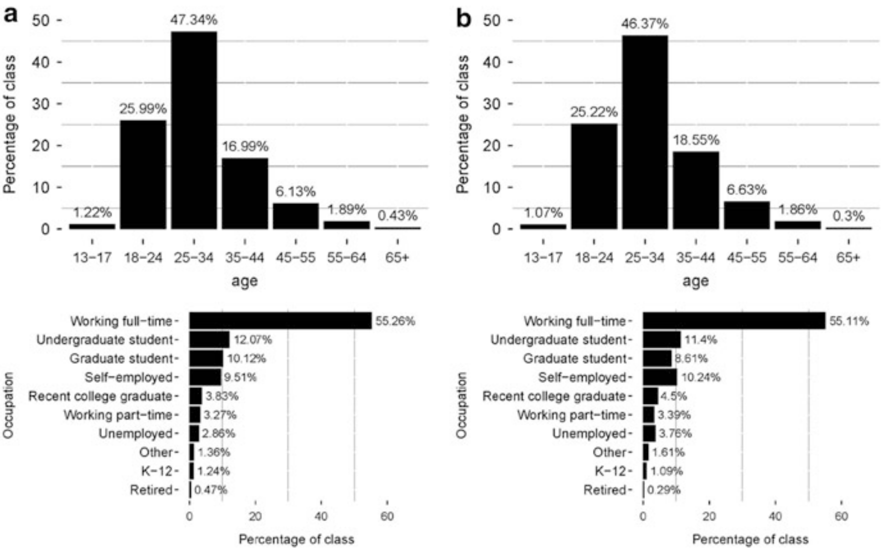


Fig. 2 Online classes attract students who cannot use traditional universities, such as those working fulltime. The age distribution of the class is remarkably similar across both iterations. (**a**) Spring 2012 (iteration 1), 10,190 participants, (**b**) Fall 2012 (iteration 2), 17,915 participants

iterations, students from 124 countries registered for the class and roughly 71 % were from outside the United States. Students transcribed lectures in 13 languages: English, Spanish, Brazilian Portuguese, Russian, Bulgarian, Japanese, Korean, Slovak, Vietnamese, Chinese (Simplified), Chinese (Traditional), Persian, and Catalan.

In all, 2,673 students submitted assignments in the first iteration, and 3,203 in the second (Fig. 3). The second iteration also allowed students to submit assignments in Spanish; 223 students did so. Student questions were answered exclusively through the online class forum. Across the course, the forum had 1,657 threads in the first iteration, and 2,212 in the second.

2.2 *Assignments*

All assignments were submitted online, and graded with calibrated peer assessment. Some assignments asked students to create physical artifacts like paper prototypes and upload photographs of their work.

Each assignment included a rubric that described assessment criteria (Andrade 2005). Rubrics comprised guiding questions or dimensions that student work was graded on, and gradations of quality for each dimension, from poor to excellent. Rubrics were released with the assignment, so students could refer to them while

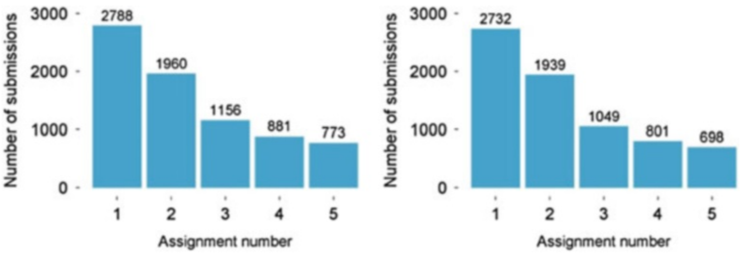


Fig. 3 Number of students who submitted each assignment

working. Table 1a, b shows a part of the rubric for the User Testing assignment, another rubric is shown in Table 3.[2]

Peers assessed using the rubric, and students were informed that peers could see all submitted work while grading. Students could also share their peers' work via class forums after grading was complete and staff used examples of student work in class announcements and lectures. Students could optionally mark their submissions as private to prevent such sharing outside the peer assessment system: over both iterations combined, 13.5 % of students chose to do so.

All assignments and rubrics were based on corresponding materials from the introductory HCI class at Stanford.[3] The in-person Stanford class uses self assessment and staff grading, but not peer assessment.

2.3 Peer Assessment

Assessment used Calibrated Peer Review (Carlson and Berry 2003). Calibrated peer review helps students learn to grade by first practicing grading on sample submissions.

Immediately after each submission deadline, staff evaluated about a dozen submissions—eight were used to train students; the rest were used to estimate accuracy of assessment. The next day, peer assessment opened for students who submitted assignments. Students had 4 days to complete peer assessment.

Peer grading for each assignment had two phases: calibration and assessment. During the first, calibration, phase, students see the staff grade for a submission they grade, along with an explanation. If the student and staff grades are close, students move to the assessment phase. Otherwise, students grade another staff-graded assignment. This process is repeated until student and staff grades match closely, with up to five such training assignments. After five submissions, students moved to the assessment phase regardless of how well they matched staff grades.

[2] All assessment materials are also available in full at http://hci.st/assess

[3] https://cs147.stanford.edu/

Table 1a A fragment of the original rubric for the last assignment

Guiding questions	Bare minimum	Satisfactory effort and performance	Above and beyond
		. . .	
Alternate redesign—Extra credit. Have you created a fully functional alternate prototype?	0: No URL to functional prototype	3: URL present, but prototype only partially functional	5: URL present, Alternative prototype is complete
User testing. Photographs—extra credit. Did you submit photos from all three user testing sessions?	0: No photographs were uploaded	3: Some photographs were uploaded (but less than 3), OR photos don't show an interesting moment in the experiment (e.g. photograph of participant signing consent form is not an interesting photo)	5: At least three photographs are uploaded and all photographs show interesting moments in the evaluation. Photos have meaningful captions
		. . .	

Table 1b Fragment of revised rubric for the same questions

Category	Unsatisfactory	Bare minimum	Satisfactory effort and performa	Above and beyond
		. . .		
Extra credit: Electronic prototype of redesign	0: No URL to functional prototype	1: The prototype is incomplete and barely interactive	3: The prototype is somewhat interactive, but not ready for user testing	5: The alternative prototype is fully interactive and ready for user testing
Photos/ sketches	0: No photographs were submitted that showed interesting moments in the user testing process	1: One photograph was submitted that showed an interesting moment in the user testing process	3: Two photographs were submitted that showed interesting moments in the user testing process	5: Three or more photographs were submitted that showed interesting moments in the user testing process
		. . .		

Only two of six questions are shown, the rest are above and below these (shown as *ellipses*)
The new rubric uses *categories* instead of guiding questions, introduces a new column for completely missing and unsatisfactory work, and uses a parallel sentence structure

Then, students assessed five peer submissions. Unbeknownst to the students, one submission was also graded by staff to provide a measure of assessment accuracy. By symmetry, this means that at least four randomly-selected raters saw each student's submission, and that each student saw one staff-assessed submission per assignment. Immediately after assessing peers, students assessed their own work. Self assessment and peer assessment used identical interfaces.

Time spent on assessment varied by assignment. Depending on assignment, 75 % of assessments were completed in less than 9.5 min to 17.3 min. On the median assignment, 75 % of assessments took less than 13.1 min.

One pedagogical goal of the class was to have students understand and have some influence on their grades. At the same time, we didn't want to reward dishonesty or delusions. To balance these goals, when the self-assessed score and the median peer score differed by less than 5 %, the student got the higher score. If the difference was larger, the student received the median peer-assessed score. This policy acknowledges 5 % to be a margin of error and gives the student the benefit of doubt. Peer grades were anonymous; students saw all rater-assigned scores, but not raters' identities. Similarly, submitters' names were not shown to raters during assessment, i.e. the assessment system was double-blind.

Because assignments built on each other, it was especially important to get timely feedback. Grades and feedback were released 4 days after the submission deadline (the subsequent assignment was due at least 3 days after students received feedback). Students who didn't complete either the self assessment or peer assessment by grade-release time were penalized 20 % of the assignment grade. Students were allowed to assess more than five submissions if they wanted to (Fig. 7 shows the distribution of assessments completed). These additional submissions were also chosen randomly, exactly like the first five submissions.

3 How Accurate Was Peer Assessment?

3.1 Methods

To establish a ground-truth comparison of self and staff grades, each assignment included 4–10 staff-graded submissions in the peer assessment pool (these were randomly selected). Across both iterations, staff graded 99 ground-truth submissions. Each student graded at least one ground-truth submission per assignment; a ground-truth assignment had a median of 160 assessments. (Some students graded more than one ground-truth submission per assignment because the system would give them a fresh ground-truth assignment when they logged-out without finishing assessment and returned to the website after a long time).

This paper's grading procedure assigns the median grade from a small number of randomly selected peers (e.g. 4–5). We evaluated the accuracy of this grading process using the 99 assignments with a staff grade. To simulate the median-grade approach, we randomly sampled (with replacement) five student assessments for each ground-truth submission, and compared the sample's median to the staff grade.[4] We present results for 1,000 samples of five assessments per submission.

[4] Staff comprised graduate students from Stanford. The second iteration had Community TAs chosen among top-performing students in the previous iteration in addition to Stanford staff.

This sampling method is essentially a bootstrapped statistical analysis (Efron and Tibshirani 1993). It allows staff to only evaluate a small set of randomly selected submissions, and still provides an estimate for every peer-rater's agreement with their grade (since all peers see at least one staff-graded submission.) Repeatedly sampling five grades from the pool of peer grades provides an approximate distribution of agreement between staff and peer grades.

We also compared students' self grade with their median peer grade to measure whether students rate themselves differently than their peers.

To enable comparisons, we present results for both iterations separately. The second iteration of the course had grading rubrics improved using data from the first iteration (discussed in Sect. 6.1). The general similarity in accuracy across both iterations (with improvements in the second) suggests that the peer assessment process produces robust results. The second iteration also allowed students to submit assignments in Spanish. For consistency, our analysis does not include those submissions.

At the end of the class, students were invited to participate in a survey; 3,550 students participated in all. Participation was voluntary, students were not compensated, and the survey did not count towards course credit.

3.2 Results: Grading agreement

Here, we present percentage differences between peer and staff grades (summarized in Table 2). Most assignments in this class were out of 35 points. Therefore, a 5 % difference represents 1.5 points (grades could only be awarded in multiples of half a point).

For the first iteration, 34.0 % of submissions had a median peer grade within 5 % of the staff grade, and 56.9 % within 10 % (Fig. 4). The second iteration improved to 42.9 % within 5 % of the staff grade, and 65.5 % within 10 %. In the first iteration of the class, 48.2 % of samples had a peer median lower than staff grade, 40.2 % had it higher. The second iteration had 36 % of samples had a peer median lower than staff grade, 46.4 % had it higher. Students tended to get better at grading over time (See Sect. 3.8).

In the first iteration of the class, 28.7 % of submissions had their median peer grade within 5 % of the self-assessed grade, and 44.9 % within 10 % (Fig. 5). The median submission had a self grade 6 % higher than the median peer grade. In the second iteration, 24.0 % of submissions had their median peer grade within 5 % of the self-assessed grade, 40.63 % had the median peer-grade within 10 %. The median submission had a self-grade 7.5 % higher than the median peer grade. (We discuss possible re sons for this lowered agreement in Sect. 6.3.)

Table 2 Summary of grade agreement

Metric	Iteration 1 (%)	Iteration 2 (%)
Peer-staff agreement (within 5 %)	34.0	42.9
Peer-staff agreement (within 10 %)	56.9	65.5
Peer < Staff	48.2	36.0
Peer > Staff	40.2	46.4
Peer-self agreement (within 5 %)	28.7	24.0
Peer-self agreement (within 10 %)	44.9	40.6

In the second iteration of the class, peer-staff agreement increased, while peer-self agreement decreased

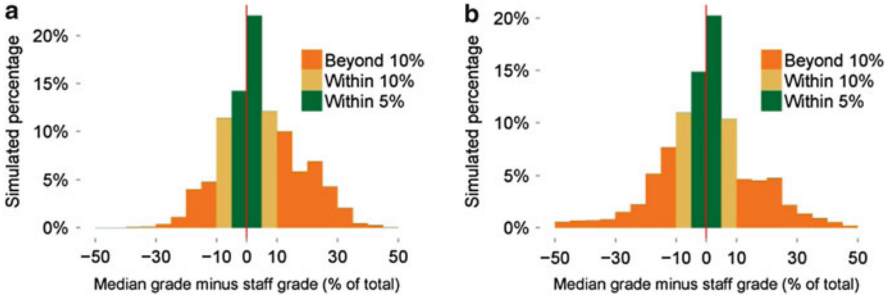


Fig. 4 Accuracy of peer assessment for submissions that were graded independently by teaching staff and peer assessors (all five assignments). Graph accuracy of random sample of 5 graders against staff. (**a**) Iteration 1: 34.0 % of samples within 5 % of the staff grade, and 56.9 % within 10 %, (**b**) Iteration 2: 42.0 % of samples within 5 % of the staff grade, and 65 % within 10 %

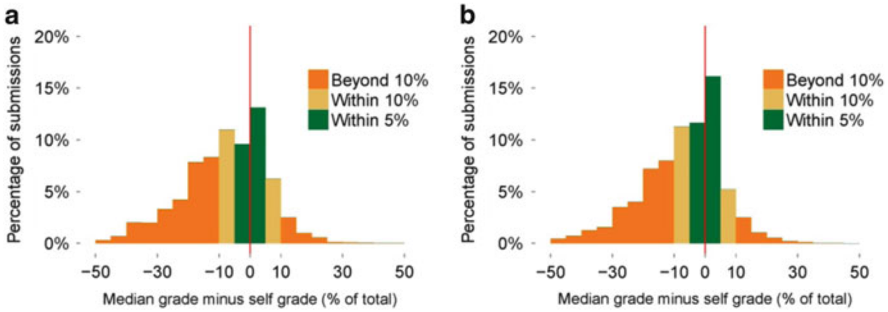


Fig. 5 (**a**) Comparison of median peer grades against self grades. In the first iteration 28.7 % of such samples were within 5 % of the staff grade, and 44.9 % within 10 %. (**b**) Same graph for second iteration of the class. 24.0 % of such samples were within 5 % of the staff grade, and 40.63 % within 10 %

3.3 Results: Grading Agreement Between Staff

The first two iterations of the class had only one staff member grading each ground-truth submission. To get an idea of how well staff grades agree amongst themselves, in the third iteration of the class we asked multiple staff members to rate each submission.

Submissions were randomly assigned to three staff members (there are six staff members in all). Staff rated 50 submissions over the course.

For these submissions, the average disagreement between staff raters (defined as the median difference between a staff grade, and the mean staff grade) was 6.7 %. 28 % of submissions had all staff grades within 5 % of the assignment grade, and 42 % within 10 %. In contrast, over the second iteration of the class, the average disagreement between peer raters was 25.0 %. Only 4.0 % of submissions had all peer grades agreeing within 5 %, and 16.9 % within 10 %.

These results suggest that correlation amongst staff grades is many times higher than agreement amongst peer raters. They also suggest that aggregating peer grades leads to a remarkable increase in agreement with staff grades (Sect. 3.2).

Staff differences in grading were usually due to differing judgments or interpretation. For example, an early assignment asked students to create storyboards of user needs without constraining to a particular design. Staff members differed in how constraining they thought storyboards were.

Such differences suggest the inherent limitations of independent assessment via rubrics due to differences in judgment. Consensus-based mechanisms that encourage sharing perspectives may improve agreement (Amabile 1982).

3.4 Comparison to In-Person Classes

These accuracy numbers also compare well to accuracy in in-person classes. The Fall 2012 version of the in-person class (cs147) that this class is based on used self assessment, but not peer assessment. The in-person class had 32.8 % of submissions with a self grade within 5 % of staff grade, and 60.8 % of submissions within 10 % (Fig. 6).

3.5 Results: Student Reactions

Student reactions to the peer assessment system were generally positive, and 20 % of students completed more peer assessments than the class required them to (Fig. 7). We infer from this that students found rating their peers valuable or enjoyable, and/or they believed it would help their peers.

42 % of students cited seeing other students' work as the biggest benefit of peer assessment, 31 % reported learning how to communicate their ideas as a benefit. Students reported both self assessment and peer assessment to be valuable, and that they played different roles. Evaluating peers was useful for inspiration and to see other perspectives. Self assessment provided students an opportunity to look at their own work again, and encouraged comparing it with others' work they had assessed. It was also useful for identifying mistakes and reflection (Fig. 8). Overall, students

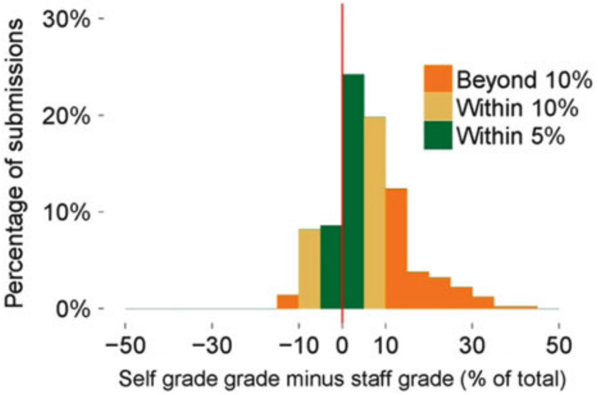


Fig. 6 Agreement of self and staff grades in an in-person class

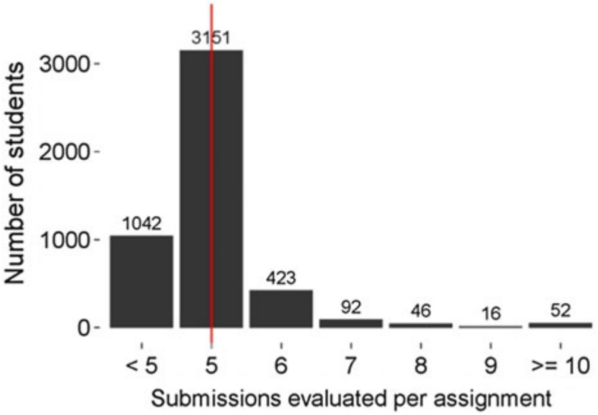
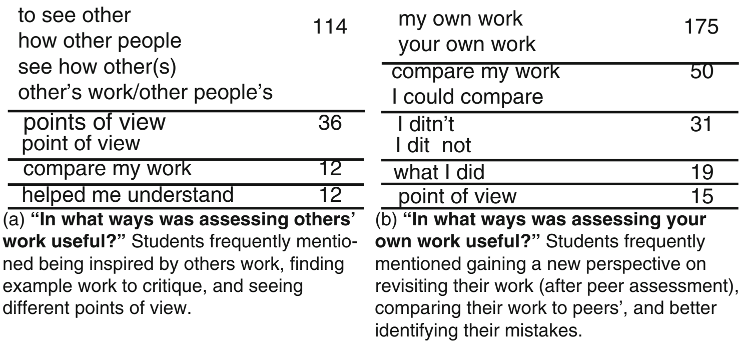


Fig. 7 Average number of submissions assessed per assignment (both iterations). Students were required to assess five, and 20 % of students evaluated more than required



to see other how other people see how other(s) other's work/other people's	114
points of view point of view	36
compare my work	12
helped me understand	12

(a) **"In what ways was assessing others' work useful?"** Students frequently mentioned being inspired by others work, finding example work to critique, and seeing different points of view.

my own work your own work	175
compare my work I could compare	50
I ditn't I dit not	31
what I did	19
point of view	15

(b) **"In what ways was assessing your own work useful?"** Students frequently mentioned gaining a new perspective on revisiting their work (after peer assessment), comparing their work to peers', and better identifying their mistakes.

Fig. 8 The most frequent trigrams (three word phrases) in students' self-report (over both iterations of class): students reported both peer and self assessment to be valuable for different reasons

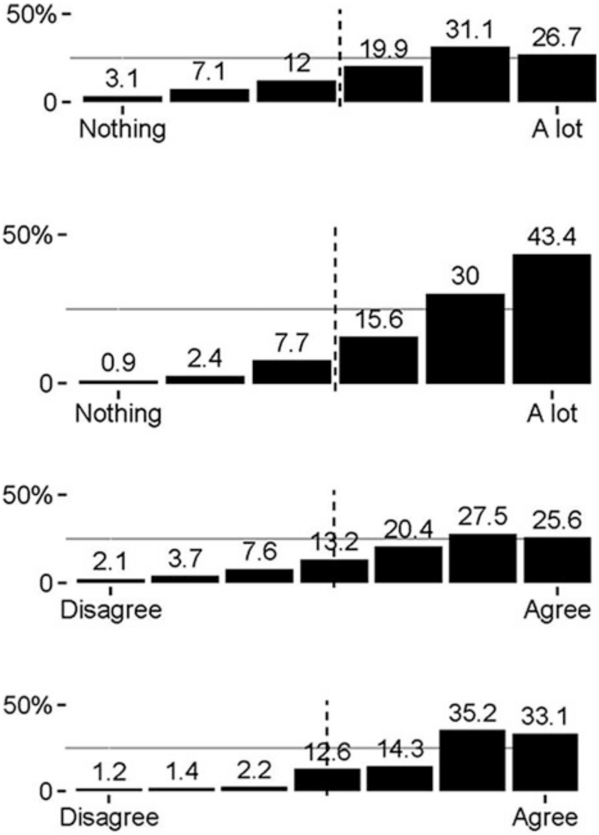


Fig. 9 (continued)

reported learning more by assessing their peers than by assessing themselves: mean ratings were 4.97 and 4.51 respectively for peer and self assessment (6-point Likert scale, 6: "agree strongly (sufficient effort)"), on a Mann-Whitney U-test $U = 580, 562$, $p < 0.001$.

However, students also reported that they felt their peers put in less effort into peer assessment than they did (Fig. 9). On a Mann-Whitney U-test, mean ratings were 4.57 for peer-effort and 5.46 for their own effort (6-point Likert scale, 6: "learnt a lot"), $U = 610, 728$, $p < 0.001$. Reasons for this bias are probably similar to the illusory superiority effect (Ehrlinger et al. 2008). Designing peer assessment interfaces that emphasize reciprocity and minimize this bias remains future work.

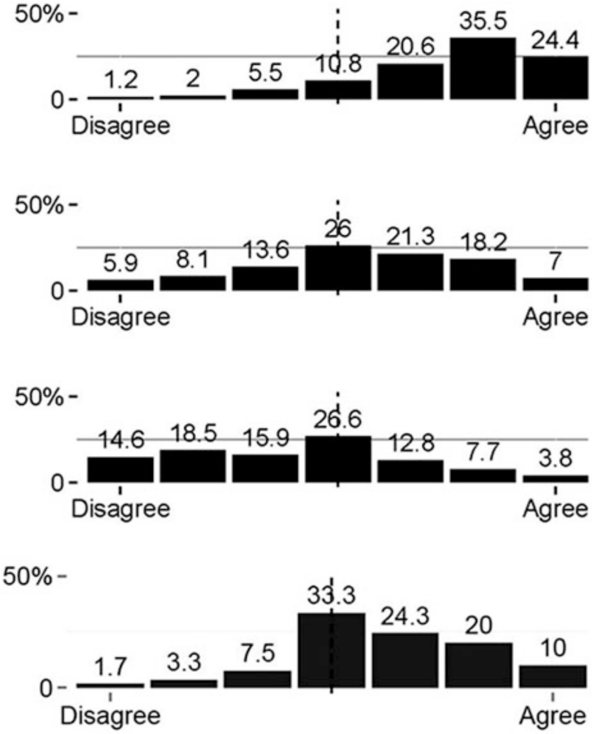


Fig. 9 End course survey results (n = 3,550) about student perceptions on peer assessment. Students reported learning from assessing others' work than their own, and putting effort into grading fairly

3.6 *Does a Different Weighting of Peer Grades Help?*

Using the median of peer grades is simple, easily explainable, and robust to outliers. Would a different weighting of peer grades more accurately mimic staff grades?

Method To find the best linear combination of weights, we built a linear regression on the staff grade with five peer grades in increasing order as the predictors, and with no intercept. This regression seeks weights on peer grades that maximally predict the staff grade.

Results The best linear regression doesn't materially improve accuracy. The linear model weighted the five peer grades from lowest to highest at 15.6 %, 13.6 %, 21.3 %, 27.6 %, 18.3 %. Holding out 10 % of ground truth grades, and testing on samples drawn from them, the regression model yields an accuracy of 35.8 % of samples within 5 %, and 58.8 % within 10 %. In contrast, using the median yields an accuracy of 35 % of samples within 5 %, and 58.7 % within 10 %.

Similarly, the arithmetic mean, geometric mean, and a clipped arithmetic mean (that only considers the middle three grades) all do worse than the median. In addition, errors are approximately evenly spread across the median, so adding a constant correction term to the median grade does not significantly improve accuracy either.

In summary, the simple median strategy seems to be surprisingly effective at identifying the most plausible grade. Is this accuracy sufficient? For a class with letter grades, greater accuracy is needed (because currently about 40 % of assignments are a full letter grade away). However, a student's grade for the entire course is generally more accurate due to positive and negative errors canceling out. Using repeated sampling, we estimate more than 75 % of students got a course grade within 5 % of staff grade (assuming grades in different assignments are uncorrelated). Consequently, for a pass/fail class (such as many current MOOCs, including ours), this accuracy is sufficient for the vast majority of students. We estimate that less than 45 students (approx. 6 %) were affected by grading errors in each iteration of the class.

3.7 Would More Raters Help?

Increasing the number of raters per submission helps accuracy, but quickly yields diminishing returns (Fig. 10). A large number of students rated staff-graded assignments. These allow us to simulate the effect of having more raters. Increasing the number of assessments per submission from 5 to 11 increases the number of assignments that were graded within 5 % of the staff grade by 3.8 %, and those graded within 10 % by 3.6 %. Increasing the number of assessments to an (unreasonable) 101 per submission increases the number of submissions graded within 10 % of the staff grade by 8.1 %.

3.8 Do Students Become Better Graders Over Time?

Agreement of peer grades with staff grades generally increases across the class. This increase is seen both for the class as a whole, and for students who submit all assignments, i.e. excluding students that drop out. This suggests that, regardless of individual differences in perseverance and motivation, familiarity and practice with peer assessment leads to more accurate assessments.

Using the repeated sampling scheme described in Sect. 3.1, five assignments had 26.4 %, 36.2 %, 36.9 %, 43.9 %, and 36.8 % of submissions estimated within 5 % of the staff grade. Within a 10 % range, the assignments had respectively 49.1 %, 53.6 %, 60.9 %, 68.5 %, and 64.3 % within 10 % (Fig. 11a). If we only consider raters that finished the class (and exclude those that dropped out), we see that staff agreement increases as well. The five assignments in order had 23.7 %, 29.4 %, 38.4 %, 39.5 %, 37.1 % within 5 % of staff, and 47.4 %, 63.8 %, 61.8 %, 63.3 %, 64.2 % (Fig. 11b). Note that both these numbers are based on repeated sampling

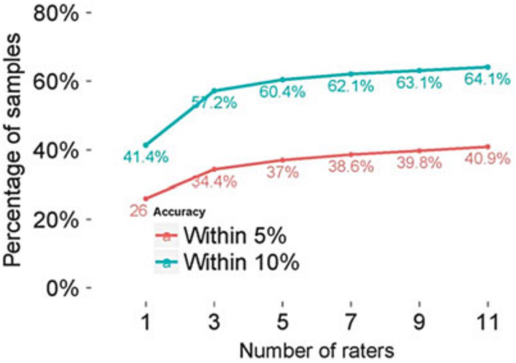


Fig. 10 Increasing the number of raters quickly yields diminishing returns

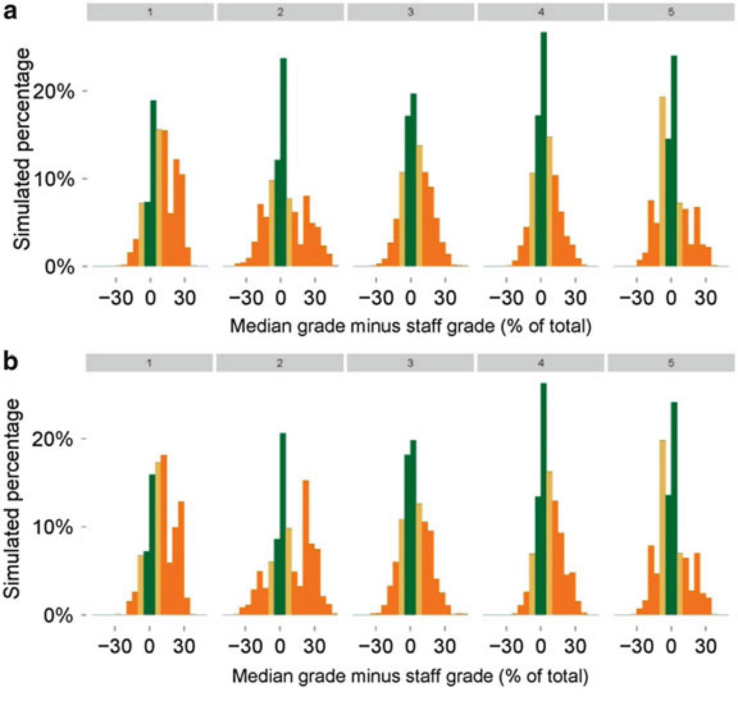


Fig. 11 Agreement of median peer grades and staff grades across different assignments. (These agreement distributions are more susceptible to variations in staff grades for a particular submission because they are based on repeated sampling from a smaller number of staff-graded assignments.). (**a**) All raters, (**b**) only raters who finished the class

from a smaller number of staff-graded assignments. As such, they are more susceptible to variations in staff grades for a particular submission.

3.9 What Is the Right Granularity of Grades?

Sections 3.3 and 3.4 shows that the grading agreement between staff members, and between staff and students in an in-person class are similar. These differences may approximately represent the smallest discernible differences in quality.

Recall that a 5 % difference in grades is 1.5 points in a 35 point assignment, i.e., three times a "just-noticeable" difference in quality (0.5 points, the minimum granularity of grades). Indeed, the in-person version of the class adopted the current 35 point grading scheme (replacing its 100 point scheme from prior years) to better balance accuracy with meaningful differences in quality.

3.10 "Patriotic" Grading?

On average, raters grade students from their own country 3.6 % higher than those from other countries: $t(27,067) = 3.98$, $p < 0.001$. This effect is consistent when the raters and submitters from the largest student enrollment (United States) are removed, but is smaller (the mean difference drops to 1.98 %, $t(12,863) = 2.0$, $p < 0.05$). We remind the reader that grading was double-blind, so raters did not see the names of submitters.

We see four possible explanations for this "patriotism" bias. One is that raters better understood applications designed for their local environment and so rated them more highly. Another is that raters were "voting" for applications that they inferred were from the same country—by the content of the application or the style of the presentation. A third possible explanation is that different cultures consider differing attributes of design, as in Kim and Hinds' work on cross-cultural creativity (Kim and Hinds 2012). Finally, assessment materials may be understood by students in different countries in subtly different ways. Understanding this effect remains future work.

4 Providing Students Feedback on Grading Accuracy Improves Subsequent Performance

So far, this paper has characterized the accuracy of large-scale calibrated peer assessment. This section explores a feedback intervention to improve graders' accuracy. Prior work has demonstrated that feedback improves the quality of crowd work (Dow et al. 2012), but can it help raters overcome their (possibly unintentional) grading bias? This section describes an experiment that provided students feedback whether they were grading either "too high," "too low," or "just right," based on how well their grade agreed with staff grades for the previous assignment. We hypothesized that providing students grading feedback would help improve accuracy. We conducted a controlled experiment on the course website that measured the impact of this feedback on accuracy.

4.1 Participants and Setup

We randomly sampled 756 participants from students who had completed the second assignment of the second iteration of the class.

The between-subjects experimental setup had two conditions: a no-feedback control condition where students received no feedback on the accuracy of their grading, and a feedback condition that provided feedback on their grading bias: too high, too low, or just right (Fig. 12).

To generate bias feedback, the system compared the participant's rating and the staff rating of the previous assignment's ground-truth submission.

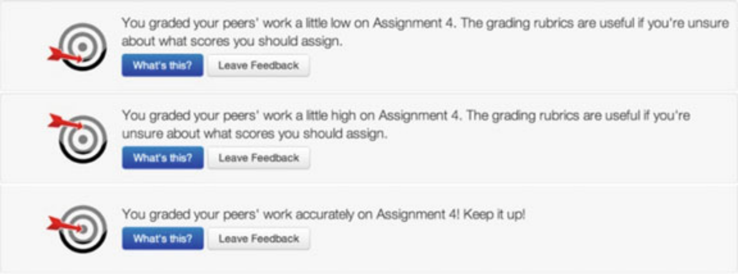
If the rating differed by more than 10 %, then feedback was shown as too high/too low; otherwise the feedback was "just right." In the feedback condition, high/low/just right feedback appeared just above the grading sheet (Fig. 13). In the control condition this space was blank.

4.2 Results: Feedback Reduces Grading Errors

Using a repeated sampling analysis (as in Sect. 3), we compared staff grades to a random sampling of peer grades from participants in each condition for ground-truth submissions. The difference between the median peer grade obtained by sampling from the feedback condition and the staff-grade was 6.77 %, compared to 7.74 % in the no-feedback condition (Fig. 14). We built a linear model that predicts grading error using experimental condition as fixed effect, and each rater as a fixed-intercept random effect.

The effect of the presence of feedback is significant: $t(4{,}998) = \_3.38$, $p < 0.01$. 4.4 % more samples in the feedback condition obtained a grade within 5 % of the staff grade than those without feedback. Notably, 55 students left comments expressing their appreciation or receptiveness to this feedback; none expressed resentment.

This experiment tested the mere presence of accuracy feedback. Future work can assess the effects of richer feedback, such as the amount of bias or change over time. It can also explore bi-directional communication between the submitter and the assessor.

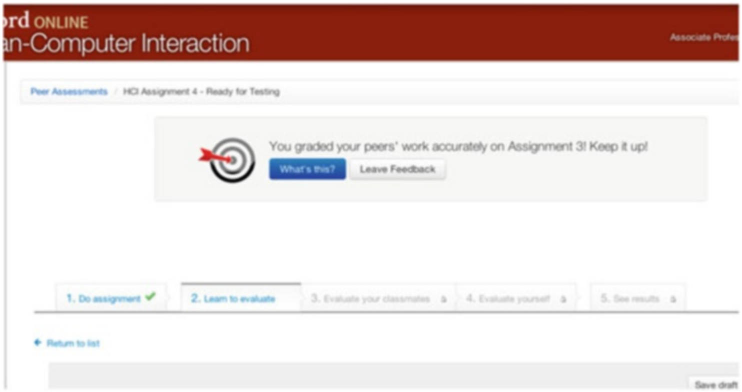


You graded your peers' work a little low on Assignment 4. The grading rubrics are useful if you're unsure about what scores you should assign.

You graded your peers' work a little high on Assignment 4. The grading rubrics are useful if you're unsure about what scores you should assign.

You graded your peers' work accurately on Assignment 4! Keep it up!

Fig. 12 In the feedback condition, students received feedback about how well they were grading



You graded your peers' work accurately on Assignment 3! Keep it up!

Fig. 13 Students improved grading when provided accuracy feedback

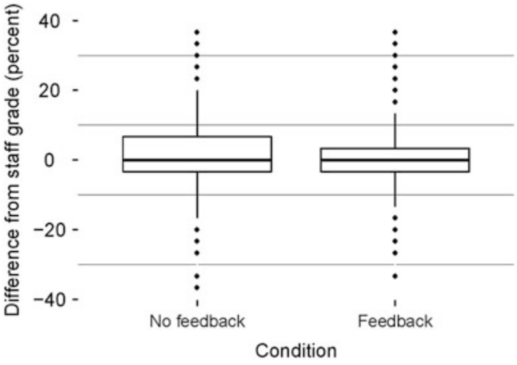


Fig. 14 Feedback on grading accuracy reduced the overall error in assessment and made the range of errors smaller

5 Providing Personalized, Qualitative Feedback on Assignments

Accurate, actionable feedback helps students improve their work (Nicol and Macfarlane-Dick 2006; Boud 2000). Actionable feedback is most useful if it is personalized, and targets the student's recent work (Gallien and Oomen-Early 2008).

Rubrics provide feedback through quality gradations for each dimension. For instance, students can look at rubric items they did poorly on to find areas for improvement. However, using rubric item scores as feedback has two important limitations. First, students must reflect on why they did poorly on some topic. Unfortunately, these are often topics the student understood poorly in the first place. Second, rubrics only point out areas for improvement, not how to improve.

Can peers provide actionable, personalized feedback? We introduce one method that captures broadly applicable yet specific feedback in short snippets. On the assessment form, raters select which snippets apply to the current assignment, and optionally fill in a "because . . ." prompt (Fig. 15). Inspired by (Dow et al. 2010), we call the result "fortune-cookie feedback" for its brevity and general applicability. Figure 16 shows some examples.

5.1 Methods: Creating Fortune Cookies

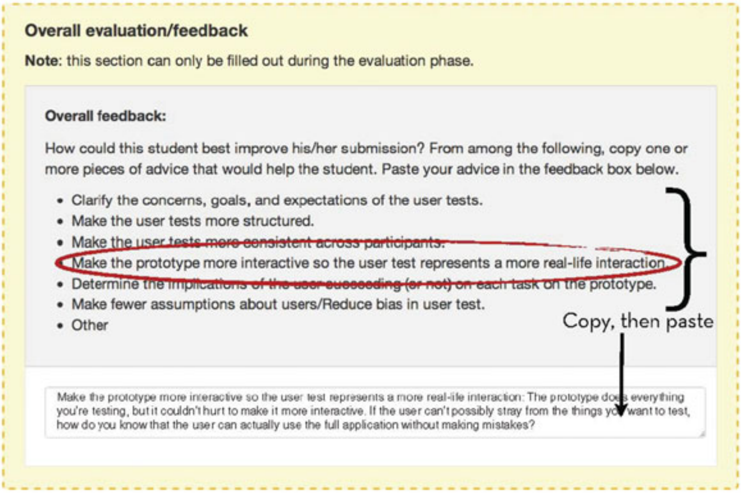
We wanted fortune cookies to help with two common patterns in student performance.

First, we wanted to find places where committed students did poorly, and retroactively generate useful advice. To find committed students (and keep the number of submissions manageable), we restricted our analysis to students whose initial performance was above the 90th percentile. Then, we compared students who subsequently got the median grade to those that got grades above the 90th percentile.

Second, we wanted to highlight strategies that students used to improve. We compared submissions from students that improved their performance from median grade to excellent (above 90th percentile) on a subsequent assignment against those that obtained median grades on both assignments.

We then manually wrote feedback for each submission separately. For each assignment, we looked at an average of 15 submissions, five each that showed improved, reduced and steady performance. Combining related feedback from different submissions led to our final list of warning signs and improvement strategies. Creating fortune cookies took a teaching assistant 3–4 h per assignment.

We created fortune cookies based on submissions in the first iteration of the class, and tested them in the second iteration. As the last question on the grading sheet, we asked "which of these suggestions would improve this submission the



Overall evaluation/feedback

Note: this section can only be filled out during the evaluation phase.

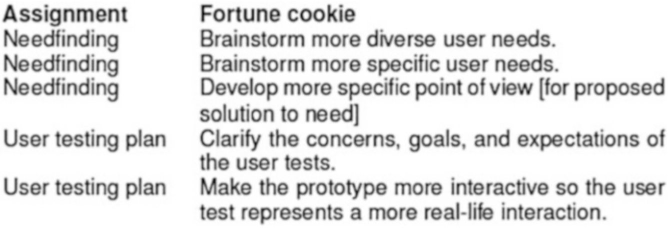
Overall feedback:

How could this student best improve his/her submission? From among the following, copy one or more pieces of advice that would help the student. Paste your advice in the feedback box below.

- Clarify the concerns, goals, and expectations of the user tests.
- Make the user tests more structured.
- Make the user tests more consistent across participants.
- Make the prototype more interactive so the user test represents a more real-life interaction.
- Determine the implications of the user succeeding (or not) on each task on the prototype.
- Make fewer assumptions about users/Reduce bias in user test.
- Other

Make the prototype more interactive so the user test represents a more real-life interaction: The prototype does everything you're testing, but it couldn't hurt to make it more interactive. If the user can't possibly stray from the things you want to test, how do you know that the user can actually use the full application without making mistakes?

Fig. 15 Students copied snippets of feedback (fortune cookies), pasted them in a textbox and optionally added an explanation



Assignment	Fortune cookie
Needfinding	Brainstorm more diverse user needs.
Needfinding	Brainstorm more specific user needs.
Needfinding	Develop more specific point of view [for proposed solution to need]
User testing plan	Clarify the concerns, goals, and expectations of the user tests.
User testing plan	Make the prototype more interactive so the user test represents a more real-life interaction.

Fig. 16 Example "fortune cookie" feedback

most?" Students copied appropriate fortune cookies from a list and pasted it in to a textbox below. Students were not required to use these snippets for feedback—they could type in their feedback into the textbox as well.

5.2 *Results: How Well Do Fortune Cookies Work?*

Overall, 36.2 % of assessments included feedback (compared to 36.4 % in the previous iteration without cookies). A chi-square test on the number of assessments that contained feedback suggests that fortune cookies do not encourage more students to leave feedback (_2 = 0.1, p = 0.75). Because submissions were assessed by multiple students, 94.9 % of submissions received at least one piece of written feedback (compared to 83 % without cookies); 67.2 % of students received at least one "fortune cookie"; and 65 % of students received one or more fortune-cookies with a "because..." explanation (Fig. 17).

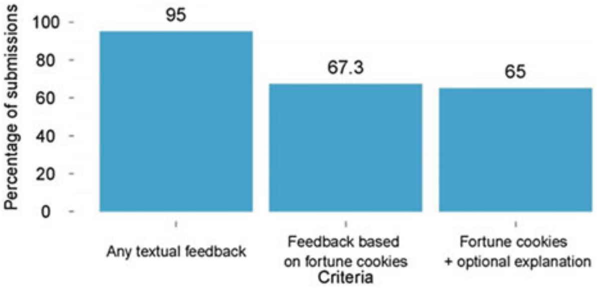


Fig. 17 Most students received at least one piece of textual feedback. Most fortune cookie feedback was personalized

Raters typed the same amount of feedback whether or not an assignment contained fortune cookies. If we subtract the text of the cookie itself, there was no significant difference in comment lengths whether or not cookies were used (t(10,673) = 0.44, p > 0.6). If the text is included, comments that used fortune cookies were longer (t(10,673) =3.61, p < 0.05). This suggests that students expend the same amount of effort writing feedback, and using fortune cookies allows this effort to be used to add to the fortune cookie text.

5.3 Discussion

Reusable pre-canned prompts encourage students to direct their effort to providing feedback beyond the cookie text. While we do not demonstrate this improves feedback in the current article, we see three reasons why fortune cookies may provide better quality feedback than non-cued feedback. First, providing raters a list of potential feedback items changes a recall/identification task into a recognition task. This reduces the cost of giving feedback (Anderson and Bower 1972; Nielsen 1994). Second, showing a list of common, assignment-specific problems that the submission could have potentially reduces inhibition, and encourages peers to think critically (Galinsky and Moskowitz 2000). Third, because fortune cookies sometimes used terminology learned in class, they may have triggered cued-recall of these concepts (Little and Bjork 2012), leading to more conceptual comments.

Future research could investigate this idea further. In addition, it could also explore if fortune cookies confer differential benefits to different students and how best to leverage this.

6 Overall Discussion

6.1 Using Data to Improve Assessment Materials

Iterative design often pays big dividends (Nielsen 1993), and assessment systems are no exception. The large scale of online classes allows data-driven iterative

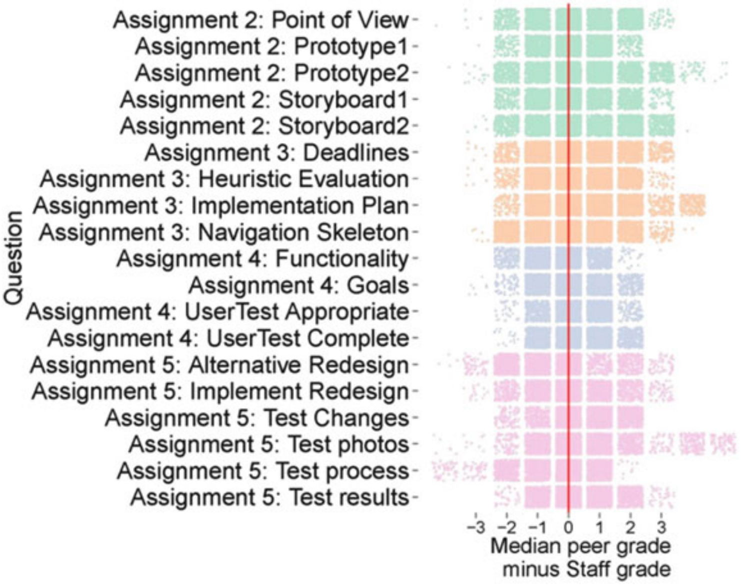


Fig. 18 Comparing variance of rubric items can help teaching staff find areas that may need improvement. For example, this figure shows the variance for four assignments of the HCI course between staff grade and median peer grade. A narrow, dense band indicates higher agreement. For example, Assignment 4 (*blue*) has generally higher agreement

improvements of classroom materials in ways that small classes may not. Below, we describe some data-driven changes we made.

One can use low rater agreement to find questions that might benefit from revisions. We found that peer and staff raters agreed far more on some questions than others (Fig. 18), and that questions with low staff agreement also had low peer agreement ($r = 0.97$, $t(24) = 19.9$, $p < 0.05$). We reviewed such questions and revised them with feedback from the forum. Most rubric revisions centered around making rubrics more easily readable.

Improving Readability Some rubrics sometimes used a non-parallel grammatical structure across sentences. This is not uncommon: even examples in prior work on using rubrics suffer from this problem [e.g. (Andrade 2005)]. We hypothesized that using a parallel sentence structure would better help students understand conceptual differences (Markman and Gentner 1993). We found that rubric items with parallel sentence structure in the first iteration had lower disagreement scores ($F(1, 39) = 2.07$, $p < 0.05$) (Fig. 19). We revised all rubrics to use parallel sentence structure. We also made other changes to improve readability, such as removing duplicate information from assignments, and splitting up rubric items that asked students to make a complex judgment (e.g. “Is the prototype complete and functional?” to “Is the prototype complete?” and “Is the prototype functional?”).

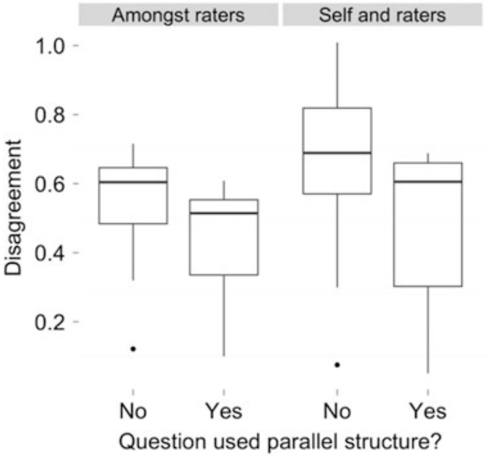


Fig. 19 In iteration 1, questions with parallel structure had lesser disagreement, both amongst peer graders, and between the median grade and the self-assessed grade. We changed all assignments to use parallel structure across rubric items

Word Choice Although the rubrics had been revised for 3 years in the in-person class, many forum posts asked for clarifications of ambiguous words. Words like “trivial”, “interesting”, “functional”, and “shoddy” may be correctly interpreted by the on-campus student with a lot of shared context, but are ambiguous online. The revised version replaces these words with more specific ones (which may help on-campus students as well).

The revised rubrics were used in the second iteration of the class. Overall, the peer-staff agreement was 2.5 % higher than the previous iteration.

6.2 *Going Beyond Pass/Fail*

Peer assessment as described in this paper works reasonably for a pass/fail class. How might peer assessment be used in classes that award more fine-grained grades? Beyond having iteratively-refined rubrics (as above), one possibility is to involve community TAs in grading submissions that are estimated to have low grading accuracy (e.g. with large differences between self and peer grades). In addition, our early experiments suggest that greater accuracy is possible by weighting different raters’ grades differently, an important topic for future work. Lastly, our experiments suggest that machine-grading approaches (such as those for essay grading) may be combined with peer assessment to provide accurate assessment.

6.3 *Inflating Self-Grades and Other Gaming*

Many types of cheating are currently possible and unchecked in online classes. For example, someone else could simply take a course on your behalf. To the extent that

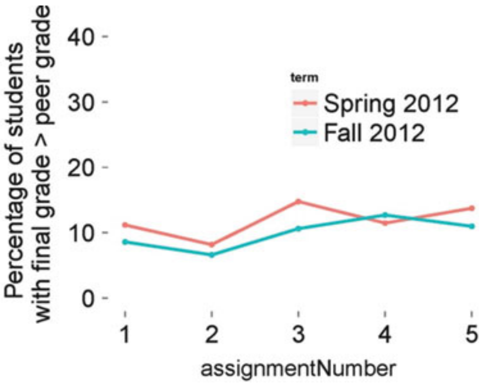


Fig. 20 Students in the second (Fall 2012) iteration of the class reported a self grade > 5 % higher than peer grade more frequently, and so got their self grade less frequently

participation in the online classroom is based on intrinsic motivations (such as a desire to learn), students rarely blatantly cheat (Mazar et al. 2008). (Anecdotally, several instructors in early online classes have reported that some students appear to be cheating, but that it doesn't currently appear to be widespread.)

To date, large-scale online classes, including our own, have primarily emphasized learning, rather than certification (Widom 2012). Students do not receive much in the way of credit. (Though on social media like Facebook and LinkedIn, some students report having "attended" Stanford.) Still, some students probably attempted to game their score by strategically over-reporting their grade (Fig. 20). As online classes count for more benefits, such gaming may increase.

Gaming also has a silver lining. A valuable skill for success is the theory of mind to intuit how others perceive one's performance (Boud 1995), and gaming may help students develop this skill. Cheating may also arise if the value of officially recorded performance in these classes increases [e.g. (Kurhila 2012; Lewin 2013b)]. To combat this, several organizations have proposed solutions like in-person testing facilities [e.g. (Lewin 2012b)], or verified-identity certification (Lewin 2013d). Others remain focused on teaching for students who want to learn (Widom 2012).

6.4 Limitations of Peer Assessment

While peer assessment offers several benefits, it also has limitations. First, peers and experts (e.g. staff) may interpret work differently (see Appendix 1.2). Such differences are well-known in related fields: Experts and novices both robustly reach consensus about creativity, but their consensual judgments differ from each other (Conti et al. 1996). This may be because novices and experts differ in their tacit understanding of value (Kaufman et al. 2008). Peer assessment addresses this problem by providing raters with expert-made rubrics, but some differences may persist. In addition, independent assessment via rubrics and subsequent aggregation may not assess "controversial" work well.

Second, peer assessment imposes a particular schedule on class, and limits student flexibility. In our class, several students complained in class forums about being unable to complete peer assessments in time. Lastly, while peer assessment works well for the large majority of students, students who receive an unfair assessment may lose motivation. Anecdotally, we have noticed that students are generally satisfied with their overall grade, but are frustrated by inaccurate qualitative feedback from some peers. Addressing these motivational aspects remains future work.

6.5 The Changing Role of Teachers

Peer assessment fundamentally changes the role of staff. When peer assessment provides the primary evaluative function, the staff role shifts to emphasize coaching (Kuebli et al. 2008). Students sometimes believe that teachers grade on personal taste, and focus on currying favor. By contrast, when teachers coach but do not grade, students focus more on conceptual understanding (Perry 1970). Also, providing explicit grading criteria (especially in advance) helps convey to students that grading is fair, consistent, and based on the quality of their work.

Peer assessment also changes how instructors spend their time. When staff assess student work, their effort is focused on doing the grading. By contrast, with peer assessment, the instructor's main task is articulating assessment criteria for others to use. Because of the diversity of submissions, this can be extremely difficult to do a priori. Teachers should plan on revising rubrics as they come across unexpected types of strong and weak work. After revision, these rubrics can scale well for both students and other teachers to use. For online education to blossom, it will be important to teach the teachers best practices for rubric creation, and to create effective design principles and patterns for creating assessments.

While the scale and medium of online education poses new challenges, it also offers new solutions. In key areas, online education encodes pedagogy into software, which increases consistency and supports reuse—and defaults have a powerful impact on behavior (Palen 1999).

The role of teaching staff (TAs) changes too. Instead of spending a majority of their time grading, they spend a large fraction of their time fielding student questions, mentoring students, boosting student morale and autonomous perspective, and making data-driven revisions to class materials.

6.6 The Changing Roles of Students

One of the most remarkable results from our experience was that students reported that assessing others' work was an extremely valuable learning activity. Can online classes provide an avenue not just for peer assessment, but for peer learning as well?

The second iteration introduced Community TAs recruited among students from the first iteration [Armando Fox and David Patterson's Software-as-a-Service online class used a similar program (Fox and Patterson 2012)]. We invited students who did well in class, assessed many submissions voluntarily, and participated actively in class to become Community TAs. Community TAs volunteered their time, and were not paid. Their duties comprised grading assignments, answering student questions, and helping iteratively improve assignments. Five students from across the world participated. Together, community TAs answered 547 questions on the forum, staff (3 local TAs and the instructor) answered 582 questions. In addition to providing factual answers and assignment clarifications, Community TAs also leveraged their personal experience to offer advice and cheerleading.

We hypothesize that Community TAs are effective for the same reasons as undergraduate teaching-assistants at a university (Roberts et al. 1995). First, because community TAs had done well in the class, they possessed enough knowledge to effectively offer information and guidance. Second, because they had taken the class recently, they could easily empathize with issues students faced and also could effectively offer social support.

Massive online classes also offer individual students an opportunity to have large-scale positive impact. For example, when the first assignment of the Spring 2012 class had fewer peer assessments than needed, one student rallied her peers to finish a large number of assessments over a single day (the top ten students assessed an average of 48 submissions: nearly ten times their required number) so that students could get feedback in time. She also participated heavily in the forums, and gathered staff-like respect from her peers.

6.7 The Changing Classroom

The online classroom is distinctly different from its in-person counterpart. Recent research has discovered some of these differences: students in online classrooms are much more diverse both demographically, and in their objectives in taking the class, and platforms make some kinds of data, such as engagement with course material, more plentiful and finer grained, while making other information, such as facial expressions of confusion, completely inaccessible (Breslow et al. 2013).

These differences require rethinking the design of the classroom. For instance, students often have work commitments, and holidays are at different times around the world. This reflects in class scheduling: the first iteration of the class spanned 7 weeks, mirroring the time these topics take in the Stanford course. Although university-like deadlines helped generate interest in online classes (Lewin 2013c), we found that campus-paced deadlines are too rigid online. Consequently, the second iteration spanned 9 weeks to give students more time and flexibility.

While class diversity requires adaptations, it also inspires new opportunities. How can teachers support student leadership and community learning more directly in the online classroom? Again, the design studio offers inspiration (Schön 1985; Pendleton-Jullian 2010). By making not only the results of work, but also the process of creation highly visible, it helps students learn and build awareness through observation (Klemmer et al. 2006). In addition, a studio facilitates dialogue between students, instructors and artifacts that helps students collaboratively learn difficult concepts and solve problems (Schön 1985).

The opportunity here is twofold. First, online learning can be blended with co-located learning. Even though this was a completely online class, students self-organized to meet up in ten locations around the world including London, San Francisco, New York City, Buenos Aires, Aachen (Germany), and Bangladesh.

Second, we can build online experiences that are inspired by the physical studio. By removing the constraints of the physical classroom, online classes have made education accessible to many new kinds of students—the new mother, the full-time professional, and the retiree. Preserving this accessibility, while providing the benefits of the in-person classroom online offer a promising area for future work.

More generally, online education requires us to re-conceptualize what it means to be a student in many ways. One has to do with enrollment and retention (Kizilcec et al. 2013). Typing one's email address into a webpage is not the same as showing up for the first day of a registrar enrolled class. It's more like peeking through the window, and what the large number of signups tell us is that lots of people are curious. How can we convert this curiosity into meaningful learning opportunities for more students?

7 Conclusions and Future Work

This paper described our experiences with the largest use of peer assessment to date. This paper also introduced the "fortune cookie" method for peers to provide each other with qualitative, personalized feedback. We demonstrated that providing students feedback about their rating bias improves subsequent accuracy. There are many exciting opportunities for future work.

First, systems could allocate raters and aggregate their results more intelligently to increase accuracy and decrease work. Crowdsourcing techniques suggest initial steps. After assessment is complete, systems could differentially weight grades based on raters' past performance, for instance, extending approaches like (Ipeirotis et al. 2010). Also, the number of raters could be dynamically assigned to be the minimum required for consensus, extending e.g. (Guo et al. 2012). Furthermore, an algorithm could adaptively select particular raters based on estimated quality, focusing high quality work where it's most needed, as in (Dai et al. 2010). Finally, as with standardized essay grading (Hearst 2000), peers could be used together with automated grading algorithms [such as (Socher et al. 2012; Zaidan and Callison-Burch 2011)]. This hybrid approach can achieve consensus while minimizing duplicated effort. Ideally, these grading schemes should be understandable as well as accurate. Should the system show students how their grade was generated? And if so, how?

Second, current online learning platforms suffer from sensory deprivation relative to a human teacher. They receive final work products, but have no knowledge of students' process. Cognitive tutoring software has shown that attending to students' process can improve learning through personalization—adapting questions, pacing, and guidance (Corbett et al. 2002). Integrating rich learner models with peer assessment offers many exciting opportunities.

Third, physical universities employ many structural levers to keep students motivated and engaged. In our experience, only a quarter of approximately 3,000 students who completed a time-intensive first assignment did all five assignments. Needless to say, at a physical university the completion rate for an equivalent class is much higher. How can online settings provide greater motivation support? Future work could draw both on research on commitment strategies in online communities [e.g. (Kraut and Resnick 2011)] and resources used at physical universities, such as mentoring and orientation courses (Murtaugh et al. 1999). More generally, online learning platforms could benefit students by incorporating known best practices about learning and moving to a more evidence-based approach.

Fourth, peers can help instruction itself. One promising approach is to use social mechanisms to highlight good student work and build connections, such as (Marlow et al. 2013). Another is to leverage peers in physical meet-ups to augment instructor teaching (Cadiz et al. 2000). This approach also creates technology and pedagogy design opportunities for a "flipped" classroom—what should class time look like at a university when students can watch the professor on video? Already, several universities are teaching physical classes augmented with online materials (Martin 2012). How would different roles change with such a model?

Fifth, future work has the potential to tie student work in class to skilled crowd work (Kittur et al. 2013). For instance, students in the HCI class could build prototypes and design websites for clients, or students studying Machine Learning could compete to build predictive models. How can the pedagogical goals of the class be intertwined with potentially productive work? This future work will offer students around the world an opportunity to learn in ways previously impossible.

Acknowledgements We thank Coursera for implementing this peer assessment system, and enabling us to use the data. We thank Sébastien Robaszkiewicz, Joy Kim and our Community TAs for helping revise assignments, assess student submissions, and provide forum support; Nisha Masharani for helping collect data and for designing fortune-cookie feedback; and Sébastien Robaszkiewicz and Julie Fortuna for rating fortune cookies. We thank Greg Little for discussions about online peer assessment and Michael Bernstein for comments on drafts. We thank Jane Manning and colleagues at Stanford for supporting this class's development. We thank our editor Marti Hearst and anonymous reviewers for their valuable feedback and suggestions. Human subjects research was reviewed by the Stanford Institutional Review Board through protocol 25001. We thank all of the students in the HCI online class for their foolhardiness enthusiasm in participating in this experimental class.

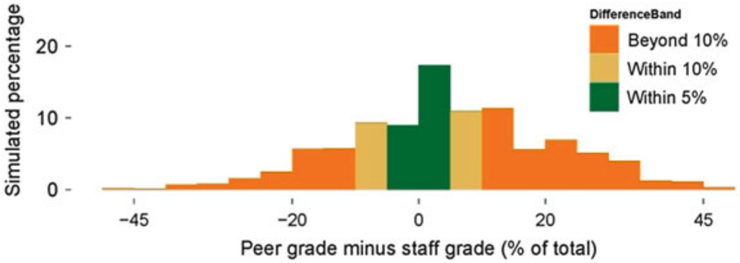
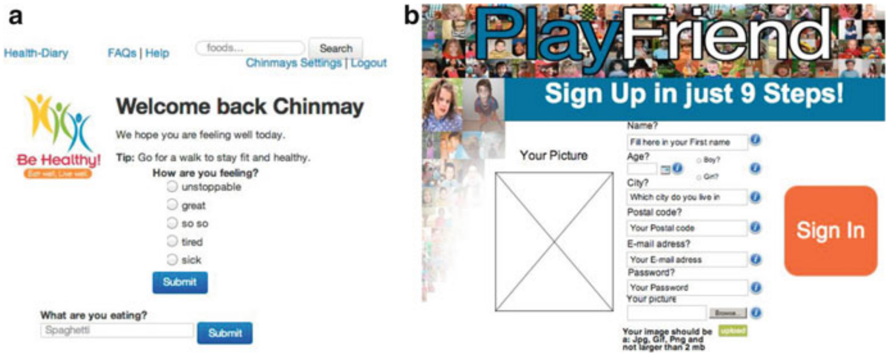


Fig. 21 Agreement of unaggregated peer grades and staff grades. Agreement is much lower than between median peer grades and staff grades



Welcome back Chinmay

We hope you are feeling well today.

Tip: Go for a walk to stay fit and healthy.

How are you feeling?

○ unstoppable
○ great
○ so so
○ tired
○ sick

What are you eating?

Spaghetti

Your Picture

Name?
Fill here in your First name
Age?
Boy?
Girl?
City?
Which city do you live in
Postal code?
Your Postal code
E-mail adress?
Your E-mail adress
Password?
Your Password
Your picture

Your image should be a: Jpg, Gif, Png and not larger than 2 mb

Fig. 22 Student submissions with large differences between staff and peer grades. (**a**) Submission where peers grade higher than staff, (**b**) Submission with staff grade higher than peers

Appendix 1

1.1 Agreement Between Peer Grades and Staff Grades Without Aggregation

Comparing the peer grades (not their medians) with staff grades demonstrates the value of aggregating peer grades (Fig. 21). 26.3 % of grades were within 5 % of staff grades, and 46.7 % within 10 %. (Recall that the median agreement was 42.0 % and 65.5 %, respectively).

1.2 Grading Differences

1.2.1 Where Peers Graded Higher

Figure 22a shows an application a student created as “an interactive website which helps people tracking their eating behavior and overall-feeling, to find and be able

Table 3 Rubric for "Ready for Testing" assignment

Category	Unsatisfactory	Bare minimum	Satisfactory effort and performance	Above and beyond
List of changes	0: No changes or completely irrelevant changes	1: The student only identified a few changes from the heuristic evaluation feedback and a large amount of feedback is ignored in the new prototype; the new prototype has some HE violations	3: Many of the simpler suggested changes were made, but some of the more complex or difficult issues were not addressed; the new prototype does not have any obvious HE violations	5: The user made several insightful and specific changes based on the heuristic evaluation feedback. It is hard to find any HE violations at all in the new prototype
Interactive prototype	0: No prototype or irrelevant prototype	1: The prototype is not interactive, lacks many features, and has many bugs; the design does not work with the goal. OR, the student submitted a prototype URL, but the prototype wasn't viewable	3: The prototype is mostly interactive, with only a few features missing and only one or two bugs; the design accomplishes the minimum requirements of the goal	5: The prototype is completely interactive, reflects the feel of the final prototype, and is ready for user testing; the design accomplishes the entire goal
User evaluation plan: completeness	0: No plan or irrelevant plan	1: User testing evaluation plan exists, but is minimal, unclear, and is not well thought out	3: The evaluation plan is mostly complete, but does not cover all questions about testing thoroughly (what is tested, what you want to learn, when, where, participants)	5: The evaluation plan is complete, answers all questions specifically, and shows a clear process for user testing
User evaluation plan: appropriateness	0: No plan or irrelevant plan	1: The student's evaluation plan does not choose to evaluate aspects of the design related	3: The evaluation plan is designed to produce some useful data, but is not justified by the student	5: The evaluation plan is very clearly motivated or innovative in a way that will ensure

(continued)

Table 3 (continued)

Category	Unsatisfactory	Bare minimum	Satisfactory effort and performance	Above and beyond
		to the design goals	(e.g. why are you doing what you are doing?– why six participants? Why in a school? etc)	rich and interesting data to address the design goals
Development goals	0: No goals met that were laid out on the development plan	1: The student met a few of the goals laid out in the development plan	2: The student met most, but not all, of the goals laid out in the development plan	3: The student met all of the goals found in the development

Students have created a paper prototype of their application in the previous assignment. Note some items have objective criteria (Did the student meet her goals?), others require subjective interpretation (Is this evaluation plan appropriate?)

to avoid certain foods which causes discomfort or health related problems." Peers rated the prototype highly for being "interactive". Staff, rated it low, because "while fully functional, the design does not seem appropriate to the goal. The diary aspect seems to be the main aspect of the app, yet it's hidden behind a search bar."

1.2.2 Where Peers Graded Lower

Figure 22b shows an application a student created as an "exciting platform, bored children can engage (physically) with other children in their neighborhood." Staff praised it as "fully interactive, page flow is complete", while some peers rated it "unpolished", and asked the student to "Try to make UI less coloured."

Appendix 2: Sample Rubric

Table 3 shows a rubric for the "Ready for testing" assignment. All other rubrics are available as online supplementary materials.

References

Alben L (1996) Defining the criteria for effective interaction design. Interactions 3(3):11–15

Amabile TM (1982) Social psychology of creativity: a consensual assessment technique. J Pers Soc Psychol 43(2):997–1013

Anderson JR, Bower GH (1972) Recognition and retrieval processes in free recall. Psychol Rev 79 (2):97–123

Andrade HG (2005) Teaching with rubrics: the good, the bad, and the ugly. Coll Teach 53 (1):27–31

Bennett RE (1998) Validity and automated scoring: it's not only the scoring. Educ Meas Issues Pract 17:4

Bennett RE, Steffen M, Singley MK, Morley M, Jacquemin D (1997) Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests. J Educ Meas 34(2):162–76

Boud D (1995) Enhancing learning through self assessment. Routledge, London

Boud D (2000) Sustainable assessment: rethinking assessment for the learning society. Stud Contin Educ 22(2):151–167

Breslow LB, Pritchard DE, DeBoer J, Stump GS, Ho AD, Seaton DT (2013) Studying learning in the worldwide classroom: research into edX's first MOOC. Res Pract Assess 8:13–25

Buxton B (2007) Sketching user experiences: getting the design right and the right design. Morgan Kaufmann, San Francisco, CA

Cadiz JJ, Balachandran A, Sanocki E, Gupta A, Grudin J, Jancke G (2000) Distance learning through distributed collaborative video viewing. In: ACM conference on computer supported cooperative work, ACM, pp 135–144

Carlson PA, Berry FC (2003) Calibrated peer review and assessing learning outcomes. In: frontiers in education conference, vol 2. STIPES

Carter S, Mankoff J, Klemmer SR, Matthews T (2008) Exiting the cleanroom: on ecological validity and ubiquitous computing. Hum Comput Interact 23(1):47–99

Cennamo K, Douglas SA, Vernon M, Brandt C, Scott B, Reimer Y, McGrath M (2011) Promoting creativity in the computer science design studio. In: Proceedings of the 42nd ACM technical symposium on computer science education, ACM, pp 649–654

Cheshire C, Antin J (2008) The social psychological effects of feedback on the production of Internet information pools. J Comput Mediat Commun 13(3):705–727

Chi EH (2009) A position paper on 'living laboratories': rethinking ecological designs and experimentation in human-computer interaction. In: Proceedings of the 13th international conference on human- computer interaction. Part I: new trends, Springer, pp 597–605

Chinn D (2005) Peer assessment in the algorithms course. ACM SIGCSE Bullet 37(3):69–73

Conti R, Coon H, Amabile TM (1996) Evidence to support the componential model of creativity: secondary analyses of three studies. Creat Res J 9(4):385–389

Corbett AT, Koedinaer KR, Haaley W (2002) Cognitive tutors: from the research classroom to all classrooms. In: Goodman PS (ed) Technology enhanced learning: opportunities for change. Lawrence Erlbaum Associates, Mahwah, NJ, p 235

Dai P, Mausam D, Weld DS (2010) Decision-theoretic control of crowd-sourced workflows. In: In the 24th AAAI conference on artificial intelligence (AAAI10), Citeseer

Dannels DP, Martin KN (2008) Critiquing critiques a genre analysis of feedback across novice to expert design studios. J Bus Tech Commun 22(2):135–159

De la Harpe B, Peterson JF, Frankham N, Zehner R, Neale D, Musgrave E, McDermott R (2009) Assessment focus in studio: what is most prominent in architecture, art and design? Int J Art Des Educ 28(1):37–51

Dow SP, Glassco A, Kass J, Schwarz M, Schwartz DL, Klemmer SR (2010) Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans Comput Hum Interact 17(4):18

Dow S, Kulkarni A, Klemmer S, Hartmann B (2012) Shepherding the crowd yields better work. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, ACM, pp 1013–1022

Drexler A, Chafee R et al (1977) The architecture of the ecole des beaux-arts. MIT, Cambridge, MA

Efron B, Tibshirani R (1993) An introduction to the bootstrap, vol 57. Chapman & Hall, New York

Ehrlinger J, Johnson K, Banner M, Dunning D, Kruger J (2008) Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent. Organ Behav Hum Decis Process 105(1):98–121

Falchikov N, Goldfinch J (2000) Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. Rev Educ Res 70(3):287–322

Fallman D (2003) Design-oriented human-computer interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 225–232

Feldman EB (1994) Practical art criticism. Prentice Hall, New York

Forlizzi J, Battarbee K (2004) Understanding experience in interactive systems. In: Proceedings of the 5th conference on designing interactive systems: processes, practices, methods, and techniques, ACM, pp 261–268

Fox A, Patterson D (2012) Crossing the software education chasm. Commun ACM 55(5):44–49

Galinsky AD, Moskowitz GB (2000) Counterfactuals as behavioral primes: priming the simulation heuristic and consideration of alternatives. J Exp Soc Psychol 36(4):384–409

Gallien T, Oomen-Early J (2008) Personalized versus collective instructor feedback in the online courseroom: does type of feedback affect student satisfaction, academic performance and perceived connectedness with the instructor? Int J E-Learn 7(3):463–476

Gerdeman RD, Russell AA, Worden KJ (2007) Web-based student writing and reviewing in a large biology lecture course. J Coll Sci Teach 36(5):46–52

Greenberg S (2009) Embedding a design studio course in a conventional computer science program. In: Kotzé P et al (eds) Creativity and HCI: from experience to design in education. Springer, Boston, MA, pp 23–41

Guo S, Parameswaran A, Garcia-Molina H (2012) So who won?: dynamic max discovery with the crowd. In: Proceedings of the 2012 international conference on Management of Data, ACM, pp 385–396

Hearst MA (2000) The debate on automated essay grading. IEEE Intell Syst Appl 15(5):22–37

Hsi S, Agogino AM (1995) Scaffolding knowledge integration through designing multimedia case studies of engineering design. In: Proceedings of the frontiers in education conference, 1995, vol 2, IEEE, p 4d1–1

Huang SW, Fu WT (2013) Enhancing reliability using peer consistency evaluation in human computation. In: Proceedings of ACM: computer supported collaborative work, ACM

Ipeirotis PG, Provost F, Wang J (2010) Quality management on amazon mechanical turk. In: Proceedings of the ACM SIGKDD workshop on human computation, ACM, pp 64–67

Kaufman JC, Baer J, Cole JC, Sexton JD (2008) A comparison of expert and nonexpert raters using the consensual assessment technique. Creat Res J 20(2):171–178

Khatib F, DiMaio F, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, Zabranska H, Pichova I, Thompson J, Popović Z et al (2011) Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat Struct Mol Biol 18(10):1175–1177

Kim H, Hinds P (2012) Harmony vs. disruption: the effect of iterative prototyping on teams creative processes and outcomes in the West and the East. In: Proceedings of the ICIC: international conference on intercultural collaboration, ACM

Kittur A, Nickerson J, Bernstein M, Gerber E, Shaw A, Zimmerman J, Lease M, Horton J (2013) The future of crowd work. In: ACM conference on computer supported coooperative work (CSCW 2013)

Kizilcec RF, Piech C, Schneider E (2013) Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. Proceedings of the Third International Conference on Learning Analytics and Knowledge, LAK'13, Leuven, Belgium, pp 170–179, ISBN 978-1-4503-1785-6, ACM, New York

Klemmer SR, Hartmann B, Takayama L (2006) How bodies matter: five themes for interaction design. In: Proceedings of the 6th conference on designing interactive systems, ACM, pp 140–149

Kraut RE, Resnick P (2011) Evidence-based social design: mining the social sciences to build online communities. MIT, Cambridge, MA

Kuebli JE, Harvey RD, Korn JH (2008) Critical thinking in critical courses: principles and applications. In: Dunn DS, Halonen JS, Smith RA (eds) Teaching critical thinking in psychology: a handbook of best practices. Wiley, New York, p 137

Kurhila J (2012) Human-computer interaction by coursera opened for credit for the students of the department. http://www.cs.helsinki.fi/en/uutiset/72025

Lawson B (2006) How designers think: the design process demystified. Architectual Press, Oxford

Lewin T (2012a) Education site expands slate of universities and courses. The New York Times

Lewin T (2012b) One course, 150,000 students. The New York Times, July 2012

Lewin T (2013a) College of future could be come one, come all. January 2013, The New York Times

Lewin T (2013b) Five online courses are eligible for college credit. February 2013, The New York Times

Lewin T (2013c) Students rush to web classes, but profits may be much later. January 2013, The New York Times

Lewin T (2013d) Universities abroad join partnerships on the Web. February 2013, The New York Times

Little JL, Bjork EL (2012) Pretesting with multiple-choice questions facilitates learning. In: Proceedings of the annual meeting of the cognitive science society

Markman AB, Gentner D (1993) Splitting the differences: a structural alignment view of similarity. J Mem Lang 32(1993):517–517

Marlow J, Dabbish L, Herbsleb J (2013) Impression formation in online peer production: activity traces and personal profiles in github. In: Proceedings of the 2013 conference on computer supported cooperative work, ACM, pp 117–128

Martin FG (2012) Will massive open online courses change how we teach? Commun ACM 55 (8):26–28

Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: a theory of self-concept maintenance. J Marketing Res 45(6):633–644

Murtaugh PA, Burns LD, Schuster J (1999) Predicting the retention of university students. Res High Educ 40(3):355–371

Nicol DJ, Macfarlane-Dick D (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ 31(2):199–218

Nielsen J (1993) Iterative user-interface design. Computer 26(11):32–41

Nielsen J (1994) Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 152–158

Palen L (1999) Social, individual and technological issues for groupware calendar systems. In: Proceedings of the SIGCHI conference on human factors in computing systems: the CHI is the limit, ACM, pp 17–24

Pendleton-Jullian A (2010) Four (+1) Studios. CreateSpace Independent Publishing, New York, NY

Perry WG (1970) Forms of intellectual development in the college years. Holt, New York

Pintrich PR (1995) Understanding self-regulated learning. New Dir Teach Learn 63:3–12

Pintrich P, Zusho A (2007) Student motivation and self-regulated learning in the college classroom. In: Perry R, Smart J (eds) The scholarship of teaching and learning in higher education: an evidence-based perspective. Springer, New York, pp 731–810

Reimer YJ, Douglas SA (2003) Teaching HCI design with the studio approach. Comput Sci Educ 13(3):191–205

Roberts E, Lilly J, Rollins B (1995) Using undergraduates as teaching assistants in introductory programming courses: an update on the Stanford experience. ACM SIGCSE Bullet 27(1):48–52

Schön D (1985) The design studio: an exploration of its traditions and potential. Royal Institute of British Architects, London

Snodgrass A, Coyne R (2006) Interpretation in architecture: design as a way of thinking. Routledge, London

Socher R, Huval B, Manning CD, Ng AY (2012) Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the 2012 conference on empirical methods in natural language processing (EMNLP)

Stanley CA, Porter ME (2002) Engaging large classes: strategies and techniques for college faculty. Anker Publishing Company, Bolton, MA

Surowiecki J (2005) The wisdom of crowds. Anchor, New York

Szpir M (2002) Clickworkers on mars. Am Sci 90(3):13–25

Tinapple D, Olson L, Sadauskas J (2013) CritViz: Web-based software supporting peer critique in large creative classrooms. Bullet IEEE Tech Committee Learn Technol 15(1):29

Tohidi M, Buxton W, Baecker R, Sellen A (2006) Getting the right design and the design right. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 1243–1252

Tomayko JE (1991) Teaching software development in a studio environment. ACM SIGCSE Bullet 23(1):300–303

Topping K (1998) Peer assessment between students in colleges and universities. Rev Educ Res 68 (3):249–276

Uluoglu B (2000) Design knowledge communicated in studio critiques. Des Stud 21(1):33–58

Venables A, Summit R (2003) Enhancing scientific essay writing using peer assessment. Innov Educ Teach Int 40(3):281–290

Widom J (2012) From 100 students to 100,000, ACM SIGMOD blog. http://wp.sigmod.org/?p=165

Winograd T (1990) What can we teach about human-computer interaction?(plenary address). In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 443–448

Zaidan OF, Callison-Burch C (2011) Crowdsourcing translation: professional quality from non-professionals. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, vol 1. pp 1220–1229

Zimmerman BJ, Schunk DH (2001) Reflections on theories of self-regulated learning and academic achievement. Self-regulated learning and academic achievement. Theor Perspect 2 (2001):289–307

Tagging User Research Data: How to Support the Synthesis of Information in Design Teams

Raja Gumienny, Steven Dow, Matthias Wenzel, Lutz Gericke, and Christoph Meinel

Abstract In user-centered design processes, one of the most important tasks is to synthesize information from user research into insights and a shared point of view among team members. This paper explores the synthesis process and opportunities for providing computational support. First, we present interviews on the common practices and challenges of information synthesis from people with different levels of experience. Based on these interviews we developed digital whiteboard software for sorting individual segments of user research. The system separates out individual and group activity and helps the team to externalize and synthesize their different views of the data. Through a case study, we explore the differences between computer-supported group interaction and an individual clustering condition. We learned that participants really appreciated an individual working phase before entering a group synthesis phase. Novice designers tended to prefer the structured computer-supported synthesis process that externalizes the different views of each team member. More experienced designers preferred to freely arrange information segments and create clusterings on their own.

R. Gumienny (✉) • M. Wenzel • L. Gericke
Hasso-Plattner-Institute, University of Potsdam, Potsdam 14482, Germany
e-mail: tele-board@hpi.uni-potsdam.de

S. Dow
HCI Institute Carnegie Mellon University, 3615 Newell Simon Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: spdow@cs.cmu.edu

C. Meinel
Hasso-Plattner-Institute for Software Systems Engineering, Campus Griebnitzsee, Potsdam 14482, Germany

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_10, © Springer International Publishing Switzerland 2015

1 Introduction

Conducting in-depth user research is a vital part of user-centered design processes (Goodman et al. 2012; Beyer and Holtzblatt 1998; Rogers et al. 2011). However, user research usually produces large amounts of data and it is difficult to integrate the collected data and information into design ideas and solutions (Kolko 2011; Brown 2009).

A common way of dealing with this large amount of qualitative data is writing the observations and interview results on sticky notes or paper cards and afterwards grouping them according to their semantic affinity (Beyer and Holtzblatt 1998; Curtis et al. 1999; Harboe et al. 2012; Hinman 2011).

Usually, this is a team-based activity with the goal of "making sense out of the data" (Kolko 2011; Oehlberg et al. 2012a). While this is easily said, it is a very difficult task to generate new insights and knowledge (as opposed to basic information) from the collected data (Kolko 2011; Hinman 2011; Oehlberg et al. 2012a). And it is particularly difficult to develop a *shared* understanding of the data and insights among all team members (Oehlberg et al. 2012a; Hey et al. 2007; Oehlberg and Roschuni 2011).

Several research projects deal with the question how to transfer the process of synthesizing qualitative data with paper notes to the digital world in order to take advantage of functions like saving and sending the content to remote colleagues, e.g. (Harboe et al. 2012; Judge et al. 2008). There are fewer research projects that deal with insight generation as such. Our goal was to understand why it is so difficult for design teams to "make sense" out of their user research data and how we can support them during this synthesis of information.

In this chapter, we present results from seven interviews we held with people of different levels of experience regarding information synthesis. Based on their needs and combined with findings from other research, we developed a tool to support the collaborative synthesis process for design teams.

With the help of this tool, we tested the influence of two factors for synthesizing information. These factors are (1) a phase of working individually with the data and (2) applying different *perspectives* or *tags* to the data in order to reveal and discuss different points of view among team members. We will discuss the underlying theory of these factors and present the results of a study with six design teams.

2 Background and Related Work

User research or field research is the basis for all user-centered design processes. Finding out more about the interests of users and developing empathy for their needs helps to develop more useful and innovative products (Brown 2009). However, in industry, clients may question the purpose of the time-consuming and costly user research efforts. A reason for this could be the missing visibility and tangibility of how this data is analyzed and integrated in the future design of a product or service (Kolko 2011). In comparison, other parts of a design process are more easy to understand by non-designers as they can see what happens when the design team sketches new ideas or builds a prototype (Kolko 2011).

Besides the fact that the synthesis of information is hard to present, it is also cognitively challenging for the design team. Filtering, organizing and making sense of uncertain and ambiguous information is complicated and exhausting (Hey et al. 2008; Kolfschoten and Brazier 2012). Working in a team can give assistance and is important for the following steps in the design process, but it also introduces the difficulty of creating a common ground and making decisions that all team members support (Hey et al. 2007). Based on background and experience, people have different views on situations and interpret or frame them in different ways. Everybody carries individual *frames* that consist of implicit knowledge structures (Hey et al. 2007; Schön 1984). By sharing these different frames with each other, they can be aligned in order to develop a shared understanding (Hey et al. 2007). Especially when dealing with ambiguous information collected by different people during user research, different individual frames lead to different points of view and it is particularly important that design teams share their views for developing a shared understanding. Thereby, communication and discussions play an important role in generating new meaning (Hill et al. 2002). Teams that synthesize their knowledge into a shared understanding tend to have more successful design processes and outcomes (Hey et al. 2008; Hill et al. 2002).

The term *synthesis* is also used in other contexts,[1] however what we call the *information synthesis* is the practice of integrating, organizing, filtering and evaluating external information into the design process as described by Jon Kolko (Kolko 2011). In other related work, this understanding of synthesis may be referred to by the terms *collaborative synthesis* (Robinson 2008), *framing* (Hey et al. 2007, 2008; Schön 1984), *sensemaking* (Pirolli and Card 2005; Naumer et al. 2008), *collaborative sensemaking* (Novak 2007; Umapathy 2010), or *information analysis* (Isenberg et al. 2008).

Though *sensemaking* describes the act of "making sense of user research information" pretty well, related work with this title mainly use the term in the way Russell et al. (1993) defined it: the process of searching for and organizing information. These research projects focus on analyses that make sense of large amounts of data in the internet (Sharma 2011; Qu and Furnas 2005), other large networks (Chau et al. 2011), or document analysis (Wright et al. 2006). The described data are often "hard facts" that need to be combined, such as facts about digital cameras (Sharma 2011; Shrinivasan and van Wijk 2008) or neighborhoods characteristics (Cheng and Gotz 2009). In this understanding, the term also

[1] Some authors use this term to refer to all activities of assembling or creating the form of the design solution – in contrast to the term analysis that refers to the activity of investigating and defining the design problem (Alexander 1964; Lawson 2006; Bamford 2002)

involves the seeking and searching for information (Pirolli and Card 2005) and not only the act of condensing information in order to create new knowledge.

Several research projects focused on searching for and navigating through huge amounts of data and developed tools for improved information visualization, searching and tagging. They studied how different devices, such as large displays (Andrews et al. 2010), tabletop displays (Morris et al. 2010), or personal and shared devices (Wallace et al. 2013) can improve the sensemaking process. Novak (2007) and Umapathy (2010) also acknowledge the importance of knowledge exchange in interdisciplinary teams and study how teams come to a shared understanding during sensemaking. With his tool, Novak suggests that visualizing implicit knowledge structures improves the knowledge exchange among team members (Novak 2007).

There are fewer tools that focus on the synthesis of qualitative user research data—this is "information synthesis" as we understand it. This research often focuses on the question of how to transfer paper notes to the digital world or augment them in order to make use of digital functions. For example, Judge et al. (2008) study how multiple display environments can improve affinity diagramming. Harboe et al. (2012) augment paper notes with barcodes for locating the notes via text search.

Though these approaches are certainly useful for affinity diagramming or working with qualitative data on sticky notes, they do not tackle the problem of synthesizing information. That is, how design teams can be supported in the task of condensing information and developing a shared understanding. A big challenge is the ambiguous nature of qualitative data in design tasks and the great amount of tacit knowledge that is important for the process. Hey et al. (2007, 2008) studied how design teams deal with the different frames among team members and how they come to a shared understanding. They developed a framework and design principles for design team framing. Kolko (2011) also offers several methods intended to help design teams during synthesis. Some of the few researchers who focused on developing a tool for the synthesis of user research information are Oehlberg et al. (2012b). Their tool, Dazzle lets design teams share their collected files, annotate them, and capture whiteboard images. However, the sharing of information stays on the file level and does not go to the level that deals with (the) individual pieces of information and how they relate. With affinity diagramming the most difficult part is making sense out of the interrelationships between paper notes.

In order to broaden our own understanding and experience on the general needs of users during synthesis and to relate them to the findings of other researchers, we interviewed people with different levels of experience on how they manage the information overload and how they synthesize their insights.

3 Interviews on Information Synthesis

We conducted seven interviews with two design students, four professional designers (graphics and interaction designers) and one design professor. The interview length ranged between 20 and 45 min. We used interview guidelines focusing on how people condense, select and decide when synthesizing information and how they evaluate the approaches they employ. All interviews were taped with a voice recorder. We used open-coding techniques to discover patterns and recurring topics (Corbin and Strauss 1990). For each interview, we wrote various memos on sticky notes and first clustered them on separate boards, afterwards we analyzed similarities and differences between the interviews.

The main topics we identified in the analysis are described briefly in the following.

3.1 Relevance for the Entire Design Process

When we asked how people processed information from user research, we found that some interviewees did not understand what the question was about. The expert designers mainly assimilated information "on the fly" and most of the time alone. In contrast, other interviewees stated that the synthesis was a very crucial point within the whole design process itself and its importance should not be underestimated. It helps to identify general statements, principles, trends, needs and requirements with regard to the design task.

3.2 Sequence and Characteristics of Subtasks

People with a developed understanding of information synthesis generally talk about their user research results with other people. It may be a colleague or a whole team, depending on company or school structure. During these conversations, people usually take notes, either on normal paper or sticky notes. Some participants summed it up under the term "storytelling". Afterwards, they try to find similarities of what they have heard and try to group them under general terms ("clustering"). Important topics are sometimes displayed in different frameworks or diagrams, such as a process diagram to show workflows or relationships. In the end, people write down their most important insights or principles. This relates to Kolko's methods of synthesis as e.g. "prioritizing" or "concept-mapping" (Kolko 2011) or the observations of other researchers (Hinman 2011; Robinson 2008). However, not everybody follows an elaborate structure when synthesizing information, but pursues a more intuitive, coincidental sequence of steps.

3.3 Decision Making

Decision making occurs when designers have to prioritize or select between different pathways. We learned that intuition plays an important role when making decisions in information synthesis. When we asked our interview partners how they identify and define insights or decide on their priority, nobody could give a clear answer. In particular, experienced interviewees said they follow their intuition and state that especially the gradually growing experience of designers enables good intuitive decision making. Interviewees with little experience stated that decision making is important and also very difficult as they do not have a lot of experience about how to decide. Literature also suggests that the role of intuition is supported by experience (Beyer and Holtzblatt 1998; Kolko 2011; Cropley 2006). Accordingly, experience helps to develop tacit knowledge about different situations and implications.

3.4 Extent of Discursivity

Our interviews suggest that discourse between the members of a design team is seen as a decisive part of information synthesis. Some interviewees even defined the synthesis as "a team process with a lot of discussions". On the contrary, other interviewees stated that they collect and synthesize information in general on their own and talk about their observations with only a few people—generally expert designers—later on. Thus, we could observe that the extent of discursivity varies with teams and design situations. In literature, discourse among design teams is seen as rather important for user-centered design (Hill et al. 2002; Krippendorff 2006; Lloyd 2000).

3.5 Forms of Media

Our interview partners use different kinds of media to communicate and process information, though analog media such as paper, sticky notes and traditional whiteboards are the most commonly used. Nevertheless, especially interviewees who work in companies (instead of education) stated that at some point digital media in the form of word processors, presentation programs or wikis is used as well.

3.6 Information Trade-Off

Converging information and finding design principles with a higher degree of abstraction is one of the goals of the synthesis phase. However, we observed different levels of information trade-off among our interview partners. Some interviewees try to keep and externalize as much information as possible, partly because they are afraid to lose information and partly because their stakeholders set these restrictions. Others stated that it is not possible and also not desirable to save all information in the design process, as it is important to quickly focus on the most important points. Most interviewees agreed that it depends on the level of experience when deciding which and how much information is important to process in the design process.

3.7 Team Interaction

We observed through the interviews several incidences in which implicit team dynamics influences the synthesis process rather unconsciously. For instance, interviewees mutually agreed that the basis for joint decisions is only possible if team members share a common ground of trust and respect cf. (Schumann et al. 2012). In another example, an interviewee stated that people who prefer to enforce the own view strongly influence the whole synthesis process. In addition, the synthesis is described as exhausting and its success highly depends on the motivation of the team members. Therefore, we regard the area of team interaction, with a special focus on team dynamics, biases and motivation, as important for a deeper understanding of information synthesis.

3.8 Communicating Preliminary Results to External Persons

Interviewees who work in companies stated that customers and stakeholders complain that they hardly see what happens during the synthesis phase cf. (Kolko 2011). Several clients want to understand where the design ideas and solutions originate and whether the budget for e.g. user research has been spent reasonably. However, such requirements generally presume a view of the relationship between design solutions and user research data, which is normally only possible towards the end of the design process. In particular in early stages of the design process, designers often face communication gaps that make it difficult to tell outsiders about the progress of the design process. In this context, information synthesis can help to create presentable states of knowledge. However, our interviews suggest that this seems to be less of a problem for the more experienced designers, as the relationship between clients and designers then rather builds upon trust. This shows that external communicability requirements depend on the relationship between designers and clients and how much confidence they have in the respective design approach.

3.9 *Organizational Restrictions and Enablers*

Especially in companies where people are working on several projects at the same time, a challenge presents itself in the overall lack of time for synthesis, as well as the many disruptions. In this case, teams also face the problem that one or more team members are missing and it is difficult for them to catch up afterwards. Sometimes there are strict rules on how the synthesis should be carried out. On the other hand, research goals are often not clearly defined and this leads to problems between the designers and clients.

In summary, we can draw the following conclusions from the interviews and our own experience on information synthesis: The synthesis is important and necessary but also difficult and exhausting or—as one of our interviewees said—"The synthesis is a necessary evil". Mostly it is a stressful team process, which depends on well-functioning team dynamics. Especially for beginners it is challenging because it heavily depends on experience and intuition. Much uncertainty and ambiguity is involved, making the whole process neither visible nor tangible for observers from outside. Last but not least it takes a lot of time, which is often not provided or scheduled.

For a more detailed analysis of the interviews and the framework we derived from them, as well as a literature review related to design research see (Gumienny et al. 2011).

4 Support for Collaborative Synthesis

Combining the insights from interviews and related work, we sought to improve different aspects that seem to influence the process and outcome of the synthesis. First, we want to help team members have a better understanding of the information they collected during user research. Each team member should have time to familiarize and engage with the data collected, especially with the notes written by other people. Therefore, we want to give each team member some time to work with the data individually at the beginning of the synthesis.

Second, we want to support teams in forming a *shared* understanding. We think that visualizing the personal views of each team member is an important prerequisite for providing insight into community perspectives cf. (Novak 2007). We want to create explicit representations of knowledge structures (Umapathy 2010) and let the team compare and analyze different representations. If the team is unaware of these differences before a decision, this may result in conflicts that hinder the ongoing progress. It is also important to understand each other's perspectives (Hey et al. 2008).

Third, as stated above, we learned that in some teams, rather dominant team members lead the whole synthesis and decision-making process. As a result the outcome does not necessarily reflect the opinion of the whole team, possibly leading to later conflicts and disregarding the advantages of multidisciplinary team work. Therefore, we want to assure all team members are involved equally and show the contribution of each person.

Fourth, we want to give novice designers more guidance and encourage them to work with the data. They should have support in getting started instead of discussing how to deal with the huge amount of sticky notes as we have often observed in student teams. People told us that they perceive the whole process as very exhausting. We hope that it feels less stressful and more managable if we divide the synthesis into different steps, designers may follow one after another.

4.1 Synthesis Guide

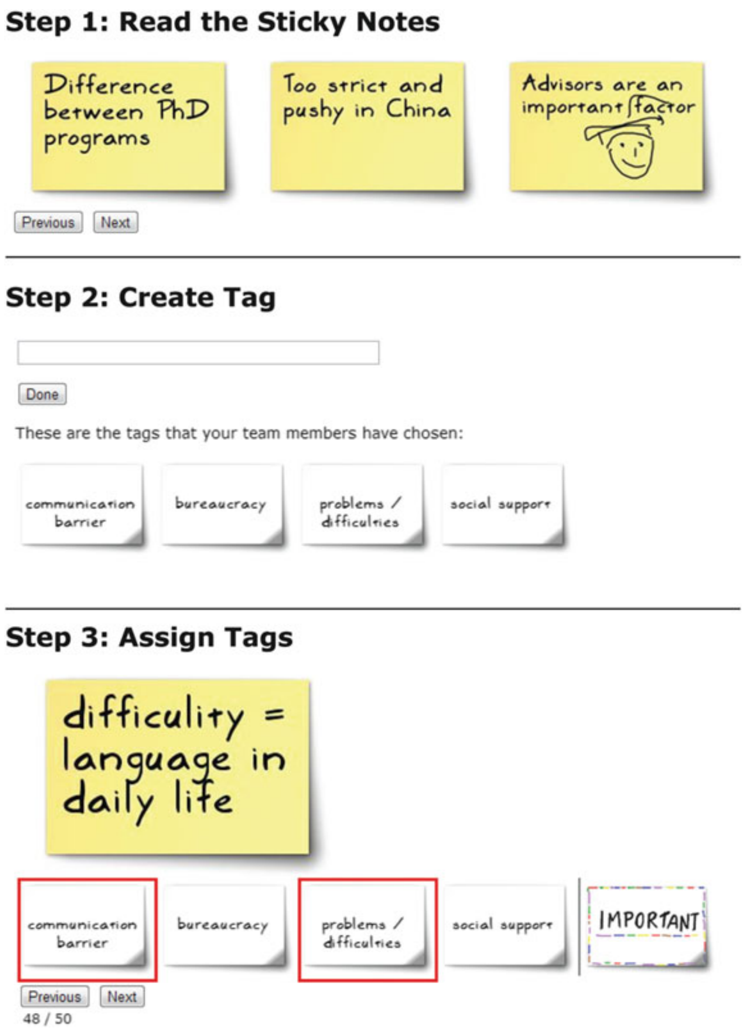
Combining these topics, we created a Synthesis Guide for a digital whiteboard system, which provides guidance and lets people work individually first and externalizes the team members' points of view in the end.

The main instrument of the Synthesis Guide is the act of applying different "perspectives" or "tags" to the user research data. The process of applying the tags is similar to the worker tasks of the Cascade system by Chilton et al. (2013). In this system, crowdworkers generate tag categories for a set of items. The best tags are picked, and then they are given to other workers who apply the tags to the items. This crowdsourcing approach could also be used for user data here. However, we think it is important that the members of the design team do these tasks because they also carry implicit knowledge from user research. Additionally, they should use the knowledge they generate during synthesis for idea generation and prototyping later on.

We see a "perspective" or "tag" as a frame or point of view that can be applied to situations or data (Kolko 2011). This is based on the assumption that different people generally have different mental models (Klimoski and Mohammed 1994; Lim and Klein 2006). These mental models are based on what we have learned and experienced, and we see the world from this perspective (Kolko 2011; Hey et al. 2007). With the task of applying perspectives to pieces of information, the different perspectives of team members are externalized and thus make them aware of the different views they have [especially in interdisciplinary design teams (Brown 2009)].

During discussions, we noticed that the term "perspective" is not understood immediately and instead of "applying perspectives", people preferred the term "tagging". Therefore, we continued to use "tags" instead of "perspectives".

In the first step of the Synthesis Guide, each team member gets an overview of all sticky notes written by the team after conducting user research. In order to reduce the overload of seeing all sticky notes at once, the notes are presented in



Step 1: Read the Sticky Notes

Difference between PhD programs

Too strict and pushy in China

Advisors are an important factor

Step 2: Create Tag

These are the tags that your team members have chosen:

Step 3: Assign Tags

difficulity = language in daily life

Fig. 1 The three steps of the Synthesis Guide. First, each user should read all sticky notes. Second, each user creates a tag that should be applied to the sticky notes. Third, each user tags every sticky note with the tags of the team. The "important" tag is highlighted by the system to indicate that it (they) should get special attention afterwards

groups of three on each page. With the help of the "next" button, users can flip through all notes, see Fig. 1, top.

In the second step, each user is to create a perspective or tag related to the sticky notes they have seen, Fig. 1, middle. Alternatively, they may choose one of the example tags offered. Each user creates his own tag that he or she considers interesting. To avoid duplicates and foster a broader range of tags, users will see tags already created by their team members.

After all team members enter one perspective, the Synthesis Guide will lead to the third step—the tagging view, see Fig. 1, bottom. Each sticky note will be displayed on one page together with all tags the team has chosen. Additionally, the tag "important" is offered to indicate that a note is important even though it does not fit to any of the chosen tags. Users shall now select all tags that fit to the displayed sticky note. They can select as many tags as they like, or none at all. After pressing the "next" button, the tags are saved and the next sticky note is displayed. Each team member should do the tagging individually. By going through the steps of the Synthesis Guide everybody is "forced" to engage with the data and cannot leave this to other team members. Additionally, the point of view of each member is collected.

After each team member has completed the three steps, the system offers a result view for each tag, see Fig. 2. Sticky notes selected by all team members appear on the highest level and are enlarged. Depending on the number of selections, the other notes are displayed on a lower level and smaller. Sticky notes not selected at all are not displayed. The result pages are intended to give an overview of how the team understands the collected information. For example, in Fig. 2 right, the team obviously had different opinions on which sticky notes are important (i.e. what information on them). The team can now discuss why they think certain information is important or not. On the other hand, they share the same view on "communication barriers", Fig. 2, left. This may strengthen the team spirit and sense of community.

4.2 Design Objectives

The system had the following design objectives:

4.2.1 View Data from Different Angles

When people apply different tags to the sticky notes, they think about the relationship between the respective tag and the data set of the respective tag with the data set. This way, they must see the data from another angle or frame. While people contemplate the data and try to view it from new angles, they engage with the data in a way they would not during standard clustering. This in-depth engagement with the data may lead to a better understanding.

4.2.2 Externalize Different Points of View

When people apply tags individually, they do it without being influenced by their co-workers. On the results pages of the Synthesis Guide, the different opinions are visualized. We assume that people are often not aware of their different points of

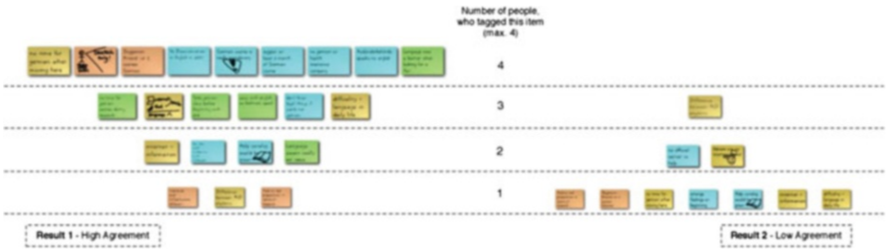

Number of people, who tagged this item (max. 4)
4
3
2
1
Result 1 - High Agreement
Result 2 - Low Agreement

Fig. 2 Examples of the result screens after the tagging. On the left the team has a very similar understanding with regard to this tag ("communication barrier"), because several sticky notes have been tagged by all team members. On the *right*, the understanding is pretty diverse, as the majority of notes have only been tagged by one person (with the tag "important")

view. In the results pages of the Synthesis Guide the team sees the similarities and differences of views from the different tags. Based on these views, they can start a discussion and come to one shared point of view.

4.2.3 Involve All Team Members Equally

When each team member is tagging the notes it is necessary to engage with the data as something that cannot be left to fellow team members. The input of all team members is counted equally and displayed on the results pages. This way, no team member is shut out, and everybody is involved equally.

4.2.4 Give Guidance to Novice Designers

Through its predefined steps, the Synthesis Guide is intended to give guidance to the team and help them get started with the synthesis. We observed that especially novice design teams often do not know how to start and waste time trying to agree on a method or framework to use for the synthesis.

5 Evaluation

To evaluate whether the Synthesis Guide really improves the synthesis and helps a team come to a shared understanding, we conducted a case study. We ran a series of pilot studies and then created a within-subjects study with two conditions, see Fig. 3. We tested the Synthesis Guide condition, i.e. a structured way of doing the synthesis, and compared it to an unstructured clustering condition. Our purpose was to find out what effect the structure and the tagging functionality have for the team process during synthesis.

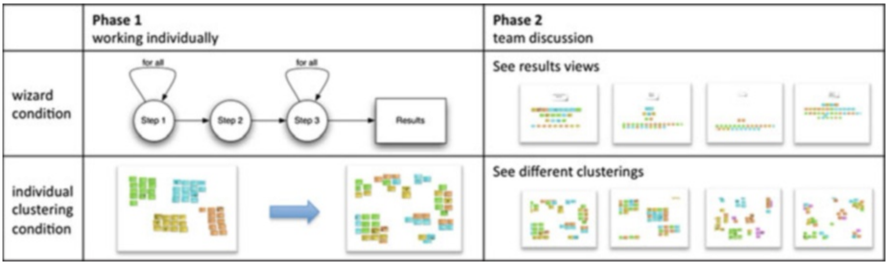


Fig. 3 Study setup for evaluating the difference between a structured synthesis process with the help of a Synthesis Guide and the unstructured clustering where each user could arrange the sticky notes freely. In both conditions, the team members were working on their own in the first phase. Afterwards, they created a point of view of the given data in a team discussion

5.1 Participants

We recruited 24 participants for six teams. Three out of the six teams started with the structured condition, three with the unstructured clustering condition. All participants had previous experience with design thinking and synthesis, but on different levels. All teams were interdisciplinary, i.e. participants had different academic backgrounds, and consisted of four people each. Of the 24 participants, 14 were female. All teams were mixed-gender and the average age of members was 28. Most participants did not know each other previously.

5.2 Procedure

In both conditions, the teams received a dataset of sticky notes and were to create a point of view (POV)—a sentence that summarizes the most important findings from the sticky notes. In a real situation, these notes are written by the team members themselves. Due to time constraints, we offered notes created by other design thinking teams after interviewing people about two challenges. These challenges were: “How to improve the arrival experience of foreign researchers coming to a foreign university” and “How to improve the airport check-in and boarding process”. The datasets consisted of 50 sticky notes each. Each challenge was done three times with the structured condition and three times with the unstructured clustering condition.

In the structured condition, teams used the Synthesis Guide. First, each team member read all sticky notes on a laptop by clicking through the pages of step 1. Then, each person created one tag or perspective. As a last step, they tagged all notes with the tags they had created. The teams had 15 min to complete the steps of the Synthesis Guide. Afterwards, they looked at the results pages and had 13 min to discuss these results and create a team POV from the data (see Fig. 3).

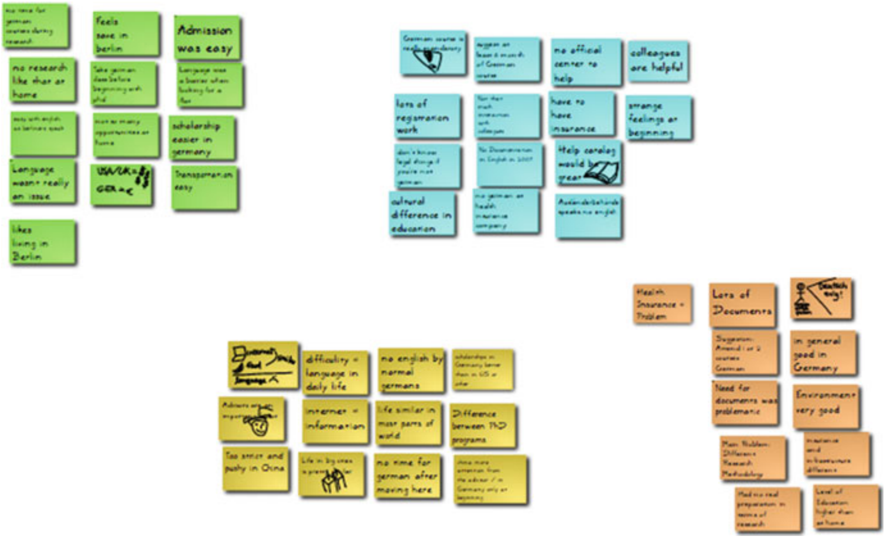

Feels save in berlin
Admission was easy
no research like that at home
scholarship easier in germany
Language wasn't really an issue
Transportation easy
likes living in Berlin
no official center to help
colleagues are helpful
lots of registration work
have to have insurance
strange feelings at beginning
Help catalog would be great
cultural difference in education
Health Insurance = Problem
Lots of Documents
in general good in Germany
Need for documents was problematic
Environment very good
difficulty = language in daily life
no english by normal germans
internet = information
life similar in most parts of world
Difference between PhD programs
Too strict and pushy in China
no time for german after moving here

Fig. 4 Start screen for the unstructured clustering condition. The sticky notes were grouped by color, representing the different interview partners. Participants could rearrange the sticky notes freely and create their own structures of the notes

In the unstructured clustering condition, each team member had the sticky note dataset in a digital whiteboard application on a laptop. The sticky notes were grouped by color (based on the interview person), see Fig. 4. People could move around the sticky notes with their mouse and cluster the notes as they liked. Additionally, they could zoom in and out. All other functions of the whiteboard application were turned off to make people focus on the content instead of the functions. Each team member had his own laptop and 15 min to work with the data individually. We gave the instruction: “get an understanding of the information on the sticky notes”. Afterwards, the team sat together and had 13 min to discuss what they learned from the data and to create a team POV as in the structured condition.

In the first phase of both conditions, when the teams worked individually at their laptops, they were sitting around a table and could not see their fellow team members’ screens. In the second phase, when discussing their findings and creating the POV together, we had them now turn around the laptops and place them in one row. This way, everybody could see all screens at the same time and they could point to sticky notes on the screens, see Fig. 5.

After each condition, team members separately filled in three forms: one with his or her most important insights from the data, one with comprehension questions related to the respective challenge, and a post-task questionnaire. After conducting both conditions, each participant additionally filled in a post-test questionnaire. The post-test questionnaire included Likert-Scale questions as well as free response questions. Each experiment lasted about 2 h.

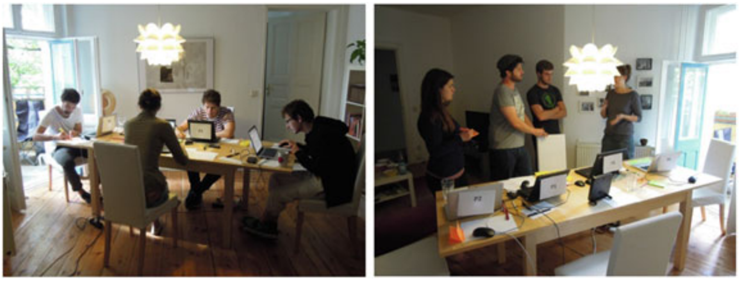


Fig. 5 Study setup for the individual phase (*left*) and the team discussion phase (*right*). In both conditions the teams' first had a phase of working individually (*left*). Afterwards, the laptops were put in one row in order to see the clusterings of each participant in the individual condition or the tagging result views in the structured condition (*right*)

5.3 Quantitative Results

In the post-task questionnaire, different questions addressed similar understanding among team members. The areas covered were: how easy it was to understand the team member's points of view, participants' satisfaction with their own contributions and with those of their team members, time management, whether participants felt they were on the right track, and the general difficulty creating the POV. We performed an analysis of variances (ANOVA) with the condition (structured or unstructured clustering) as a factor and the responses to the Likert-scale questions as dependent variables. Between conditions, there were no significant differences for all questions.

In the post-test questionnaire we asked in which condition the common ground or understanding was best. We saw a marginally higher value for the structured, Synthesis Guide condition (on average 1.46, SD = 0.51 opposed to 1.54, SD = 0.51 in the unstructured clustering condition, values ranging between 1, best and 2, worst). We also asked for factors like efficiency and fun. The mean values hardly differed: 1.54, SD = 0.51 (unstructured clustering); 1.46, SD = 0.51 (structured clustering) and 1.42, SD = 0.5 (unstructured clustering); 1.58, SD = 0.5 (structured clustering).

Additionally, we analyzed the insights questionnaires regarding general quality and similarity among team members. For both attributes, we could not find differences depending on the teams' condition. In general, some teams had more similar insights than other teams—independent from their condition.

To test a team's comprehension of the given data we created five sample questions for each data set. Each correct answer received one point. Then we calculated the amount of points a team earned per condition. We could see slightly better results for the unstructured clustering condition with the foreign researchers challenge (on average 12.67 points as opposed to 11.00 points in the structured clustering condition), but these values are not significant. For the airport challenge there was no difference: on average 17.00 points in the unstructured clustering condition, 17.33 points in the structured clustering condition.

For all of the reported measures, we could not find significant differences between the conditions. Therefore, we focused our analysis more on qualitative data, that is, the free response questions of the post-test questionnaire.

5.4 Qualitative Results

Overall, we observed that all teams in all conditions created POVs within the given time frame. All team members participated, though some were more active than others. In the Synthesis Guide condition, each team created a perspective as directed. Some participants did this very quickly; others needed some time. Some teams asked each other clarifying questions about the perspectives, especially to dissociate them from each other.

In the post-test questionnaire, we asked which condition the participants preferred and for what reasons. We also asked which advantages and disadvantages they saw for each of the conditions. The results are summarized with regard to the main findings.

5.4.1 Showing All User Data Supports Overall Comprehension

For overall comprehension, people preferred the unstructured clustering condition. They liked the ability to structure the sticky notes on their own in as many clusters and hierarchies as they needed:

> It has the advantage that everyone can use as many clusters as he likes for his own sensemaking and not just 4/5 tags. (T4P1)

Participants also pointed out that they liked having an overview of the information on all sticky notes at a glance and that it was always visible:

> It is an advantage to arrange post-its directly on the screen while having an overview of all the post-its and on the same screen. (T6P3)

> During clustering you see groups emerging, and in the end you try to find a name. When you have to tag notes before clustering, you kind of have to know the names first. (T3P1)

In the structured condition, some participants were afraid of forgetting or losing important information. The reasons may be the following:

> The facts people choose the most don't have to be the best or most important ones (T5P4).

> It implies that the insights that can't be categorized as well are not as good, which isn't true. (T5P2)

In summary, having an overview of all sticky notes and being able to structure them helps people to get a better understanding of the data. In their comments, people did not point out that the tagging had an influence on seeing the data from another angle as we had anticipated. The tags were instead seen as a fixed category equivalent to cluster names. In this sense, people found it problematic to define the tags before they had worked with the data.

5.4.2 Both Ways Can Help Form a Shared Understanding

The comments regarding shared understanding are divided. Some participants said the tagging result views helped them find common ground faster because they had the overviews and needed less discussion:

> I think it can demonstrate common ground very easily and doesn't lead to so much discussion about which post-it should go where. (T4P1)

> You see the most tagged post-its. This way you get a quick overview and gain faster common ground with the team members. (T4P3)

Others saw advantages in using the Synthesis Guide but did not really know why:

> Maybe it was just the example, but the clustering felt quite natural. We had the most important facts immediately. (T2P4)

> In the end it seemed to be clearer what the interviewees said and what the others thought about it. (T2P3)

Two participants also acknowledged the "important" category because sticky notes can be highlighted without a special reason:

> Especially the important tag is interesting. Because you sometimes have a feeling this is important but don't know why. (T5P1)

However, six participants had problems creating the tags and misunderstood them. They disliked being limited to four and also that they could not change them afterwards. Although they were able to ask their team members questions about the tags, they saw problems in interpreting them: "There was some confusion about the tag-categories" (T5P4), and this user feared it was "just the least common denominator". Other participants generally saw more advantages in the unstructured clustering condition. They felt that the information sharing and discussions were more vivid and personal. "It was much more organic and encourages dialog" (T5P2). Or overall:

> It felt like the team reached a better common understanding of the challenge even though we didn't talk about it like with the [system]. But the building of clusters seemed to give us better tools to share our understanding (T4P1).

In summary, we cannot say that the tagging generally helps to come to a shared understanding. For some teams it did, but for other teams clustering the sticky notes was more useful.

5.4.3 Participants Thought Structured Synthesis Was More Balanced

Several people pointed out that the Synthesis Guide showed the overall team opinion and involved everybody:

> The [system] makes it pretty clear what the team's opinion is, also the people who were not so "loud" give a good overview. (T2P2)

> More fair, everyone's opinion counts. (T5P1)

> Balances team members dominant vs. introverts as what's mostly considered is what EVERYONE agreed on. (T6P3)

Furthermore, they liked that people were not influenced by each other:

> People are not influenced that much by others because the rating was done secretly. (T1P2)

In the unstructured clustering condition they thought it was interesting to see the different clusterings from their team members and compare them to their own clusters:

> First you can cluster it your way and then see what the other team members came up with. (T4P3)

> You can cluster and think first on your own and create a picture in your mind, so you can discuss with the team better, because you already thought about it, and talk only about the essentials. (T4P4)

> You can really see how people work and how they organize their findings. (T2P4)

On the other hand, two people saw the danger that it was easier for a dominant person to take the lead:

> A dominant person can push her view of the topic harder when explaining her way of clustering. (T6P3)

> It is easy for somebody to take control of the process alone. (T2P2)

5.4.4 The Synthesis Guide Provides Guidance for New Users

The participants of the study had different levels of experience. In their comments after the test, people who had just finished design school pointed out that the Synthesis Guide was helpful. In the free-form responses, participants also commented that they liked the guidance of the tool: "With a program like this, it's more structured and always clear what to do" (T4P3). Several participants perceived the process as easier and more structured: “It is easier to concentrate on the individual post-its” (T6P4). “You are not overwhelmed” (T3P2). “The use of a proper interface to choose among the topics made them easier to visualize” (T3P1).

5.4.5 General Preferences

In the post-test questionnaire we asked the participants about their general preference, i.e. which way of doing the synthesis they preferred. From 24 participants, 12 participants chose the structured Synthesis Guide condition and 12 participants the unstructured clustering condition. These divided opinions can also be seen in the overall comments of users:

> So all in all the [system] saves you a lot of clustering and cluster-discussion time that you can spend later to create a better POV. For me, the [system] makes the process more based on individual ratio and choice, which I like a lot. (T2P3)

> The “tagging” method is efficient but makes synthesis very scientific. There could be a danger that people just go for insights that were very clear to categorize. (T3P4)

> There was a lot of guidance, but also the feeling that one loses information, e.g. if a category is missing. (T2P2)

> I don’t like the tagging, I like to see my clusters and to think while shifting the post-its around. (T4P4)

In summary, some people preferred the new guidance and tagging result views of the Synthesis Guide because it created an equally balanced process involving all team members and helped to reach a shared common ground. Other participants preferred the unstructured clustering condition because they could freely cluster the sticky notes as they liked and thereby get a better overview and common understanding with their team members.

5.4.6 Experts Evaluation

We also gave the POVs that the teams created to design thinking experts (with coaching experience) to let them evaluate the POVs according to three characteristics (insightful, actionable and overall) on 5-point Likert-scale questions. Additionally, the experts had to choose the best POV per team. Ten experts rated the twelve POVs that were created by the six teams. For analyzing the evaluation of the characteristics, we performed an analysis of variance with repeated measures with condition as a factor with six levels (for the six teams). We could not find significant differences between the two conditions. When choosing the best POV per team the experts preferred the POV of the Synthesis Guide condition in three cases: one time the POV of the unstructured condition and two times the result was undecided. On average, the structured condition got a score of 5.5 (SD = 2.07), the unstructured clustering 4.5 (SD = 2.07).

5.5 Limitations

To identify statistical differences between these approaches, we would need more participants. It would also be interesting to test with design thinking novices only, considering they need the most support during synthesis as we identified in the interviews. Additionally, it would have been advantageous to have teams that already know each other at least a bit. Both of these prerequisites were true for some participants and teams, but not for all, as the scheduling of the studies did not allow this.

6 Conclusion and Future Work

In this chapter we analyzed the difficulties of synthesizing user research data in order to make sense out of it as a team. Through interviews we found that synthesis is perceived as a stressful team process that is especially difficult for novices because it depends on experience and intuition. Furthermore, it is an ambiguous, nontransparent process that takes a lot of time.

Based on these findings, we wanted to create a tool that has the following features: helping to get a better understanding of the user research data, providing a more balanced team process where everybody is equally involved, helping to come to a shared understanding more easily, guiding through the process and assisting in "getting started".

We presented a "Synthesis Guide" which was aimed to achieve these objectives with a phase of working alone as well as the option to apply "perspectives" or "tags" to the collected sticky notes. In a controlled experiment we tested the Synthesis Guide, i.e. structured condition versus an unstructured clustering condition. We could not find significant differences between the conditions from questionnaire data. Therefore, we draw our conclusions from the subjective free form text answers of the 24 participants. We found that while tagging sticky notes does not help to get a better understanding, an individual clustering phase for each team member was greatly appreciated by the participants. Several users confirmed that the individual tagging helped to equally involve all team members in the process and see the different points of view. However, seeing different clusterings of each team member also gives insights into the views of the others. Regarding a shared understanding or common ground, participants were divided into two groups: some said the clusterings better support discussions and in the process of reaching a consensus view. Others preferred the tagging result views as the basis for discussions. Several participants liked the different steps of the Synthesis Guide and said that it helps in the beginning, makes the process easier and reduces the feeling of being overwhelmed.

In summary, we cannot say that we found an effective way that helps all design teams during information synthesis. However, as already shown in the interviews section, there are a lot of different needs and preferences based on the experience of the people involved. For design thinking novices and teams that do not know each other very well, the Synthesis Guide and its tagging functionality seem to be a good option for the beginning of the synthesis. This way, team members can get an idea of the opinion of others and make sure everybody is involved. In a second step, the tagging results views can be the basis for clustering in a way people are used to. For experienced and well-functioning teams, the Synthesis Guide is probably not necessary because they already know how to proceed and work with each other.

Regarding the tagging functionality, people should be allowed to create more tags. Maybe the system should also introduce more “meta tags,” such as Important, Surprising, or My Favorite, to let people highlight more “fuzzy” sticky notes. However, we still wish to emphasize the notion of a “perspective” that should make it necessary to see the information from a different angle. The tags were mostly interpreted as fixed cluster categories and therefore often seen as too rigid for the process. Our idea was to apply perspectives that intentionally change the point of view on the data. If every team member proposes one of these angles, people are forced to think in a way they may not have done until now. In the future, we have to think about how to convey this meaning of tags or perspectives (and which word to choose) and how to test its influence in this intangible phase of the design process.

References

Alexander C (1964) Notes on the synthesis of form. Harvard University Press, Cambridge, MA

Andrews C, Endert A, North C (2010) Space to think: large high-resolution displays for sensemaking. In: Proceedings of CHI ’10, pp 55–64

Bamford G (2002) From analysis/synthesis to conjecture/analysis: a review of Karl Popper’s influence on design methodology in architecture. Des Stud 23(3):245–261

Beyer H, Holtzblatt K (1998) Contextual design: defining customer-centered systems. Morgan Kaufmann, San Francisco, CA, p 472

Brown T (2009) Change by design: how design thinking transforms organizations and inspires innovation. HarperBusiness, New York, p 272

Chau DH, Kittur A, Hong JI, Faloutsos C (2011) Apolo: making sense of large network data by combining rich user interaction and machine learning. In: Proceedings of CHI ’11, pp 167–176

Cheng W-H, Gotz D (2009) Context-based page unit recommendation for web-based sensemaking tasks. In: Proceedings of IUI ’09, pp 107–116

Chilton LB, Little G, Edge D, Weld DS, Landay JA (2013) Cascade: crowdsourcing taxonomy creation. In: Proceedings of CHI ’13, pp 1999–2008

Corbin JM, Strauss A (1990) Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 13(1):3–21

Cropley A (2006) In praise of convergent thinking. Creat Res J 18(3):391–404

Curtis P, Heiserman T, Jobusch D, Notess M, Webb J (1999) Customer-focused design data in a large, multi-site organization. In: Proceedings of CHI '99, pp 608–615

Goodman E, Kuniavsky M, Moed A (2012) Observing the user experience: a practitioner's guide to user research, 2nd edn. Morgan Kaufmann, Waltham, MA, p 608

Gumienny R, Lindberg T, Meinel C (2011) Exploring the synthesis of information in design processes—opening the black-box. In: Proceedings of ICED '11, vol 6. pp 446–455

Harboe G, Minke J, Ilea I, Huang EM (2012) Computer support for collaborative data analysis: augmenting paper affinity diagrams. In: Proceedings of CSCW '12, pp 1179–1182

Hey JHG, Joyce CK, Beckman SL (2007) Framing innovation: negotiating shared frames during early design phases. J DesRes 6(1/2):79–99

Hey J, Yu J, Agogino AM (2008) Design team framing: paths and principles. In: International conference on design theory and methodology, pp 1–12

Hill AW, Dong A, Agogino AM (2002) Towards computational tools for supporting the reflective team. In: Artificial intelligence in Design '02, pp 305–325

Hinman R (2011) Getting to meaning through story. Expo. Magic Des. A Pract. Guid. to Methods Theory Synth, pp 67–75

Isenberg P, Tang A, Carpendale S (2008) An exploratory study of visual information analysis. In: Proceedings of CHI '08, pp 1217–1226

Judge TK, Pyla PS, McCrickard DS, Harrison S (2008) Using multiple display environments for affinity diagramming. In: Workshop on beyond the laboratory: supporting authentic collaboration with multiple displays at CSCW '08, pp 9–12

Klimoski R, Mohammed S (1994) Team mental model: construct or metaphor? J Manag 20 (2):403–437

Kolfschoten GL, Brazier F (2012) Cognitive load in collaboration–convergence. In: International conference on system sciences, pp 129–138

Kolko J (2011) Exposing the magic of design: a practitioner's guide to the methods and theory of synthesis. Oxford University Press, New York

Krippendorff K (2006) The semantic turn: a new foundation for design. CRC Press, Boca Raton, FL

Lawson B (2006) How designers think: the design process demystified, 4th edn. Elsevier, Oxford

Lim B-C, Klein KJ (2006) Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. J Organ Behav 27(4):403–418

Lloyd P (2000) Storytelling and the development of discourse in the engineering design process. Des Stud 21(4):357–373

Morris MR, Lombardo J, Wigdor D (2010) WeSearch: supporting collaborative search and sensemaking on a tabletop display. In: Proceedings of CSCW '10, pp 401–410

Naumer C, Fisher K, Dervin B (2008) Sense-making: a methodological perspective. In: CHI 2008 workshop Sense-Making, Florence, pp 1–5

Novak J (2007) Helping knowledge cross boundaries: using knowledge visualization to support cross-community Sensemaking. In: Proceedings of HICSS '07, pp 38–47

Oehlberg L, Roschuni C (2011) A descriptive study of designers' tools for sharing user needs and conceptual design. In: Proceedings of of ASME DETC '11, pp 199–208

Oehlberg L, Simm K, Jones J, Agogino A, Hartmann B (2012) Showing is sharing: building shared understanding in human-centered design teams with Dazzle. In: Proceedings of DIS '12, pp 669–678

Oehlberg L, Simm K, Jones J, Agogino A, Hartmann B (2012) Showing is sharing: building shared understanding in human-centered design teams with Dazzle. In: Proceedings of the designing interactive systems conference on—DIS '12, pp 669–678

Pirolli P, Card S (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of international conference on intelligence analysis

Qu Y, Furnas GW (2005) Sources of structure in sensemaking. In: CHI '05 extended abstracts on Human factors in computing systems—CHI '05, pp 1989–1992

Robinson AC (2008) Collaborative synthesis of visual analytic results. In: 2008 I.E. symposium on visual analytics science and technology, pp 67–74

Rogers Y, Sharp H, Preece J (2011) Interaction design: beyond human—computer interaction. Wiley, New York, p 585

Russell DM, Stefik MJ, Pirolli P, Card SK (1993) The cost structure of sensemaking. In: Proceedings of INTERACT '93 and CHI '93, pp 269–276

Schön D (1984) Problems, frames and perspectives on designing. Des Stud 5(3):132–136

Schumann J, Shih PC, Redmiles DF, Horton G (2012) Supporting initial trust in distributed idea generation and idea evaluation. In: Proceedings of GROUP '12, pp 199–208

Sharma N (2011) Role of available and provided resources in sensemaking. In: Proceedings of the 2011 annual conference on Human factors in computing systems—CHI '11, p 1807

Shrinivasan YB, van Wijk JJ (2008) Supporting the analytical reasoning process in information visualization. In: Proceedings of CHI '08, pp 1237–1246

Umapathy K (2010) Requirements to support Collaborative Sensemaking. In: CSCW CIS workshop, 2010

Wallace JR, Scott SD, Macgregor CG (2013) Collaborative Sensemaking on a digital tabletop and personal tablets: prioritization, comparisons, and tableaux. In: CHI '13, pp 3345–3354

Wright W, Schroh D, Proulx P, Skaburskis A, Cort B (2006) The Sandbox for analysis: concepts and methods. In: Proceedings of CHI '06, pp 801–810

Part III
Supporting Information Transfer

Embodied Design Improvisation: A Method to Make Tacit Design Knowledge Explicit and Usable

David Sirkin and Wendy Ju

Abstract We present a design generative and evaluative technique that we call embodied design improvisation, which incorporates aspects of storyboarding, Wizard of Oz prototyping, domain expert improvisation, video prototyping and crowdsourced experimentation to elicit tacit knowledge about embodied experience. We have been developing this technique over the last year for our research on physical interaction design, where practitioners often rely on subtle, shared cues that are difficult to codify, and are therefore often left underexplored. Our current technique provides an approach to understanding how everyday objects can transition into mobile, actuated, robotic devices, and prescribing how they should behave while interacting with humans. By codifying and providing an example of this technique, we hope to encourage its adoption in other design domains.

1 Introduction

> *Design, by definition, is. . .mostly tacit knowledge. It has to do with people's intuitions and harnessing the subconscious part of the mind rather than just the conscious. . .If you think about the structure of the mind, there just seems to be a small amount that is above the water—equivalent to an iceberg—which is the explicit part. . .If you can find a way to harness, towards a productive goal, the rest of it, the subconscious [understanding], the tacit knowledge, the behavior—just doing it and the intuition—all those, then you can bring in the rest of the iceberg. And that is hugely valuable.*
>
> *Bill Moggridge*
> *Co-founder of IDEO, Director of Cooper-Hewitt National Design Museum*
> *Ambidextrous Magazine interview, 2007*

Project PI: Larry Leifer

D. Sirkin (✉) • W. Ju
Center for Design Research, Stanford University, 424 Panama Mall, Stanford, CA 94305, USA
e-mail: sirkin@stanford.edu; wendyju@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_11, © Springer International Publishing Switzerland 2015

1.1 Embodying Design Thinking

One of the key challenges facing designers is to unlock the tacit understanding of how they believe things should be, so that these ideas can be shared, discussed, critiqued and eventually operationalized. Nowhere is this more difficult than in the design of physical interactions, where critical aspects of a design are often neither verbalized nor materialized. And yet physical movement, behaviors and gestures can be critically important in the design of everyday objects—of cars, of robots, of doors and drawers—where autonomous motion is increasingly being incorporated, and where inexpensive controllers and batteries enable products that can lock and unlock, open and close, move around, wave, hide—*act* on their own.

We propose that, on some level, designers intuit what should be designed, but at the same time, they need ways to elicit and elucidate that knowledge in ways that are actionable. How, then, can we help designers to understand, think through and evaluate interactions during their design process? We have been developing a novel method for *embodied design improvisation* that combines storyboarding, physical and video prototyping, Wizard of Oz techniques, and crowdsourced experimentation (Table 1) to both reveal and evaluate appropriate physical interactions. Our use of improvisation has been particularly crucial in designing machines and robots that employ physical interactions, because (a) we are drawing upon motions, gestures and interaction patterns that are most often implicitly, rather than explicitly, understood, (b) the design space of possible actions, mechanisms and relevant dimensions is vast, and (c) the cost in time, money and effort to build real functional systems to evaluate is high.

In exploring and integrating these methods for physical interaction design, we believe that we are also developing a more generalizable technique for drawing out the intuitive and tacit aspects of design thought and action, which need to be articulated, recorded, codified, transmitted and reused in order to reach their full potential.

2 Background

Our use of design improvisation has evolved over several projects, and has been adapted in a number of ways to suit different research constraints.

2.1 Entryways into Embodied Interaction

We first undertook the design improvisation technique in collaboration with Bjoern Hartmann and Leila Takayama, in a field study looking at how people responded to interactive doors that gestured in different ways at passersby (Ju and Takayama

Table 1 Phases of the embodied design improvisation process applied to physical interaction

Step	Activity	Purpose
Question	Identify design and research challenges/questions	Provide a guide for design activity
Storyboard	Sketch users, devices, behaviors and scenarios	Generate initial design concepts
Prototype	Develop physical instances of device look/feel	Test critical functions and usage
Improvise	Enact impromptu/typical interaction scenarios	Explore use cases in great depth
Video record	Record a few scenarios and establishing shots	Demonstrate device interactions
Crowdsource	Deploy video prototypes as web-based studies	Learn how people perceive scenarios
Lab/field study	Live, in-person tests of prototypes and scenarios	Confirm/extend web-based findings

2009). One of the limitations of the field studies was that they were necessarily between-subjects, and study participants, caught on their way between one place and another, often gave very brief answers to our survey questions. These difficulties led us to create video re-enactments of the door gestures that got the strongest reactions and to present these videos to online study participants for reaction and feedback (Fig. 1). The video studies correlated with the results of the field studies: the effect-sizes were smaller, but the differences more statistically significant.

2.2 Increasing Remote Presence

The next project where we made significant use of this technique was in the context of robotic telepresence. To understand the effect that physical movement—and the lack of movement—has on our understanding of a remote collaborator's communicative actions, we developed looks-like prototype robots by using a video camera in one hand, and gaffer's tape and wooden dowels in the other. The dowels became a control rig, which we attached to an articulating iMac G4 computer screen. We then experimented with natural and strange juxtapositions of on-screen and in-space actions by puppeting the robot screen and acting out different expressions. We video recorded behaviors that seemed natural to us at that moment. These early experiments gave way to shorter five-second clips showing consistent and inconsistent onscreen and in-space interactions which we used in crowdsourced studies deployed through Amazon's Mechanical Turk platform (Sirkin and Ju 2012). Whereas in the previous study we were able to compare our online video prototypes with live field tests, in this project we were able to compare how people responded these short isolated clips with longer videos depicting whole scenarios of functional robotic systems (Fig. 2).

2.3 Communicating Interior Processes

The importance of depicting context and scenario in investigating embodied design interaction was also a theme in the studies on the effect of performed forethought

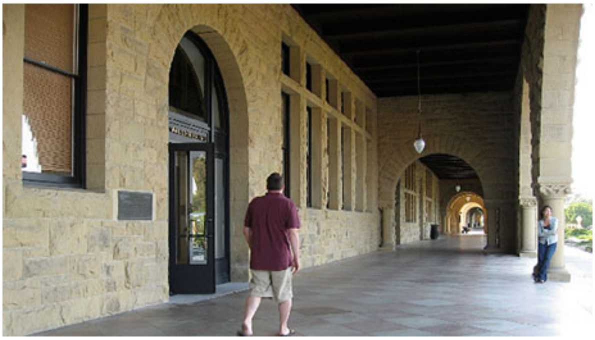


Fig. 1 Video prototype of the gesturing interactive door field study. The person standing on the right was a confederate, who signaled another confederate concealed just behind the entrance on the left. That person manually operated the door to simulate various expressive opening and closing behaviors

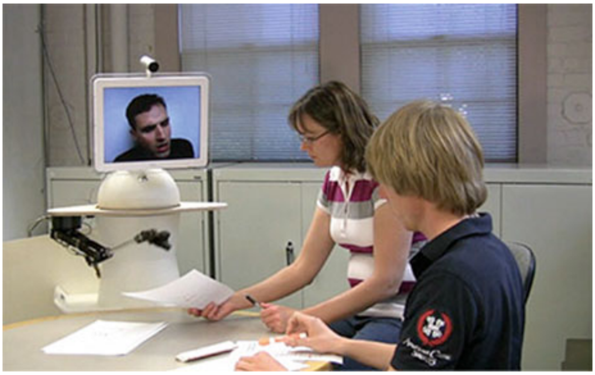


Fig. 2 Video prototype of the crowdsourced robotic telepresence study. An initial study found that perceptions improved when the remote person's onscreen actions and the robotic platform's in-space movements were consistent. The study used Wizard of Oz to actuate the robot: initially using manual levers, and later using remote teleoperation

and reaction in human-robot interaction (Takayama et al. 2011). These studies emerged from discussions of the expressive limitations of Willow Garage's PR2 robot that were observed by Pixar animator Doug Dooley. Because we expressly wanted to explore dimensions and actions that were outside of the capabilities of the actual PR2, we chose to convert our early brainstorming and physical enactments into short animated clips that depicted different scenarios in which the PR2 would perform motions that suggested contemplation and forethought (Fig. 3), and compared them to animations that showed what the PR2 actually did when it was sensing and computing prior to taking physical action—which was nothing.

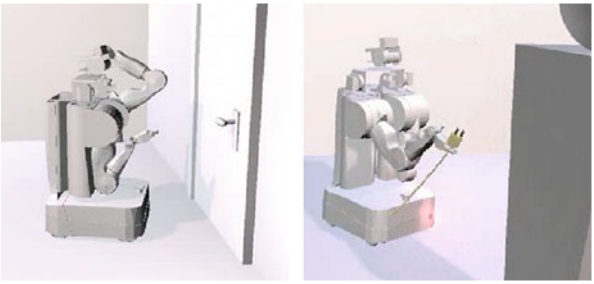


Fig. 3 Frames from an animated clip showing a Willow Garage PR2 robot contemplating its next action. A real PR2 has limited degrees of freedom that might express internal thought processes. The animation allowed us to explore alternative gestures, as well as compare these with the robot's actual movements

3 Design Process

Embodied design improvisation incorporates both divergent/generative and convergent/analytical techniques, and mixes in some classical interaction design methods. These include the use of concept videos like those of Apple (Dubberly and Mitch 1987) or the MIT Media Lab (Hoffman 2007) and video prototypes (Mackay 2002) as well as controlled experimentation (Newell and Card 1985) and crowdsourced evaluation (Kittur et al. 2008).

On the generative side, we develop simulations of would-be interactive devices by using basic puppetry and stagecraft techniques to set the context for the interaction. The process of storyboarding and blocking-out the interaction is loose and collaborative, allowing for brainstorming (Gerber 2009) and discussion of the many factors relevant to the interaction. We act out numerous scenarios (Hornecker 2005), using rapid prototyping (Hix and Hartson 1993) and Wizard of Oz techniques (Dahlbäck et al. 1993) to quickly create artifacts and to enable actions and reactions for our interactive devices. The outcomes of this research phase include physical artifacts as well as inventories of interaction stimuli and responses.

On the analytical side, we reflect specifically on ways to refine and crystallize the research questions and hypotheses that emerged from our earlier prototyping and playacting sessions. The far narrower set of designed interactions or scenarios that result from this analytical phase will fix certain aspects of the research and design, but also open questions that we expect to study further. It is at this point that we make more formal storyboards (Landay and Myers 1996) and hypothesize what variables and dimensions we will be testing in the eventual experiment (Nass and Mason 1990). We then shoot a set of video prototypes that capture key variations of our improvised designs, and generate and deploy web-based studies that allow us to get a read on how a wider swath of other people interpret our designs, and also to understand what the consequences of some of those design decisions might be.

Regarding the physical prototypes that lie at the center of our investigations, we are currently developing everyday objects that can sense and physically respond to

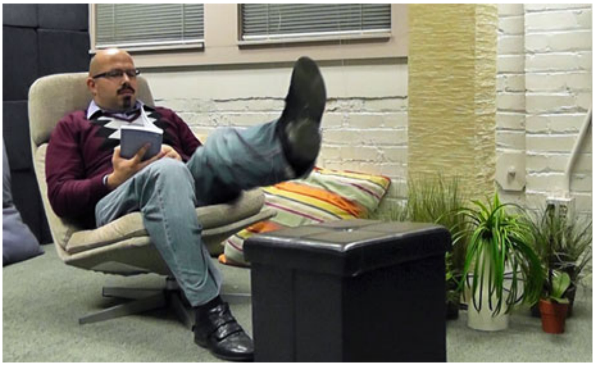


Fig. 4 The seated person is an actor and designer, who we invited to improvise with the robotic ottoman. He came up with several ways to call the ottoman over, including raising his leg, as we rolled the ottoman across the floor at different speeds and approaches. For the video, the mechanism to move the ottoman was concealed just below the frame

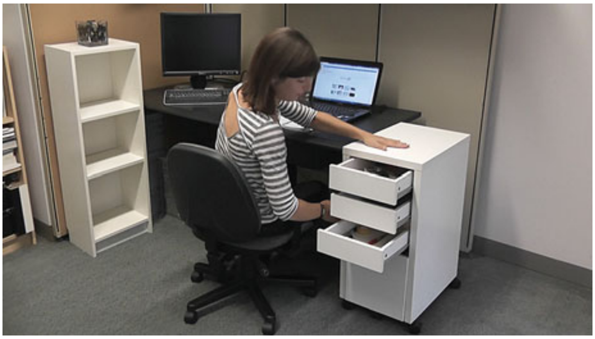


Fig. 5 The *top* three drawers were actuated, and controlled by a confederate across the room. We portrayed several interactions to explore how observers respond when the person's and the drawers' emotive behaviors were matched or mismatched (both expressing happiness here). This research question emerged from several rounds of improvising, prototyping and video recording

users' needs, and express their own intentions and emotions. The Mechanical Ottoman is a robotic footstool that interacts with users by offering or responding to request cues to place itself under their feet (Fig. 4). It can roll along the floor from across the room, rotate around itself to change direction, and lift itself up several centimeters as if poised to act.

The Emotive Robotic Drawers is intended to anticipate when users need to stow or retrieve small desktop items, and open or close in expressive ways that react to their emotional states (Fig. 5). It can open or close a specific drawer swiftly or lazily, synchronize the movement of several drawers, or even shudder as if frightened. These two prototypes represent initial forays into understanding and designing interactions between humans and *ubiquitous robotics* (Weiser 1991).

The following sections detail elements of our design process in the context of these two ongoing interaction studies. The activities are presented separately for clarity, although in practice, they overlap and may occur in some other sequence, depending on the project's design arc.

3.1 Identify a Research Question

When starting a design research program, it is important to identify one or more research questions to guide and focus efforts, even if the questions are only broadly defined at first. Research questions inform the selection of relevant interactions, participants, and the contexts in which these interactions occur. These circumstances, in turn, inform the types of research protocols to follow, the type of data that can be collected and analyzed, and the scope and validity of findings.

In particular, an open-ended approach can define a problem enough to specify initial contexts and scenarios to investigate, while leaving room for discoveries that can refocus later efforts. Research questions may therefore change over the course of a study, as investigators develop a deeper understanding of the issues involved.

With a design agenda in hand, benchmarking enabling technologies, observing individuals and environments, and brainstorming behaviors of interest can all be particularly helpful in generating ideas for what interaction scenarios to focus on. Such ideation methods should generally be considered team activities, whose value increases with several researchers involved per activity. It is important to document and record this process and its artifacts, as rich descriptions, sketches and recordings are often used in reports, and as pointers to when and where ideas initially emerged.

3.2 Storyboard People, Activities and Environments

Especially during the early stages of a design oriented research project, it can be valuable to find or create scenarios to explore and understand people, technologies and the interactions between them. The goal is to help reveal implicitly known, unstated, behaviors and understanding.

A storyboard is a visual narrative of an interaction scenario. With origins in cinema, it is a story-telling device that describes characters, the activities they engage in, the objects that they need and/or use, their motivations, emotions and reactions to interactions, and an environment for those interactions (Van der Lelie 2006).

We typically start a design project as a group huddled around a whiteboard with Post-It notes in hand, creating one or more storyboards, and blocking out the interactions that we want to explore (Figs. 6 and 7). At first, these storyboards serve as guides for how to enact scenarios or design prototypes. Later, they become

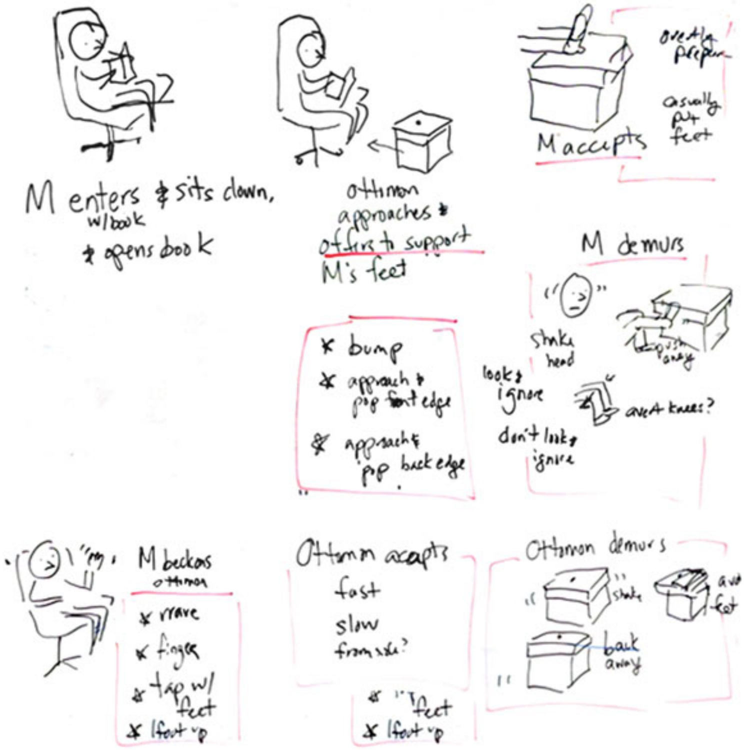


Fig. 6 Storyboard for the Mechanical Ottoman. The scene across the *top* shows our initial ideas about how the ottoman would offer to interact and the person would accept. The *middle and lower sections* show alternative ways that the ottoman could approach the person and offer to interact, or that the person could beckon it over, and how they both might accept or demur

archives of our initial thinking and indicators of how that thinking has evolved. By sketching as a group, we raise and challenge alternatives, often acting out interactions or device usage scenarios. This process unearths what individuals on the team implicitly know or anticipate about the interaction scenarios, so that others can reflect upon, and then extend or redirect, those ideas. By working at a rapid pace, at large visual and physical scale, in a situated environment, the team develops an embodied understanding of the problem (Klemmer et al. 2006; Wilson 2002) and alternative design approaches to explore it.

3.3 Prototype Technologies for People and Situations

While we discuss how interactions should unfold—usually concurrent with storyboarding—we prototype devices, technologies and situations. Prototyping

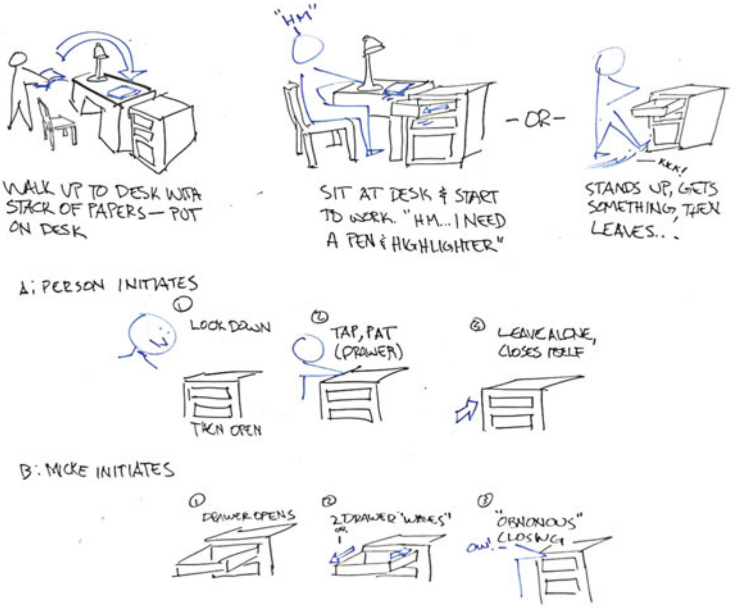

"HM"
-OR-
KKK!
WALK UP TO DESK WITH STACK OF PAPERS — PUT ON DESK
SIT AT DESK & START TO WORK. "HM...I NEED A PEN & HIGHLIGHTER"
STANDS UP, GETS SOMETHING, THEN LEAVES...
A: PERSON INITIATES
① LOOK DOWN
THEN OPEN
② TAP, PAT (DRAWER)
③ LEAVE ALONE, CLOSES ITSELF
B: MICE INITIATES
① DRAWER OPENS
② 2 DRAWER 'WAVES' OR 1
③ "OBNOXIOUS" CLOSING
OW! —

Fig. 7 Storyboard for the Emotive Robotic Drawers. Notice how the storyline across the *top* anticipates the observer's position and viewing angle used in subsequent video prototypes. The *middle and lower sections* show alternative ways that the person or drawers could initiate the interaction

and storyboarding inform each other, and alternating between them builds a deeper understanding of the design questions at hand (Beaudouin-Lafon and Mackay 2003).

At first, prototypes are hastily constructed: we recruit lamps, hand tools, cardboard boxes, furniture, or whatever else may be at hand as stand-ins for the features and functions that we need. For example, the first prototype of the Mechanical Ottoman was a half-meter square foam cube, while the first prototype of the Robotic Drawers was an IKEA mobile drawer set borrowed from a nearby office. We animate these prototypes through motion: steering objects around the floor by hand using parallel linkages made from broomsticks, lifting and lowering lids by tugging on barely-visible clear monofilament, or opening and closing drawers from behind using makeshift handles made of rolled-up gaffer's tape.

By improvising usage scenarios with these prototypes, we develop an initial set of functional requirements (such as *"the drawers should be able to open/close with variable speed"*) and design principles (*"objects should 'sit down' when they stop moving"*), which we use to purchase or construct more robust, useful prototypes, and thus iterate our way to improved designs. Over the course of iterating, these prototypes *should* begin to resemble useful devices, rather than rough-and-ready prototypes, to make sure that users or observers can focus on their designed features and functions rather than the artificial aspects of their construction.

3.4 Improvise Usage Scenarios with Experts

We recruit domain experts to help improvise these scenarios and enact our storyboards. Interactions depend strongly on context and use, so these improvisation sessions help us maintain a focus on prototyping interactions rather than devices (Brandt et al. 2012; Simsarian 2003).

In particular, we seek professionals from outside of our design context—actors, dancers, puppeteers, digital interface designers and roboticists—who offer diverse, yet deeply experienced and informed, perspectives. During any project, we conduct several improvisation sessions, each with one or (preferably) more artists and engineers present, as engaging several participants at once can raise questions and answers that might otherwise have gone unspoken. One example is whether the Mechanical Ottoman should be treated, or behave, as a trusted servant, or as an obedient pet.

By physically engaging these experts with specific design challenges *("shoo the ottoman away in as many ways as you can")*, and purposefully creating a playful environment (*"how would a silly, or hungry, or timid drawer open?"*), we hope to invite serendipity, and encourage them to open up, challenge our expectations, and reveal their own implicit understandings.

Improvisation sessions should be quite informal, although at times we may employ more structured warm-up exercises or techniques. These include (a) focusing on the obvious or immediate needs of the situation, (b) failing cheerfully by encouraging exaggeration, (c) telling stories about real or fictional characters and situations, and d) asking whether more or less detail is needed to understand the situation (Gerber 2007).

3.5 Record Video to Demonstrate Usage

Video prototypes are brief clips of how an interaction might take place. Much like storyboards, only with greater audio and visual fidelity, video prototypes help designers and researchers communicate—to others as well as themselves—the interplay between humans and novel technologies, within a specific context, over time. When combined with rapid prototyping and Wizard of Oz techniques (Mackay 1988), video prototypes allow design researchers to (a) explore and evaluate potential technologies—their roles, appearance and functions—without incurring the time and expense of building fully realized systems, and (b) distribute meaningful representations of these technologies to a much broader and more diverse population than the actual, physical prototypes and environments would allow. This is not to say that more fully realized systems are never built, just that their design extends across exploratory stages that help to get the *right design* (Buxton 2007).

By video recording improvisation sessions, we create a record of our expert interactions as well as (potential) initial video prototypes of these interactions. But most often, we recreate these scenarios more deliberately during one or more dedicated recording sessions, for which we recruit colleagues as onscreen actors and employ established videography techniques. Among these are use of an establishing shot (so the viewer understands the interaction context), multiple camera angles (typically one wide and another narrow), and framing scenes to conceal our Wizard of Oz manipulations (including linkages, cables, and humans-in-the-loop).

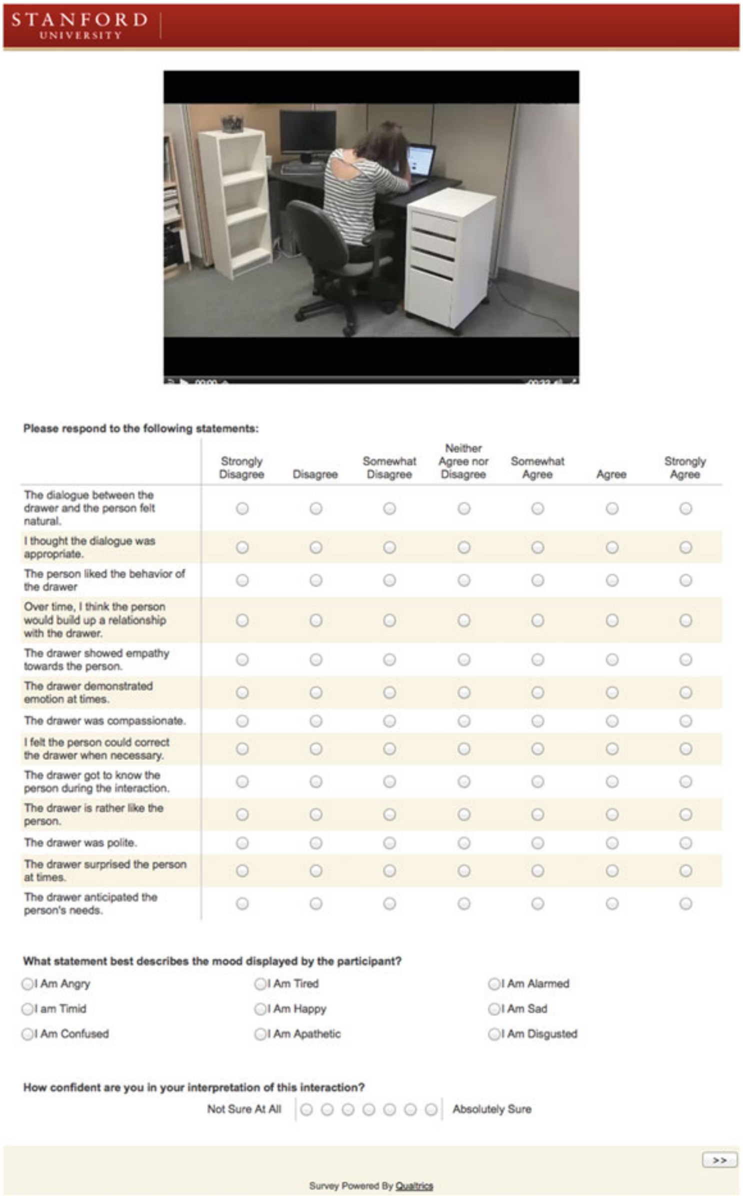
To help scenes convey the designed interactions, we create simple sets that include lighting, furnishings and props appropriate to that situation. At the same time, we are careful not to over-populate scenes in ways that distract from the important action. For example, scenes of the Mechanical Ottoman were shot in a lab corner with a lounge chair, floor lamp, cushions and houseplants (Fig. 4), while scenes of the Robotic Drawers were shot in an office with a work chair, desk, shelf and computer (Fig. 5).

We then edit this footage into several alternative clips, where each clip demonstrates a different device, behavior or scenario for observers to evaluate. Since these clips will serve as conditions in an online experiment, it is important to craft them so that only one of these variables of interest changes between conditions (Nass and Mason 1990).

3.6 Crowdsource Studies to Understand Broad Perceptions

We deploy video prototypes using Amazon's Mechanical Turk, a crowdsourcing platform that matches requesters having small online tasks to be performed with qualified workers. Each video is embedded within an online questionnaire that asks participants about aspects of the scenario before them (Fig. 8). We find sets of questions by searching prior literature, or we develop our own, to address research and design questions raised during prototyping and improvisation sessions. Finding measures within existing work has the advantage of drawing upon scales and analyses that have already been vetted by the community.

For the Robotic Drawers, more than just understanding whether one type of action/response was perceived as more appropriate than another, we wanted to understand whether the interaction between person and drawers would be perceived as a dialogue, and to what extent the drawers could project an emotional affect. This led us to a body of work on robotic conversational analysis (Benyon et al. 2008), which we adapted for our research context, and included in the online questionnaire. The resulting set of questions included, for example: *"how appropriate was the dialogue between person and drawers," "did the drawers show empathy toward the person," "are the drawers and person similar,"* and *"did the drawers and person get to know each other during the interaction?"*



Please respond to the following statements:

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
The dialogue between the drawer and the person felt natural.	○	○	○	○	○	○	○
I thought the dialogue was appropriate.	○	○	○	○	○	○	○
The person liked the behavior of the drawer	○	○	○	○	○	○	○
Over time, I think the person would build up a relationship with the drawer.	○	○	○	○	○	○	○
The drawer showed empathy towards the person.	○	○	○	○	○	○	○
The drawer demonstrated emotion at times.	○	○	○	○	○	○	○
The drawer was compassionate.	○	○	○	○	○	○	○
I felt the person could correct the drawer when necessary.	○	○	○	○	○	○	○
The drawer got to know the person during the interaction.	○	○	○	○	○	○	○
The drawer is rather like the person.	○	○	○	○	○	○	○
The drawer was polite.	○	○	○	○	○	○	○
The drawer surprised the person at times.	○	○	○	○	○	○	○
The drawer anticipated the person's needs.	○	○	○	○	○	○	○

What statement best describes the mood displayed by the participant?

○ I Am Angry	○ I Am Tired	○ I Am Alarmed
○ I am Timid	○ I Am Happy	○ I Am Sad
○ I Am Confused	○ I Am Apathetic	○ I Am Disgusted

How confident are you in your interpretation of this interaction?

Not Sure At All ○ ○ ○ ○ ○ ○ ○ Absolutely Sure

Survey Powered By Qualtrics

Fig. 8 Sample page from our online study on Emotive Robotic Drawers. Participants began by browsing a sample of the study to determine if they wanted to proceed, at which time they were redirected to our questionnaire, deployed on Qualtrics. Participants were paid through Mechanical Turk, commensurate with typical university experiment participant compensation rates

The use of Mechanical Turk—especially when combined with an online survey and analytics service such as Qualtrics—allows rapid iteration of study design: typically within a few hours. Often, before a study is complete, a cursory review of initial responses reveals how well participants understand the questions asked and the scenarios depicted, and suggests whether the study should proceed as it is, or be revised and redeployed. This revision might include re-recording video prototypes, altering the set of questions asked, or even revisiting the motivating research questions.

3.7 Conduct a Lab/Field Study to Confirm/Extend Findings

The use of video prototypes is but one tool in conducting robot interaction research (Woods et al. 2006; Ju and Takayama 2009). To confirm how well the perceptions of online observers agree with the experiences of physically co-present participants, we recreate, to the extent possible, the online study in our lab or a nearby environment (Odom et al. 2012). Alternatively, we may bring prototypes into the field to extend the findings from web-based studies. In either case, live interactions with functional robots requires much more robust construction than recorded interactions with Wizard of Oz prototypes, so we only begin lab or field studies once we build in to our devices the ability to move on their own. If we are able to do so early enough, field trials may begin before we have even recorded the video prototypes. For example, the Robotic Ottoman prototype is built upon an iRobot Create platform, so after only a few hours of software and network configuration, we could teleoperate the ottoman at a local coffee house.

For such impromptu field studies, we prefer to have in mind a detailed set of questions to answer, or tasks to perform, as this focuses our activity during the session and provides specific, actionable takeaways. For the coffee house study, these questions included the following: *"how many ways can the ottoman get someone to rest his/her feet (or drink) on it," "can the ottoman start a conversation between two strangers," "how would the ottoman calm someone,"* and *"how would the ottoman get someone to speak to it?"*

4 Design Domains and Impact

While the immediate application domain of our research is that of physical interaction design, the techniques we are using address the broader challenges of making tacit design knowledge explicit. As such, the approach can be employed by the wider design thinking research community.

Many of our techniques are, and have been, in use in pockets of design practice and research. However, the potential synergies between embodied action, improvisation, video prototyping and crowdsourced experimentation as yet remain unrealized. By adopting this method, designers begin to make explicit things that are known, but are difficult to articulate. They expose these understandings to open discourse, allowing them to operationalize the resulting insights into practices that can then be employed by others. They incorporate divergent and convergent thinking into their design research process: to help refine thinking, answer questions, test approaches and resolve questions.

The impacts of this work lie in (a) developing and systematizing our approach to design research, (b) making the approach actionable and available to the research community, and (c) exposing embodied design improvisation and video prototyping as valid and integral elements of the design research process. We expect that this work will provide a roadmap for researchers in other design domains to follow in their own explorations of embodied design thinking.

References

Beaudouin-Lafon M, Mackay W (2003) Prototyping tools and techniques. In: Sears A, Jacko J (eds) Human computer interaction—development process. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 122–142

Benyon D, Hansen P, Webb N (2008) Evaluating human-computer conversation in companions. In: Proceedings of the 4th international workshop on human-computer conversation

Brandt E, Binder T, Sanders E (2012) Ways to engage telling, making and enacting. In: Simonsen J, Robertson T (eds) Routledge international handbook of participatory design. Routledge, New York, pp 145–181

Buxton B (2007) Sketching user experiences: getting the design right and the right design. Morgan Kaufman, San Francisco, CA

Dahlbäck N, Jönsson A, Ahrenberg L (1993) Wizard of Oz studies—why and how. Knowl Based Syst 6(4):258–266

Dubberly H, Mitch D (1987) The knowledge navigator. Apple Computer Inc, Cupertino, CA

Gerber E (2007) Improvisation principles and techniques for design. In: Proceedings of CHI 2007, ACM Press, pp 1069–1072

Gerber E (2009) Using improvisation to enhance the effectiveness of brainstorming. In: Proceedings of CHI 2009, ACM Press, pp 97–104

Hix D, Hartson H (1993) Developing user interfaces: ensuring usability through product and process. Wiley, New York

Hoffman G (2007) Aur: a robotic desk lamp, Demo Video. MIT, Cambridge

Hornecker E (2005) A design theme for tangible interaction: Embodied facilitation. In: Proceedings of ECSCW 2005, Springer, pp 23–43

Ju W, Takayama L (2009) Approachability: how people interpret automatic door movement as gesture. Int J Des 3(2):1–10

Kittur A, Chi E, Suh B (2008) Crowdsourcing user studies with mechanical turk. In: Proceedings of CHI 2008, ACM Press, pp 453–456

Klemmer S, Hartmann B, Takayama L (2006) How bodies matter: five themes for interaction design. In: Proceedings of DIS 2006, ACM Press, pp 140–149

Landay J, Myers B (1996) Sketching storyboards to illustrate interface behaviors. In: Conference companion on human factors in computing systems: common ground, ACM Press, pp 193–194

Mackay W (1988) Video prototyping: a technique for developing hypermedia systems. In: Conference companion of CHI 1988, ACM Press

Mackay W (2002) Using video to support interaction design, DVD Tutorial, INRIA and ACM, New York

Nass C, Mason L (1990) On the study of technology and task: a variable-based approach. In: Fulk J, Steinfeld C (eds) Organizations and communication technology. Sage, Newbury Park, pp 46–67

Newell A, Card S (1985) The prospects for psychological science in human-computer interaction. Hum Comput Interact 1:209–242

Odom W, Zimmerman J, Davidoff S, Forlizzi J, Dey A, Lee M (2012) A fieldwork of the future with user enactments. In: Proceedings of DIS 2012, ACM Press, pp 338–347

Simsarian C (2003) Take it to the next stage: the roles of role playing in the design process, Ext. Abstracts CHI 2003, ACM Press, pp 1012–1013

Sirkin D, Ju W (2012) Consistency in physical and on-screen action improves perceptions of telepresence robots. In: Proceedings of HRI 2012, ACM Press, pp 57–64

Takayama L, Dooley D, Ju W (2011) Expressing thought: improving robot readability with animation principles. In: Proceedings of HRI 2011, ACM Press, pp 69–76

Van der Lelie C (2006) The value of storyboards in the product design process. Pers Ubiquit Comput 10(2–3):159–162

Weiser M (1991) The computer for the 21st century. Sci Am 265(3):94–104

Wilson M (2002) Six views of embodied cognition. Psychon Bull Rev 9(4):625–636

Woods S, Walters M, Koay K, Dautenhahn K (2006) Methodological issues in HRI: a comparison of live and video-based methods in robot to human approach direction trials. In: Proceedings of ROMAN 2006, IEEE, pp 51–58

Connecting Designing and Engineering Activities II

Thomas Beyhl and Holger Giese

Abstract Nowadays, innovation is an important competitive business advantage. Therefore, companies implement innovation processes or outsource them to external consulting companies. One example for such an innovation process is the methodology of design thinking, which enables the creation of innovative products or services. In Design Thinking an innovative product or service makes sense to people and for people, is likely to become a sustainable business model, and furthermore is functionally possible within the foreseeable future. Therefore, Design Thinking is considered as incubator for new innovative products and services. However, the transition from designing innovative products or services to implementing them is challenging since innovators and engineers are seldom the same people. This means a knowledge transfer between both groups is inevitable. As can be observed in practice, this knowledge transfer seldom goes smoothly since usually only the final innovative product or service is subject to the handover process. This is the case in spite of the fact that design decisions and the design path leading to this innovative outcome include important design rationales required by engineers. Thus, the design path and design decisions need to be recovered later on. We tackle this challenge with a manifold approach, which consists of (a) capturing design thinking artifacts, (b) inferring additional knowledge to recover the design path and design decisions, and (c) querying this knowledge. In this chapter we introduce our inference engine, which infers the design path and design decisions of Design Thinkers with the help of our Design Thinking inference rule set.

T. Beyhl • H. Giese (✉)
System Analysis and Modeling Group, Hasso Plattner Institute for IT Systems Engineering at the University of Potsdam, Prof.-Dr.-Helmert-Street 2-3, Potsdam 14482, Germany
e-mail: holger.giese@hpi.uni-potsdam.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_12, © Springer International Publishing Switzerland 2015

1 Introduction

Nowadays, innovation is an important competitive business advantage desired by different kinds of companies, e.g. small start-ups founded at the end of innovation processes, innovation consulting companies, and large enterprises implementing innovation processes by themselves. One example for such an innovation process is the methodology of Design Thinking (Plattner et al. 2009), which enables the creation of innovative products and services by in-depth end-user research, accompanied by the need for desirability, viability and feasibility (Brown 2009). Thus, the aim of the design thinking methodology is to yield products and services, which "*make sense to people and for people*" (desirability), "*are likely to become a sustainable business model*" (viability), and "*are functionally possible within the foreseeable future*" (feasibility), cf. (Brown 2009). Therefore, design thinking is considered an incubator for new innovative products and services (innovative ideas for short). However, the transition from designing an innovative idea to implementing this idea and bringing it on the market is challenging. Because innovators and engineers are seldom the same people a knowledge transfer between both groups is inevitable.

In our research project "Connecting Designing and Engineering Activities"[1] (Beyhl et al. 2013a, c) we observed that an information handover gap between innovation processes, e.g. design thinking, and engineering, exists. This is because innovators often only present the final innovative idea, demonstrate a final prototype (Fig. 1a), which usually only covers the most important issues of the overall idea, and hand over an informal document, which describes the final idea in isolation (Beyhl et al. 2013c). Unfortunately, the innovator's journey (Fig. 1b) leading up to this final idea is often neglected. It is particularly this journey which embodies important design rationales, product and service alternatives, end-user feedback, and accepted/rejected requirements. Consequently, innovators pass on their overall vision of what has to be built informally, while engineers rely on formal specifications about what needs to be build.

To be able to cope with this situation, engineers need to trace back through the innovator's journey to recover the design path and design decisions, which include design rationales, requirements and end-user feedback. Therefore, innovation processes need to support traceability (Beyhl et al. 2013c) (Fig. 1), i.e. "*the ability to describe and follow the life of a requirement in both a forwards and backwards direction*" (Gotel and Finkelstein 1997). However, as in other disciplines traceability is considered as beneficial only for others Arkley and Riddle (2005) and a tradeoff between the effort to create traceability information and the usefulness of this information has to be taken into account. Especially in design thinking, innovators might not know for the moment which information will become important later on.

[1] https://www.hpi.uni-potsdam.de/giese/projekte/dtr_connecting_designing_and_engineering_activities.html

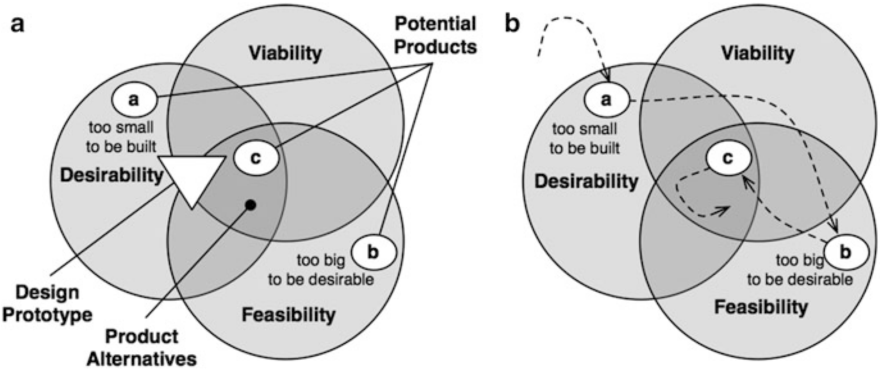


Fig. 1 Three dimensions of Design Thinking, cf. (Beyhl et al. 2013c; Brown 2009)

Since innovation processes need to support traceability, we mapped the concepts of signs, tracks and traces for requirements traceability introduced by Gotel et al. (Gotel and Morris 2011) to innovation processes based on the example of Design Thinking (Beyhl et al. 2013c). A sign is an "*identifying mark made by, or associated with for a particular purpose, an animate or inanimate object*"(Gotel and Morris 2011). "*A pattern of signs created as these signs are generated*" is a track, which can be traced later on by "*identifying a track following its pattern sign by sign*" (Gotel and Morris 2011). Figure 2 depicts these main concepts mapped to Design Thinking. Design thinkers primarily create analog artifacts, e.g. post-its, which are captured by digital artifacts, e.g. photographs, for documentation. These analog artifacts embody ideas, rationales and end-user feedback, which are primary parts of the Design Thinkers' journey. Thus, these artifacts are signed entities that need to be associated with artificial signs since they do not carry natural signs. Thereby, each sign substitutes the associated artifact for the purpose of artifact representation. During the Design Thinking process, Design Thinkers move these analog artifacts between different contexts, e.g. by clustering them concerning certain issues. This movement of analog artifacts between different contexts creates a track that represents the Design Thinkers' journey. In general, these different contexts are captured by taking snapshots of the working state at certain process milestones, e.g. when an activity ends. However, these snapshots, which represent the Design Thinkers' journey, are seldom used in the final documentation and instead the final innovative idea is described in isolation. In short, Design Thinkers are considered as sign and track makers. The primary user of these signs and tracks are engineers They trace in backward direction from the final documentation to earlier milestones to recover the design rationales neglected in the final documentation. The secondary trace users are Design Thinkers, who use the traces for synthesis and reflection. Thus, both, engineers and Design Thinkers, would profit from innovation processes which support traceability. Therefore, a suitable documentation platform with traceability support for innovation processes is required, which (a) assists Design Thinkers in making signs (capture), (b) uses these signs to establish tracks (inference), and (c) enables Design Thinkers and engineers to trace these tracks (query).

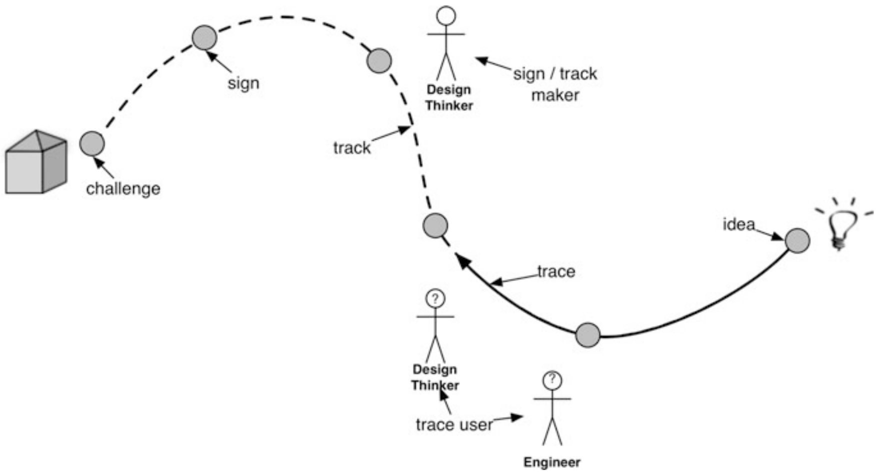


Fig. 2 The concepts of signs, tracks and traces mapped to Design Thinking

In this chapter, we present our documentation platform with traceability support for Design Thinking to capture, infer and query knowledge gained through Design Thinking. This documentation platform consists of a graphical user interface (GUI), an aggregator and an active repository. The GUI enables Design Thinkers to organize their artifacts (e.g. photographs of post-it clusters) captured by the aggregator, which loads these artifacts from arbitrary software tools (e.g. online storage services) used by Design Thinkers. The active repository is a storage whose content is steadily analyzed by an inference engine, which extracts knowledge embodied within the stored artifacts and exists between these artifacts for later reference. In other words, the inference engine pre-computes answers to potential queries stated by Design Thinkers and engineers later on. Thereby, the inference engine applies inference rules, which describe how to infer such pre-computed answers. We refer to these inference rules as a Design Thinking inference rule set.

This chapter is structured as follows: After a discussion of current documentation practices in innovation processes (Sect. 2), we introduce our documentation platform with traceability support for Design Thinking (Sect. 3). Afterwards, we introduce our inference engine and describe our Design Thinking inference rule set (Sect. 4). Finally, we discuss related work (Sect. 5) and outline future work (section “Conclusion and Future Work”).

2 Current Documentation Practices in Design Thinking

In practice, several application scenarios exist as to how Design Thinking and engineering are combined (Beyhl et al. 2012, 2013a). Figure 3 depicts these scenarios. Either (a) design thinking and engineering are decoupled,

(b) overlapped, or (c) (theoretically) applied concurrently. Commonly, enterprises outsource innovation processes to external innovation consulting companies. This leads to a clear separation of the innovation process and engineering (Fig. 3a). Therefore, an explicit knowledge transfer between both steps is required. In this *decoupled setting* the innovative idea is stable when the knowledge transfer takes place. Thus, it is known which kinds of engineers are required. On the other hand, this is a-priori unknown when an open-minded innovation process is kicked off. This setting is the common approach and is therefore considered as benchmark concerning progress speed, agility and documentation/communication effort. As is well known, the introduction of additional personal is not effortless and requires additional documentation and communication (Balzert 1997). Usually, innovation teams are a lot smaller than engineering teams. When innovation processes and engineering overlap (e.g. when the realization of a promising innovative idea is outsourced to start-ups, see Fig. 3b1), the members of the innovation team act as knowledge carriers. Thereby, they transfer the knowledge gained during the innovation process to the engineering team, which again requires documentation and communication. In a similar setting (depicted in Fig. 3b2) clients explicitly hire the involved design thinkers after the end of the Design Thinking project. However, also in this setting Design Thinkers have to look up design rationales themselves. Thus, the knowledge transfer/lookup is just shifted in the overall process. The overlapping of both steps is only beneficial when the innovative product or service idea is stable enough to avoid wasted engineering effort in case the innovation team changes their innovation scope. Thus, the overlapped setting does not provide advantages over the decoupled setting concerning progress speed, agility and documentation/communication effort.

While it is theoretically possible to apply innovation processes and engineering in parallel (Fig. 3c), in a practical sense it is unrealistic. This is because innovation teams are usually small, engineers usually focus on feasibility, which can be counter-productive (cf. brainstorming rule “defer judgement”), and engineering should not start before the innovative idea is stable. To sum it up, all three settings suffer from the same knowledge transfer challenge, but at different points in time. Consequently, documentation is inevitable and a need for a documentation platform with traceability support exists in all three settings.

We investigated how innovation processes are documented in educational and business settings. In business settings, e.g. at innovation consulting companies such as D-LABS,[2] documentation is created faithfully because professional innovators know they must rely on their documentation to create business value. In contrast, students in educational settings, e.g. at the HPI School of Design Thinking,[3] are usually faced with documenting their innovation process for the first time and regard documentation as worthless and obstructive. To guide students, templates about what to document are provided and milestone presentations are given by the

[2] http://www.d-labs.com/english/

[3] http://www.hpi.uni-potsdam.de/d_school/home.html?L = 1

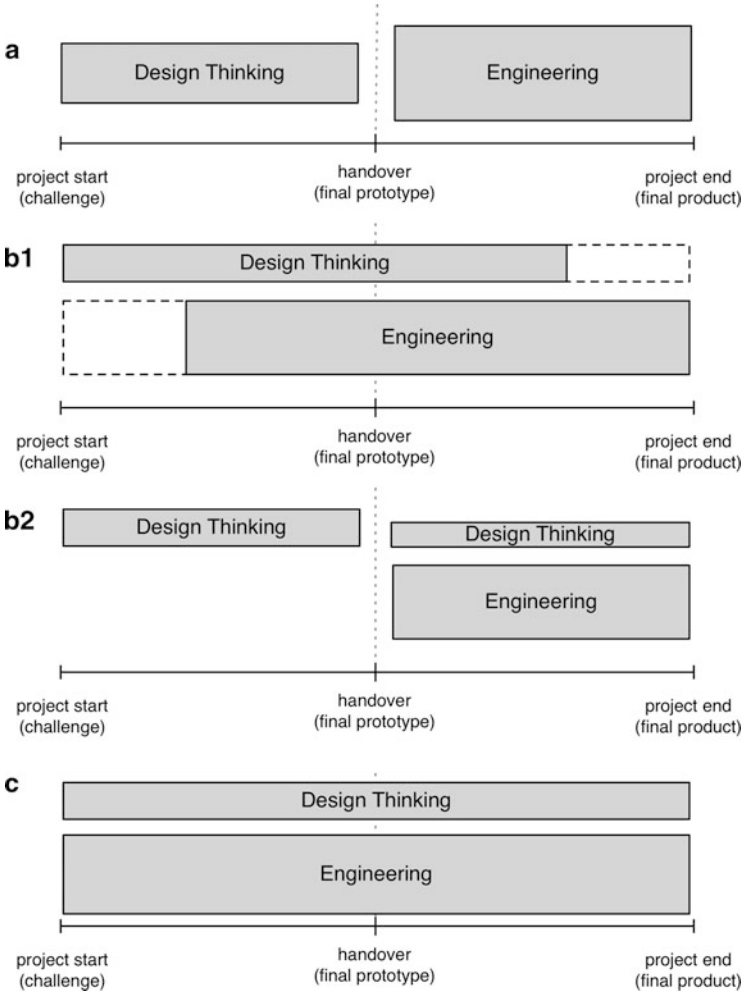
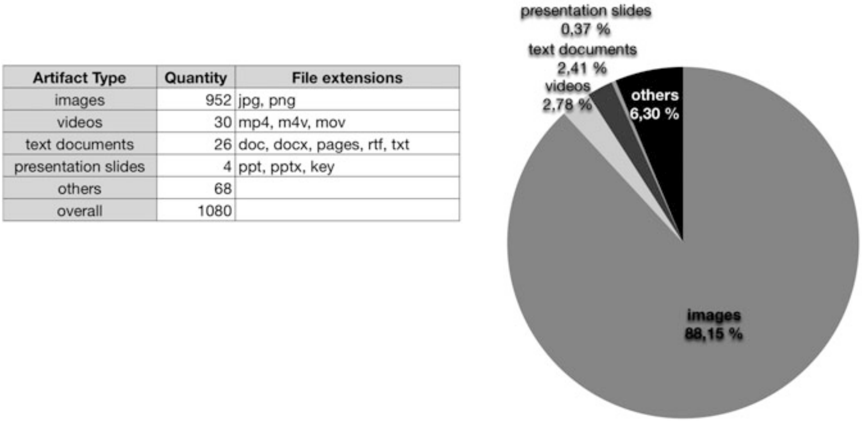


Fig. 3 Combination of Design Thinking and engineering: (**a**) decoupled setting, (**b**1) and (**b**2) overlapped setting, (**c**) concurrent setting, cf. (Beyhl et al. 2013a)

students to share their knowledge and progress. These templates are particularly applied as synthesis instruments at the end of each process step and indicate follow-up research directions. However, providing students with a well-structured template of what to document can be insufficient when they only look at this template *after* they have finished their work (Gabrysiak et al. 2012). Thus, the students are encouraged to document their journey by taking photographs of their post-it walls and prototypes. These collected artifacts are stored and processed with arbitrary software tools, e.g. online storage services. Figure 4 depicts which kinds of artifacts are used for documentation purposes in educational design thinking projects. We counted the number of artifacts from 16 educational design thinking 3-week



Artifact Type	Quantity	File extensions
images	952	jpg, png
videos	30	mp4, m4v, mov
text documents	26	doc, docx, pages, rtf, txt
presentation slides	4	ppt, pptx, key
others	68	
overall	1080	

Fig. 4 Overview of artifact types used in educational Design Thinking projects

projects at HPI School of Design Thinking. Documentation in educational settings mainly consists of images and seldom descriptive texts, making it even more difficult to look up important artifacts later on since these images are seldom renamed appropriately nor is their content easily searchable. While documentation in educational settings does not follow strict rules, documentation in business settings is subject to regulations. For example, in business settings often well-defined file systems and document structures are used, regular snapshots of post-it walls are taken, which are transcribed afterwards into text documents for later look-up. Furthermore, interview notes are taken in a pre-defined structured way and are immediately synthesized to minimize information loss. Thereby, different kinds of software tools are used, e.g. local file shares. In summary, both settings suffer from the same core problem of collecting all artifacts at one central searchable repository for an intelligent organization and later lookup of these artifacts.

3 Approach

We tackle this documentation challenge with a manifold approach. This consists of (a) *capturing* Design Thinking artifacts [cf. Beyhl et al. (2013b)], (b) *inferring* additional knowledge to recover design paths and design decisions, and (c) making it possible to *query* this knowledge. Figure 5 depicts the overall architecture of our documentation platform with traceability support. This platform consists of a *repository* that stores the artifacts fetched by an *aggregator* from several *storage* and *editing tools* used by Design Thinkers. Design Thinkers and engineers can organize these artifacts graphically using a stakeholder-specific *frontend* later on (Beyhl et al. 2013a, b). An inference engine enhances the repository content, i.e. captured artifacts and relationships between these artifacts. Therefore, the

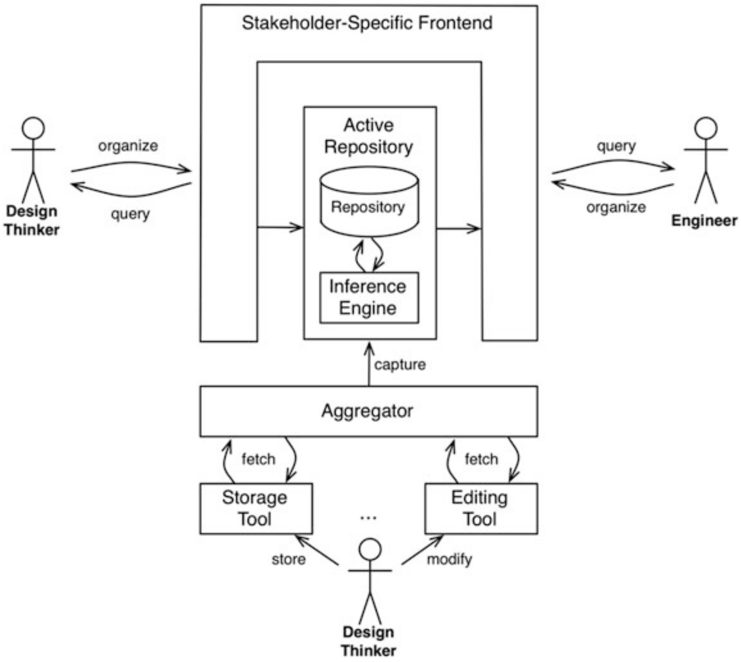


Fig. 5 Overall architecture of our documentation platform with traceability support

inference engine exploits *low-level properties* of raw documentation data, e.g. creation dates of artifacts, files names indicating process steps and activities, file and folder hierarchy, and the content of artifacts. These extracted low-level properties are combined into *high-level properties* by the inference engine, e.g. the identification of temporal clusters due to extracted creation dates of artifacts or the identification of handover artifacts between process steps and activities since these artifacts are assigned to different process steps or activities at the same time. To sum it up, the inference engine infers high-level properties, i.e. combined properties, from low-level properties, i.e. atomic properties, or already inferred high-level properties. We refer to this repository and inference engine as *active repository*.

In Sect. 3.1 we introduce Project-Zoom designed to support Design Thinkers in capturing and organizing their artifacts, which embody the knowledge gained during the innovation process. These artifacts are stored in an active repository (Sect. 3.2). This repository applies inference rules to make *implicit* undocumented, knowledge contained by and existing between these artifacts, *explicit*, i.e. documented.

3.1 Data Capturing and Visualization

The first part of the documentation challenge is how to capture and to collect artifacts at one central repository for further processing without changing the Design Thinkers' habits concerning documentation, i.e. how to capture and collect artifacts unobtrusively (Beyhl et al. 2013b). We investigated this challenge in cooperation with HPI School of Design Thinking and developed a software tool called Project-Zoom. Project-Zoom is a collaborative and intuitive application, which fetches artifacts, e.g. photographs of post-it walls and prototypes, text documents or presentation slides, from Design Thinkers' favorite storage and editing tools. While automatically creating a backup of these artifacts, different artifact versions are captured over time. Project-Zoom enables the innovators to organize these artifacts in a graphical manner as depicted in Fig. 6. The collected artifacts are depicted on the left- hand side and can be moved to the canvas on the right-hand side by drag & drop. Furthermore, these artifacts can be combined to clusters. Each cluster can belong to one of the six steps of the design thinking methodology. Additionally, links between artifacts can be added and comments can be attached to artifacts and clusters within the canvas. This kind of artifact organization and the meta data additionally captured by the aggregator and contained by the repository, e.g. creation dates of artifacts, file system structures and names, internal structure of artifacts etc., serve as basis for our inference system. Moreover, Project-Zoom can be the basis for additional techniques to capture additional meta data, such as mouse click frequency on artifacts, number of incoming and outgoing edges of artifacts [cf. (Voget 2013)], and a visual documentation language [cf. Forbus and Usher (2002)].

3.2 Active Repository

Design Thinkers are encouraged to document their design process in a reasonable manner, although it might be that it is unknown in the process which piece of knowledge will be important later on a-priori. Therefore, usually innovators explicitly document the most important insights from their current perspective, e.g. they move artifacts, which embody the most important insights, to process step clusters within Project-Zoom. While this explicit, i.e. documented, knowledge is visible to engineers and design thinkers later on, the implicit, i.e. undocumented, knowledge cannot be recovered ad-hoc. For that reason we apply an inference engine (Fig. 7), which executes an *inference rule set* to infer *high-level properties* from *low-level properties*. Low-level properties are *atomic properties*, which can be extracted directly from the artifacts *raw data*, while high-level properties are *combined properties*, which can be inferred by combining atomic properties or already inferred high-level properties. For example, a low-level property is a classification of an artifact concerning the process step it belongs to. When such an artifact can be

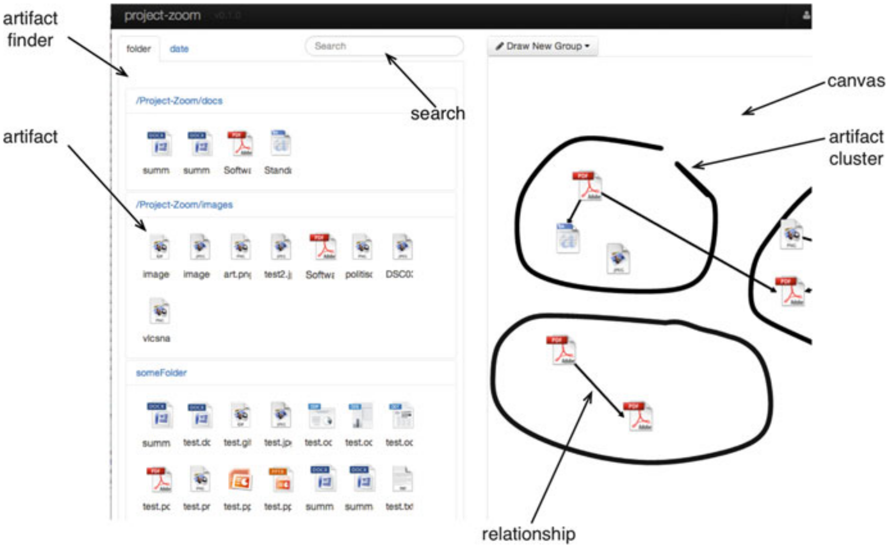


Fig. 6 Screenshot of Project-Zoom

assigned *with virtual certainty* to two different process steps it might be a handover artifact between these two process steps. It might include rationales for design decisions and next steps in the innovation process. An analyst, e.g. a domain expert, creates such inference rules. Thereby, these inference rules can search for well-known structures in the provided documentation data, extract meta data, or encode hypotheses in terms of finding new yet known structures in the documentation data. In the latter case, new inference rules can be derived, if the encoded hypotheses are accepted. Inference rules and their dependencies are described within an inference rule set.

Figure 8 depicts the ratio between raw Design Thinking documentation (input) and the number of answerable queries stated by Design Thinkers and engineers (output). While the ratio between input and answerable queries is *idealized* proportional with the current manner of documenting (q_{current}), because only the raw Design Thinking documentation is available for information lookup, the ratio between input and answerable queries (q_{inferred}) is expected to be significantly better when the inference rule set is applied. We assume that the enriched documentation contains more relevant knowledge. This makes it possible to answer more queries stated by Design Thinkers and engineers correctly in comparison to raw Design Thinking documentation. The raw Design Thinking documentation is considered as *inference benchmark* and we refer to the difference between the amount of raw Design Thinking documentation and enriched Design Thinking documentation as *inference performance*. The inference performance is more than the amount of inferred documentation since the inference process also has to infer the knowledge correctly. Raw Design Thinking documentation is considered as fundamental truth and therefore is always correct. This fundamental truth is used

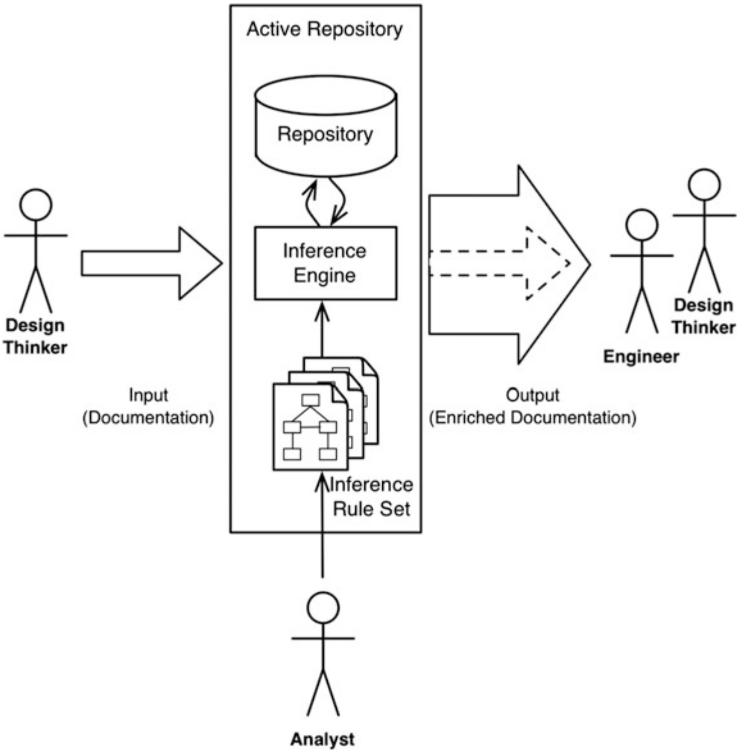


Fig. 7 Overview of active repository

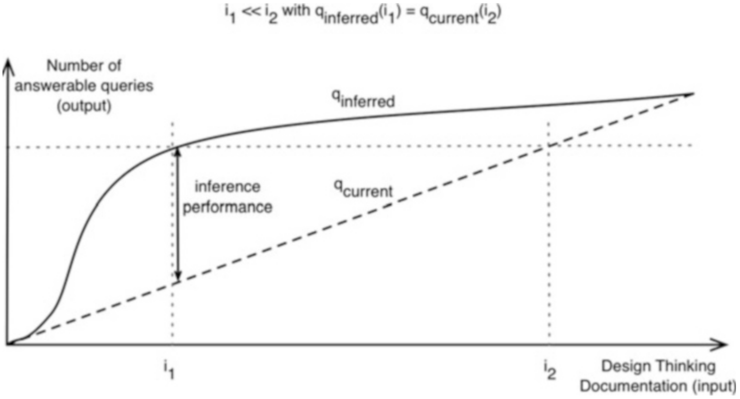


Fig. 8 Comparison of the ratios between answerable queries and raw (*dashed*) respectively inferred (*solid*) Design Thinking documentation data

for *inference performance assessment* in two ways (a) to evaluate which issues of explicit documentation can be omitted when documenting, because these issues can always be inferred with a high precision and (b) to evaluate which issues could be inferred additionally (cf. Fig. 7 solid arrow on right-hand side). In case (b) explicit evaluations whether these additionally inferred issues are correct or are not correct are required.

4 Inference System

In this section we introduce our inference system approach and describe our inference meta model in Sect. 4.1. It describes the main concepts of our inference models. In Sect. 4.2 we introduce our graphical notation for these inference models. We then introduce an exemplary Design Thinking inference rule set in Sect. 4.3 and apply this inference rule set retrospectively to educational design thinking project documentation in Sect. 4.4.

4.1 Inference Meta Model

In general, design thinking artifacts are heterogeneous and do not carry natural signs. This makes a generic processing of these artifacts a difficult task. Therefore, we introduce an *inference model*, which consists of elements (signs) that substitute Design Thinking artifacts for traceability and inference purposes (cf. Sect. 1). In scientific literature, this kind of model is called a mega model (Barbero and Bézivin 2008). Figure 9 depicts a simplified[4] version of our *inference meta model*, which describes the concepts of our *inference models*. The *Inference* class is a container for the following elements. The meta model consists of artifacts, i.e. representations of Design Thinking artifacts, and annotations, i.e. representations of pre-computed answers to (potential) queries stated by Design Thinkers and engineers. A*rtifacts* (cf. *Artifact* class) represent a physical artifact referenced by the *uri* attribute (i.e. file location) for inference purposes, while *annotations* (cf. *Annotation* class) represent knowledge inferred by inference rules. Artifacts and annotations consist of an artifact type, respectively an annotation type. An *artifact type* (cf. *ArtifactType* class) describes the type of an artifact, since this is necessary to deal with heterogeneous artifacts. Thus, the inference model is application domain independent. An *annotation type* (cf. *AnnotationType* class) describes the type of an annotation, since it is necessary to assign a semantic to each annotation in order to express what kind of knowledge is embodied within the annotation. Moreover,

[4] We omit artifact type and annotation type hierarchies. Every meta model element consists of a name attribute excepting the class *Attribute*. We have chosen to omit name attributes for clarity.

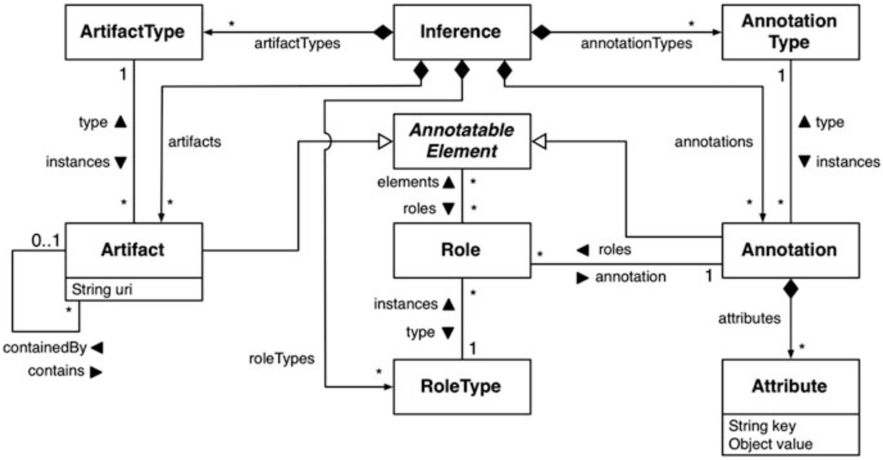


Fig. 9 Excerpt from our inference meta model (artifacts, roles, and annotations)

artifacts (cf. *contains* reference) are organized in a hierarchic structure. Artifacts and annotations are annotatable elements (cf. *AnnotatableElement* class) and can take on a certain role (cf. *Role* class) of a certain role type (cf. *RoleType* class). Not only artifacts can be part of an annotation scenario but also previously created (high-level) annotations can be part of such a scenario. Since annotations represent pre-computed answers to (potential) queries, they can consist of additional *attributes* to store arbitrary key-value pairs, e.g. belief values which describe the probability of whether the inferred annotation is correct or not.

The second part of our inference meta model is depicted in Fig. 10 and covers the *inference rule graph*. An inference rule graph (cf. *RuleGraph* class) consists of inference rules (cf. *Rule* class). An inference rule creates annotations between artifacts or annotations within the inference model, in the case this rule matches a certain pattern or if a certain property is fulfilled. An inference rule can be a *composed rule* (cf. *ComposedRule* class), which consists of additional inference rules. Moreover, rules that infer high-level properties depend on rules that infer low-level properties (cf. *dependencies* reference between *Rule* class). Thus, these inference rules constitute a directed acyclic graph (DAG), since cycles between rules are not allowed. Furthermore, inference rules consist of attributes that attach arbitrary key-value pairs, e.g. values describing the belief in an inference rule (i.e. general percentage of created true positive annotations) or the effort required to execute an inference rule.

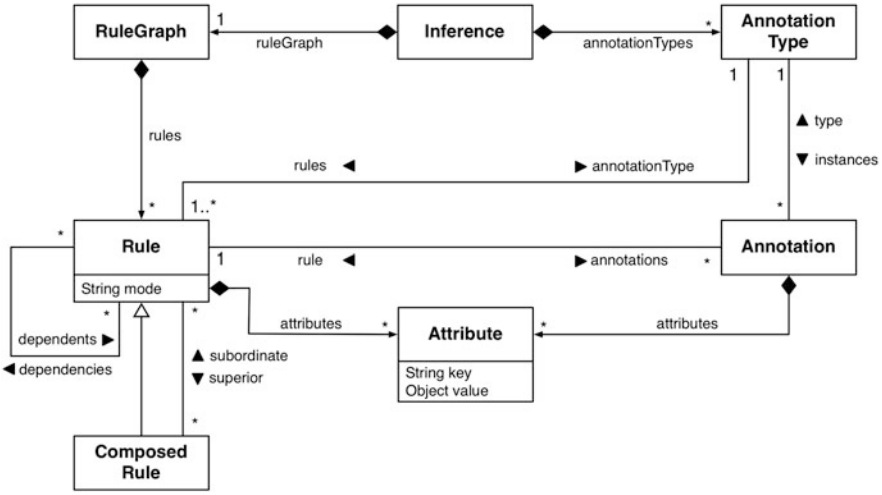


Fig. 10 Excerpt from our inference meta model (inference rule graph and inference rules)

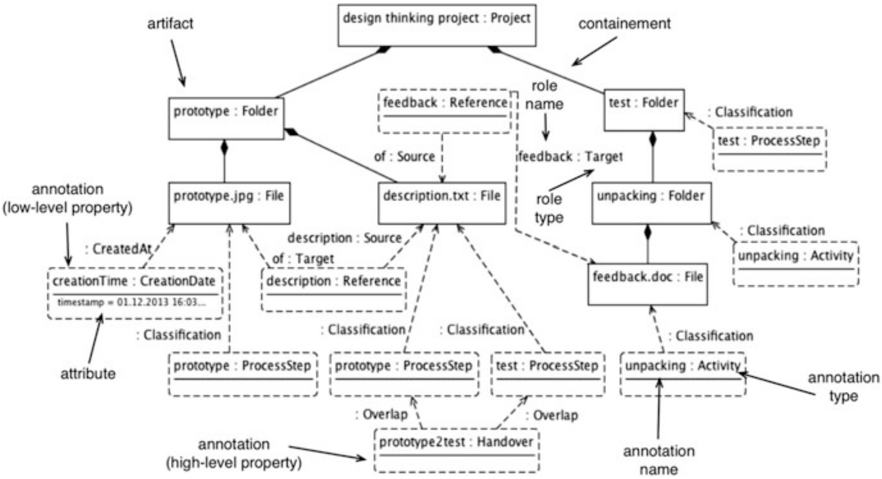


Fig. 11 Exemplary annotation scenario

4.2 Graphical Inference Model Notation

Due to the fact that inference models can become quite complex we developed a concrete syntax for inference models, i.e. a graphical notation, by extending the notation of Niere et al. (cf. Niere et al. 2002). The notation is based on UML object diagrams. Figure 11 depicts an exemplary annotation scenario. While elements with solid lines represent physical artifacts and their containment structure, elements with dashed lines depict inferred answers to potential queries. Artifacts are depicted by solid rectangles. The name of the artifact is followed by the name of the artifact type separated by a colon. Solid lines with rhombs as arrows depict the hierarchic structure of artifacts. The rhomb depicts the superior artifact. Dashed rounded rectangles represent annotations. The annotation name is followed by the name of the annotation type separated by a colon. Attributes attached to annotations are depicted in the compartment below the name of the annotation. Dashed arrows depict roles of artifacts and annotations in a certain annotation scenario. The name attached to dashed lines is composed of the role name and role type name separated by a colon.

4.3 Inference Rule Set

The main idea of our approach is the combination of inference rules (a) to build up on already inferred answers to potential queries and (b) to cope with the diversity of how design thinking projects can be documented by inferring the same kind of knowledge in different ways. Especially in case (b) evidence and contradictions can be identified, e.g. when different inference rules, which derive from the same kind of knowledge, infer the same piece of knowledge (or do not). Tables 1 and 2 describe exemplary inference rules whose dependencies are depicted by the inference rule graph in Fig. 12. Table 1 describes inference rules, which are applicable without making explicit traceability signs during the innovation process, i.e. these inference rules can be applied retrospectively on any kind of Design Thinking documentation. Table 2 gives an overview of Design Thinking inference rules, which explicitly require additional traceability signs made during the innovation process, i.e. these inference rules require additional documentation effort during the innovation process. In general, we distinguish between *atomic inference rules* and *composed inference rules*. Atomic inference rules are inference rules, which cannot be decomposed (e.g. R00a), while composed inference rules (e.g. R00) make use of other atomic or composed inference rules. In Fig. 12 rounded rectangles represent inference rules or composed inference rules respectively. Arrows depict the trigger direction, e.g. inference rule R00 triggers inference rule R01, R02 and R03. The inference rules R01, R02 and R03 are composed inference rules and therefore trigger the contained inference rules, e.g. the rule R01 triggers the rules R01a–R01c. We consider the following inference rules as hypotheses, i.e. as an initial Design Thinking inference rule set, derived from our observations of documentation practices in Design Thinking, cf. Sect. 2.

Each inference rule consists of a certain belief, i.e. a probability whether its inferred result is correct or not, cf. (Niere 2004). Furthermore each inference rule has a certain execution effort, e.g. execution time. Analysts, e.g. domain experts, initially define these belief and effort values. The ratio between both values should be considered during inference rule execution, e.g. in case a tradeoff between execution time and correctness of computation is required. For example, inference

Table 1 Design Thinking inference rules and their purpose *without* taking explicit signs into account

Rule name	Rule no.		Rule description	Rule purpose
Creation date extraction	R00	R00a	Extract creation date from file header	Creation dates can be used to aggregate temporal related artifacts. Different extraction methods should be applied to cope with the artifacts' diversity
		R00b	Extract creation date from file name	
		R00c	Extract creation date from parents' folder name	
Process step classification for artifacts (e.g. observe, understand, point of view, ideate, test, prototype)	R01	R01a	Process step classification based on artifact name	In general, it is important when an insight or findings was elaborated and therefore its importance can differ between different process steps. Different extraction methods can be applied whose result should be consolidated
		R01b	Process step classification based on the parents' folder name	
		R01c	Process step classifica-tion based on artifact content	
Activity classification for artifacts (e.g. synthesis, unpacking, fast-forward)	R02	R02a	Activity classification based on artifact name	In general, it is important when an insight or finding was elaborated and therefore its importance can differ between different activities. Different extraction methods can be applied whose result should be consolidated
		R02b	Activity classification based on the parents' folder name	
		R02c	Activity classification based on artifact content	
Temporal clustering (process step)	R03	R03a	Guess number of process steps, e.g. by counting the number of top-level folders	A reliable temporal clustering requires the number of expected process step clusters to yield good results
		R03b	Detect temporal clusters based on the creation dates of artifacts and the number of expected process step clusters	The importance of gained insights and findings differ between process steps. Moreover, knowledge about concrete process steps enables the application of process step specific inference techniques
Cluster classification (process step)	R04		Infer the process step for each temporal cluster by combining the information about temporal clusters, creation dates of artifacts and process step information attached to artifacts	It is expected that each cluster represents a certain process step. Successive clusters with the same process step information attached can be consolidated later on

(continued)

Table 1 (continued)

Rule name	Rule no.		Rule description	Rule purpose
Detect handover artifacts between process steps	R05		Conclude handover artifacts between different process steps based on artifacts that are assigned to different process steps simultaneously	Handover artifacts may contain design rationales for design decisions and next process steps
Detect handover artifacts between activities	R06		Conclude handover artifacts between different activities based on artifacts that are assigned to different activities within the same process step simultaneously	Handover artifacts may contain rationales for design decisions and next process steps
Create temporal order of clusters (process step)	R07		Derive temporal order of process step clusters	Artifacts at the connection of two or more process step clusters may be handover artifacts between process steps, which consist of important design rationales
Temporal clustering (activity)	R08	R08a	Guess number of activities per process step, e.g. by counting the number of sub folders	The number of activities per process step is required for a reliable activity clustering within process steps
		R08b	Detect temporal cluster based on the creation dates of artifacts and the number of expected activity clusters	The importance of gained insights and findings differ between activities, e.g. it makes a difference whether an insight is the result of an observation or of a prototype testing session. Moreover, information about the applied activity/framework may enable assumptions about the content and structure of artifacts created during this activity
Cluster classification (activity)	R09		Infer the activity for each temporal activity cluster within a process step cluster by combining the information about temporal clusters, creation dates of artifacts and activity information attached to artifacts	It is expected that each temporal cluster represents a certain activity. Successive time clusters with the same activity information attached can be consolidated later on

(continued)

Table 1 (continued)

Rule name	Rule no.	Rule description	Rule purpose
Create temporal order of clusters (activity)	R10	Derive temporal order of activity clusters	Artifacts at the connection of two or more activity clusters may be handover artifacts between activities, which consist of important design rationales
Process step classification from activities	R11	Infer process step from the activities within the process step temporal cluster due to the assumption that certain activities are more likely common practices for a certain process step than for another	This is an alternative method to infer information about an applied process step. This inference rule might give evidences or contradictions for already inferred knowledge
Detect handover artifacts	R12	Artifacts in the outer zone of process step and activity clusters may be handover artifacts	Handover artifacts are more important than other artifacts, because they are assumed to contain design rationales
Find evidence and contradictions for handover artifact classification	R13	Rule R11, R06 and R05 provide different methods to derive the same kind of knowledge. This rule looks up evidences and contradictions concerning the classification of artifacts as handover artifacts	Evidence/contradictions for classifications as handover artifacts can increase/decrease the overall belief in the classification result
Lookup similar artifacts	R14 R14a	Look up equal artifacts in different formats, e.g. feedback.docx con-forms feedback.pdf	The properties inferred for one artifact may be inherited to another similar artifact or version of the artifact
	R14b	Detect artifact versions, e.g. feedback_v1.pdf, feedback_v2.pdf, etc.	
Inherit inferred knowledge to similar artifacts	R15	Inherit detected properties of similar artifacts in different file formats respectively artifacts in different versions	This can reduce computation effort and may be a workaround for artifacts where a rule application is technically difficult, e.g. proprietary file formats cannot be parsed for analysis

(continued)

Table 1 (continued)

Rule name	Rule no.	Rule description	Rule purpose
Create temporal order of artifact versions	R16	Derive temporal order of versioned artifacts and derive the differences between these artifacts	The temporal order and differences between these artifact versions reflect the Design Thinkers journey. Especially the differencing between both versions may support cognitive processes
Rate artifact importance via artifact type	R17 R17a	Infer artifact importance from file type, e.g. presentation slides may represent milestones	More important artifacts should be ranked higher than less important artifacts in query results. They can serve as entry point for backward traceability
Perform OCR	R21	Perform optical character recognition (OCR) to extract textual content embedded in images	The textual content may be used to look up keywords associated with certain process steps or activities. The textual content can be exploited by inference rules such as R01 and R02

rule R00a receives a high level of belief, because it can be assumed that the creation date of artifacts is always set correctly by software tools and photo cameras. It is also an inference rule with a high execution effort because the artifact needs to be loaded from disk and needs to be parsed to extract the creation date. Furthermore, due to inference rule dependencies the belief values should be propagated between inference rules appropriately, e.g. when the number of process step clusters is guessed incorrectly by inference rule R03a the higher-order inference rule R03b, which clusters the artifacts due to their creation dates, will not return the appropriate result. Consequently, in this case the belief in the inference result of the higher-order inference rule R03b would then be degraded as well.

4.4 Proof of Concept

We implemented a proof-of-concept prototype and an initial subset of the presented Design Thinking inference rule set. We applied this inference rule subset retrospectively to four selected Design Thinking project documentations of the HPI School of Design Thinking (two 12-week and two 3-week projects) and were able

Table 2 Design Thinking inference rules and their purpose *with* taking explicit signs into account

Rule name	Rule no.		Rule description	Rule purpose
Rate artifact importance via number of edges	R17	R17b	Infer artifact import-ance from the number of incoming and outgoing connections of artifacts in Project-Zoom, cf. [Voget 2013]	It is expected that important artifacts have more incoming and outgoing edges than less important artifacts. More important artifacts should be ranked higher than less important artifacts in query results. They can serve as entry point for backward traceability
		R17c	Infer artifact importance from the number of mouse clicks on and downloads of artifacts	
Detect graphical patterns	R18		Infer sketch meaning from low-level glyphs, cf. (Forbus and Usher 2002). An own sub inference rule set is imaginable	Exploit the visual language of design thinkers
Extraction of semantic signs	R19	R19a	Look up photographs of certain design thinking frameworks via artifact name, e.g. idea dashboard, LogCal, LogBook	Detection of framework artifacts to exploit their inner structure later on
		R19b	Look up whiteboard photographs with certain QR code in it	Enrich photographs with semantic, e.g. photograph of whiteboard with brainstorming results
Exploit artifact structure	R20		Exploit the inner structure of well-known kinds of artifacts, cf. R19, and attach semantic information to each structure element, e.g. image of prototype or purpose of prototype	Structured artifacts, e.g. idea dashboard, define semantic units whose content is an image, text or both. These semantic units can be used later on without the need to interpret the text or image semantically

to detect low-level properties and infer high-level properties from these low-level properties. Table 3 gives an overview of the inference rules applied and how *many* annotations were created by each inference rule. However, whether our Design Thinking inference rule set has a good precision and recall is beyond the scope of our investigation and subject to future work, since no gold standard is available yet. Moreover, inferred high-level properties depend on detected low-properties and thus the precision and recall of low-level inference rules have an impact on the precision and recall of high-level inference rules. Therefore, we present a qualitative analysis of the inference results by giving some examples of inferred answers to potential queries stated by Design Thinkers and engineers.

Figure 13 depicts an excerpt of the execution result of inference rule R01d, which provides evidence for the assignment of the artifact “POV, basta.docx” to the process step Point of View, because this artifact was assigned to the process step

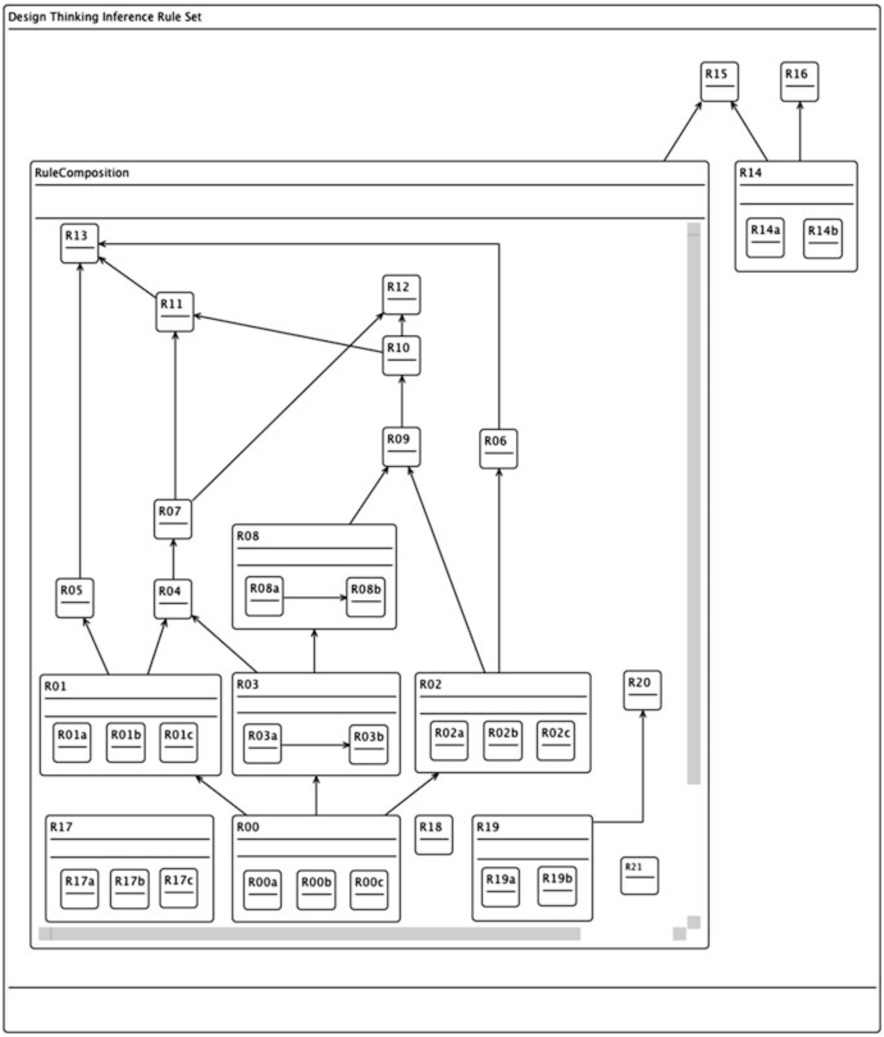


Fig. 12 Design Thinking inference rule graph

Point of View by inference rule R01a (via artifact name) and R01c (via artifact content).

Figure 14 depicts an excerpt of the execution result of inference rule R01 and R05. The artifact "POV, final docu.docx" was assigned to the Point of View and Ideate process step by inference rule R01a (via artifact name) and R01c (via artifact content). These low-level properties were used by inference rule R05 to conclude that this artifact is a handover artifact. This artifact contains the sentence: "*Building on this, we came up with our Persona Carola, which we used in the ideation process to find solutions*." This indicates the correctness of the inferred handover artifact

Table 3 Overview of inference results (number of created annotations)

Rule no.		Project #1 (12-week project, 651 artifacts)	Project #2 (12-week project, 596 artifacts)	Project #3 (3-week project, 157 artifacts)	Project #4 (3-week project, 150 artifacts)
R00	R00a	583	512	131	121
	R00b	488	216	27	80
	R00c	457	127	0	0
R01	R01a	16	9	5	3
	R01b	20	11	76	55
	R01c	6	24	0	2
	R01d	1	5	0	0
R02	R02a	27	18	8	8
	R02b	20	18	129	83
	R02c	7	24	0	5
	R02d	1	5	0	2
R03	R03b	18	18	8	8
R04		7	10	3	5
R05		1	0	0	0
R06		10	4	9	0
R14	R14a	0	78	3	6
R17	R17a	4	1	0	3
R19	R19a	0	1	1	1
R21		157	78	42	34

We defined the overall number of clusters as 18 or 8 respectively Inference rule R03 is required because inference rule R04 depends on inference rule R03

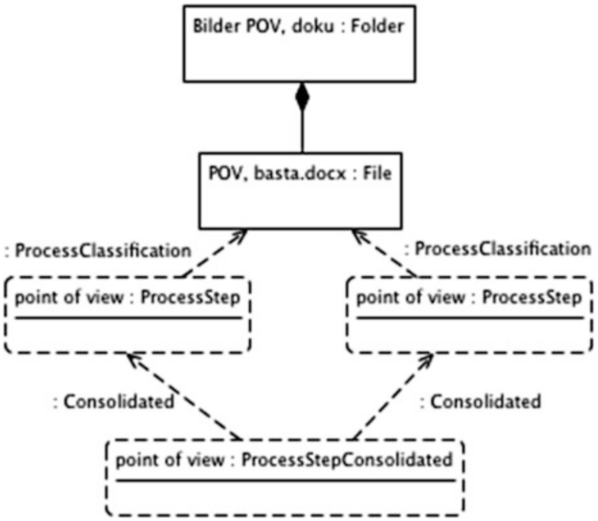

Bilder POV, doku : Folder
POV, basta.docx : File
: ProcessClassification
: ProcessClassification
point of view : ProcessStep
point of view : ProcessStep
: Consolidated
: Consolidated
point of view : ProcessStepConsolidated

Fig. 13 Example of evidenced process step knowledge

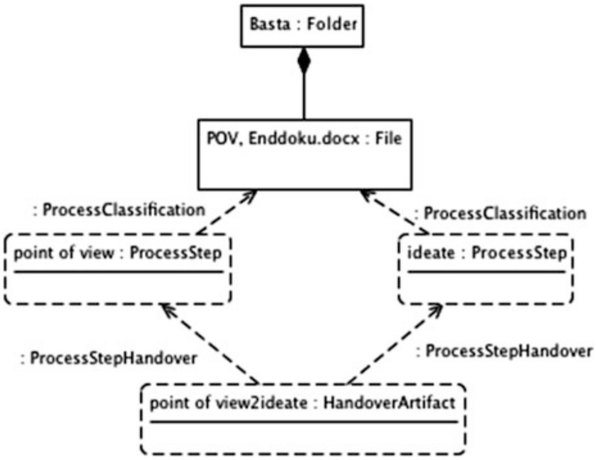


Fig. 14 Example of an identified handover artifact between process steps

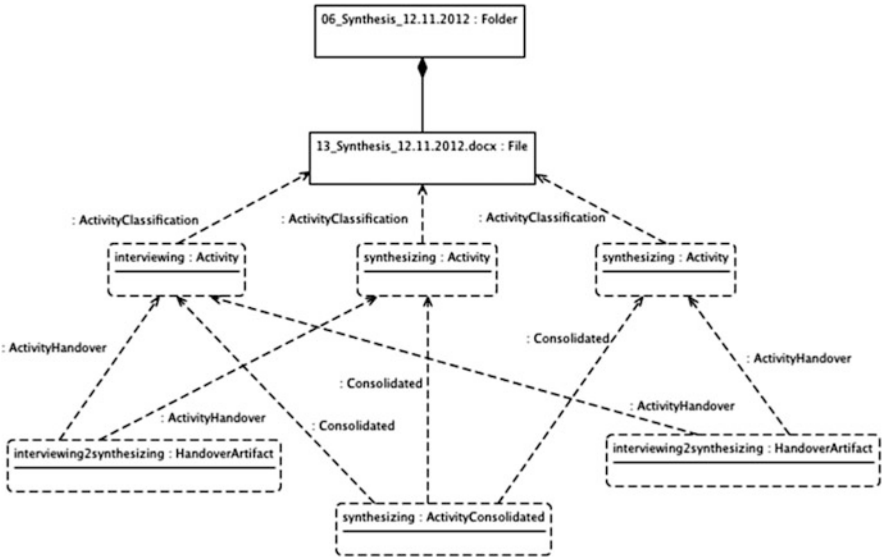


Fig. 15 Example of an identified handover artifact between an interview and synthesis activity

property together with the artifact name "POV, final docu.docx" (translated from German). This piece of knowledge can be the basis for additional inference rules, which take the internal structure of artifacts into account to infer which paragraphs contain design rationales that are important for the handover between both process steps.

Figure 15 depicts an example of an identified synthesis document, which was classified as a handover artifact between the activity interview and synthesis by

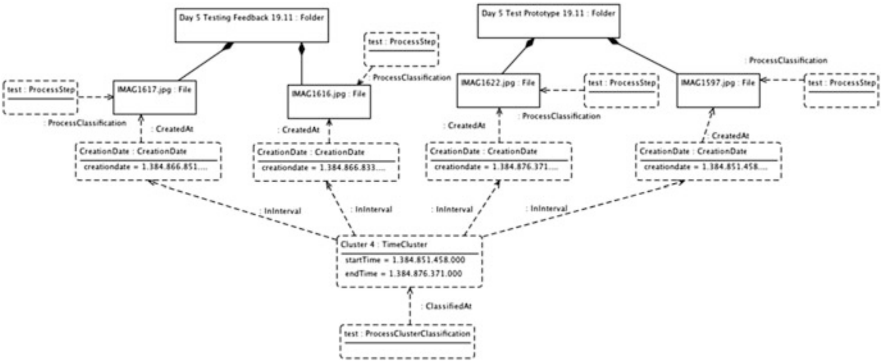

Day 5 Testing Feedback 19.11 : Folder
Day 5 Test Prototype 19.11 : Folder
test : ProcessStep
test : ProcessStep
test : ProcessStep
test : ProcessStep
: ProcessClassification
: ProcessClassification
: ProcessClassification
: ProcessClassification
IMAG1617.jpg : File
IMAG1616.jpg : File
IMAG1622.jpg : File
IMAG1597.jpg : File
: CreatedAt
: CreatedAt
: CreatedAt
: CreatedAt
CreationDate : CreationDate
creationdate = 1.384.866.851....
CreationDate : CreationDate
creationdate = 1.384.866.833....
CreationDate : CreationDate
creationdate = 1.384.876.371....
CreationDate : CreationDate
creationdate = 1.384.851.458....
: InInterval
: InInterval
: InInterval
: InInterval
Cluster 4 : TimeCluster
startTime = 1.384.851.458.000
endTime = 1.384.876.371.000
: ClassifiedAt
test : ProcessClusterClassification

Fig. 16 Example of a detected temporal cluster, which was classified as a Test process step

inference rule R06. Inference rule R06 concluded this piece of knowledge based on the assignment of the synthesis document to the interview and synthesis activity inferred by inference rule R02c (inferred interview activity) and R02a or R02b respectively (inferred synthesis activity on two different ways). Moreover, inference rule R02d concluded that the synthesis document belongs more likely to a synthesis than an interview.

Figure 16 depicts a set of images files. These files were assigned to the same temporal cluster by inference rule R03b, since the creation dates of these images extracted by inference rule R00a are temporally related to each other. Moreover, the images were assigned to the Test process step by inference rule R01b (via artifact hierarchy). This was in turn exploited by inference rule R04 to conclude that the whole temporal cluster represents a Test process step.

Figure 17 depicts an example of similar artifacts in different formats. While a creation date could be extracted for artifact “gr8-questions.pdf” by inference rule R00a, no creation date could be extracted for the similar artifact “gr8-questions.rtf”. The extracted creation date for the PDF version of this file leads to a temporal cluster inferred by inference rule R03b, which was classified as Understand process step by inference rule R04. This piece of information could not be inferred for the RTF version of this artifact, since no creation date could be extracted. However, inference rule R14a concluded that the PDF and RTF versions of this artifact are similar artifacts in different formats. This information can be exploited to inherit the properties inferred for the PDF version of this artifact to the RTF version, i.e. the RTF version of the artifact also belongs to the inferred Understand process step.

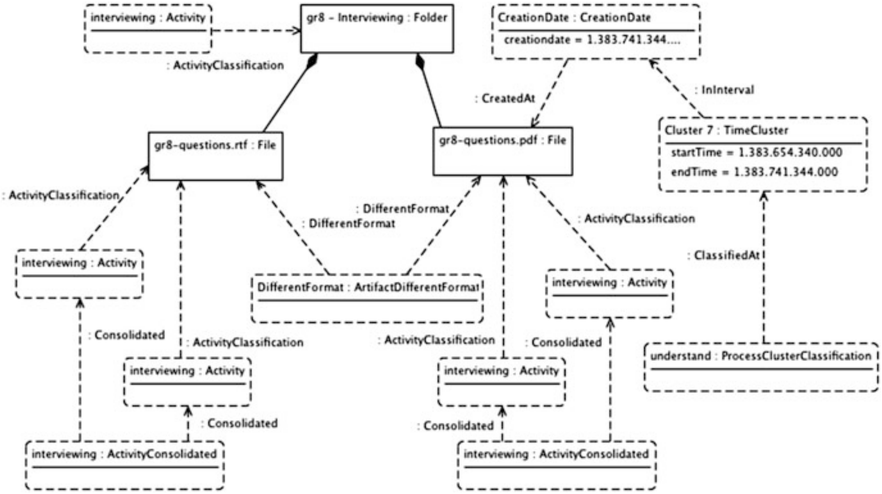


Fig. 17 Example of similar artifacts in different formats

5 Related Work

The area of related work concerning the documentation challenge and our proposed approach is manifold. We consider Design Thinking as a modern form of requirements engineering. While Brown defines desirability, viability and feasibility as the three dimensions of design thinking (Brown 2009), Pohl defines the three dimensions of requirements engineering, namely specification, agreement and representation (Pohl 1994). While the outcome of Design Thinking projects have to be desired by end-users—viable to sell and feasible to build—requirement engineers have to come up with a common agreement, complete system specification and formal representation of requirements. Traceability helps to understand and manage these requirements as well as to demonstrate whether or not they are fulfilled (Gotel and Morris 2009). Requirements are the result of content transformations between representations (Gotel and Morris 2009), e.g. use-cases derived from interview transcriptions, which are often not bidirectional what leads to an information loss. Gotel et. al (Gotel and Finkelstein 1997) define traceability as the "*ability to describe and follow the life of requirements in both a forwards and backwards direction*". In (Winkler and von Pilgrim 2010) an overview about traceability in requirements engineering is provided. Especially in design thinking traceability links (traces) of informal nature need to be managed. Related literature classifies traceability approaches into runtime creation of traces, e.g. (Jouault 2005), recovery of traceability links, e.g. (Grechanik et al. 2006), and combined approaches, e.g. (Poshyvanyk et al. 2006). Further, Seibel et al. (2010, 2011) describe an approach to capture the hierarchy and context of traceability links for efficient and scalable traceability maintenance, which ensures that the quality of traceability links does not degrade (Egyed et al. 2009). Especially such maintenance is important to our approach later on in case new artifacts are added, deleted or changed in the active repository.

Several approaches exist to capture and organize Design Thinking artifacts. For example, Tele-Board (Gericke et al. 2011; Gumienny et al. 2012) is a digital whiteboard designed for global team collaboration. ConnectingInfos (Voget 2013) is a software tool, which implements an algorithm to assess the importance of design thinking artifacts based on their incoming and outgoing connections to other artifacts. Moreover, approaches to knowledge capture such as (Klemmer et al. 2001; Ju et al. 2004) exist. In (von Ahn and Dabbish 2004) a verification game is used to label images with descriptive keywords to make these images retrievable. Moreover, Design Thinking artifacts often consists of sketches, which embody knowledge that might be extracted with the help of additional analysis tools such as (Forbus and Usher 2002) by interpreting glyphs.

While (a) knowledge discovery deals with detecting new yet unknown structures in large data sets (Goebel and Gruenwald 1999), (b) information retrieval deals with finding useful information corresponding to a user's query in large data sets (Mitra and Chaudhuri 2000). Our research has to deal with both research areas, since we have to guess patterns and properties within sets of Design Thinking artifacts (a) to translate these patterns and properties into inference rules and search for these patterns and properties later on (b). For example, software design pattern recovery is an information retrieval approach in the area of reverse engineering to infer well-known software design patterns, which are described by the Gang of Four (Gamma et al. 1994). Niere et al. (2001a, b, 2002, 2003a, b) describe an rule system to recover software design patterns as part of software documentation. While their approach is limited to software design pattern recovery, we aim at a *generic* inference system that can be applied to diverse application domains and can be used to emulate already existing inference/recovery approaches, e.g. traceability link recovery and software design pattern recovery.

6 Conclusion and Future Work

In this chapter we described how to capture Design Thinking artifacts and how to use these artifacts and associated meta data to infer additional knowledge, which includes answers to questions of Design Thinkers and engineers. This additional knowledge enables Design Thinkers and engineers to answer more of their questions than with the Design Thinking documentation raw data. We presented an initial Design Thinking inference rule set as proof-of-concept to show that additional knowledge can be extracted from artifacts and additional relationships between these artifacts exist. However, our inference system does not make ongoing Design Thinking documentation during Design Thinking projects obsolete, since our inference system requires at least a minimum quantity of Design Thinking documentation to infer additional knowledge.

Additional investigations are especially required to exploit the knowledge embedded within photographs, since these kind of artifacts is primarily used. We intend to improve our initial Design Thinking inference rule set by taking uncertainty into account, e.g. in terms of belief values. We plan to evaluate the improved Design Thinking inference rule set in an experiment where Design Thinkers rate the correctness of inferred knowledge retrospectively. Therefore, concrete metrics are required which take the dependencies between created annotations into account. Moreover, we plan to incrementally update inferred knowledge, e.g. in case artifacts are added, changed or deleted, since a batch execution of the Design Thinking inference rule set is resource-intensive and can take up to several minutes.

Acknowledgement The authors are grateful for the input of Claudia Nicolai (D-School Potsdam), Axel Menning (D-School Potsdam). The authors are also grateful for the input of Alexander Renneberg (D-LABS GmbH) and our student assistants Josephine Harzmann, Manuel Hegner and Lukas Pirl. We give thanks to our bachelor project teams *"From Creative Ideas to Well-Founded Engineering"* and *"Intelligent Organization of Design Thinking Artifacts"* for their input.

References

Arkley P, Riddle S (2005) Overcoming the traceability benefit problem. In: CORD conference proceedings, pp 385–389, April 2005

Balzert H (1997) Software-Management, Software-Qualitätssicherung, Unternehmensmodellierung. Lehrbuch der Software-Technik, 1st edn. Spektrum Akademischer Verlag GmbH, Heidelberg, November 1997

Barbero M, Bézivin J (2008) Model driven management of complex systems: implementing the macroscope's vision. In: 15th annual IEEE international conference and workshop on the engineering of computer based systems, pp 277–286

Beyhl T, Berg G, Giese H (2012) Tackling the documentation benefit problem in design thinking. In: Confestival 2012—design thinking the future, Potsdam

Beyhl T, Berg G, Giese H (2013a) Connecting designing and engineering activities. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research—building innovation eco-systems. Springer, Heidelberg

Beyhl T, Berg G, Giese H (2013b) Towards documentation support for educational design thinking projects. In international conference on engineering and product design education, pp 408–413, September 2013

Beyhl T, Berg G, Giese H (2013c) Why innovation processes need to support traceability. In: Workshop on traceability in emerging forms of software engineering 2013, San Francisco, May 2013

Brown T (2009) Change by design. HarperCollins Publishers, New York

Egyed A, Grünbacher P, Heindl M, Biffl S (2009) Chapter 14: Value-based requirements traceability: lessons learned. In: Lyytinen K, Loucopoulos P, Mylopoulos J, Robinson W (eds) Design requirements engineering: a ten-year perspective, vol 1. Springer, Germany, pp 240–257

Forbus K, Usher J (2002) Sketching for knowledge capture: a progress report. ACM, New York

Gabrysiak G, Guentert M, Hebig R, Giese H (2012) Teaching requirements engineering with authentic stakeholders: towards a scalable course setting. In proceedings of ICSE 2012 workshop on software engineering education based on Real-World experiences, Zurich, Switzerland, June 2012

Gamma E, Helm R, Johnson R, Vlissides J (1994) Design patterns—elements of reusable object-oriented software. Addison-Wesley, Boston

Gericke L, Gumienny R, Meinel C (2011) Tele-board: follow the traces of your design process history. In: Design thinking research, July 2011, Springer, Berlin, pp 15–29

Goebel M, Gruenwald L (1999) A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explor 1:20–33

Gotel O, Finkelstein A (1997) Extended requirements traceability: results of an industrial case study. In: Proceedings of the third IEEE international symposium on requirements engineering, 1997, pp 169–178

Gotel O, Morris S (2009) More than just "Lost in Translation". IEEE Softw 26(2):7–9

Gotel OCZ, Morris SJ (2011) Out of the labyrinth: leveraging other disciplines for requirements traceability. In: 19th IEEE international requirements engineering conference (RE), pp 121–130

Grechanik M, McKinley K, Perry D (2006) Recovering use-case-diagram-to-source-code traceability links. In: 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on the foundations of software engineering

Gumienny R, Gericke L, Wenzel M, Meinel C (2012) Tele-Board in use: applying a digital whiteboard system in different situations and setups. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research—measuring performance in context. Springer, Heidelberg, pp 109–125

Jouault F (2005) Loosely coupled traceability for ATL. In: Proceedings of European conference on model driven architecture workshop on traceability, pp 29–37

Ju W, Ionescu A, Neeley L, Winograd T (2004) Where the wild things work: capturing shared physical design workspaces. In: ACM conference on computer supported cooperative work workshop on shared environments to support face-to-face collaboration, pp 533–541. ACM

Klemmer SR, Newman MW, Farrell R, Bilezikjian M, Landay JA (2001) The designers' outpost: a tangible interface for collaborative web site. In: UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM request permissions, November 2001

Mitra M, Chaudhuri BB (2000) Information retrieval from documents: a survey. Inf Retr 2(2):141–163

Niere J (2004) Inkrementelle Entwurfsmustererkennung. Ph.D. Thesis, Universität Paderborn

Niere J, Wadsack J, Wendehals L (2001a) Design pattern recovery based on source code analysis with fuzzy logic

Niere J, Wadsack JP, Zündorf A (2001b) Recovering UML Diagrams from Java Code using Patterns. In: 2nd Workshop on Soft Computing Applied to Software Engineering

Niere J, Schafer W, Wadsack J, Wendehals L, Welsh J (2002) Towards pattern-based design recovery. In: Proceedings of the 24rd international conference on software engineering (ICSE), pp 338–348

Niere J, Wadsack J, Wendehals L (2003a) Handling large search space in pattern-based reverse engineering. In: 11th IEEE international workshop on program comprehension, pp 274–279

Niere J, Wendehals L, Zündorf A (2003b) An interactive and scalable approach to design pattern recovery. October 2003

Plattner H, Meinel C, Weinberg U (2009) Design thinking. Innovation lernen—Ideenwelten öffnen. mi-Wirtschaftsbuch, Finanzbuch Verlag GmbH, München

Pohl K (1994) The three dimensions of requirements engineering: a framework and its applications. In: Selected papers from the fifth international conference on advanced information systems engineering, Pergamon Press, Elmsford, NY, pp 243–258

Poshyvanyk D, Gueheneuc Y-G, Marcus A, Antoniol G, Rajlich V (2006) Combining probabilistic ranking and latent semantic indexing for feature identification. In: 14th IEEE international conference on program comprehension, pp 137–148

Seibel A, Neumann S, Giese H (2010) Dynamic hierarchical mega models: comprehensive traceability and its efficient maintenance. Softw Syst Model 9(4):493–528

Seibel A, Hebig R, Neumann S, Giese H (2011) A dedicated language for context composition and execution of true black-box model transformations. In: 4th International conference on software language engineering (SLE 2011), Braga, pp 19–39

Voget L (2013) An assocation-based documentation tool to support design thinking groups in using their acquired knowledge. Ph.D. Thesis, June 2013

von Ahn L, Dabbish L (2004) Labeling images with a computer game. In: CHI '04: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Request Permissions, April 2004

Winkler S, von Pilgrim J (2010) A survey of traceability in requirements engineering and model-driven development. Softw Syst Model 9(4):529–565

How Cost Reduction in Recovery Improves Performance in Program Design Tasks

Bastian Steinert and Robert Hirschfeld

Abstract Changing source code often leads to undesired implications, raising the need for recovery actions. Programmers need to manually keep recovery costs low by working in a structured and disciplined manner and regularly performing practices such as testing and versioning. While additional tool support can alleviate this constant need, the question is whether it affects programming performance? In a controlled lab study, 22 participants improved the design of two different applications. Using a repeated measurement setup, we compared the effect of two sets of tools on programming performance: a traditional setting and a setting with our recovery tool called CoExist. CoExist makes it possible to easily revert to previous development states even, if they are not committed explicitly. It also allows forgoing test runs, while still being able to understand the impact of each change later. The results suggest that additional recovery support such as provided with CoExist positively affects programming performance in explorative programming tasks.

1 Introduction

Changing source code easily leads to the need for recovery actions because the changes reveal implications that are not only unexpected but also undesired. They might suddenly turn out inappropriate, turn out more complex than expected, or they might have introduced an error. Programmers then need to withdraw these recent changes, recover knowledge from previous development versions, or locate and fix the error.

B. Steinert (✉) • R. Hirschfeld
Software Architecture Group, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
e-mail: firstname.lastname@hpi.uni-potsdam.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_13, © Springer International Publishing Switzerland 2015

To keep the costs for potential recovery needs low, programmers have to follow a structured and disciplined approach. This involves the regular use of testing and versioning tools, but also to perform baby steps and to work only on one thing at a time (Beck and Andres 2004; Fowler 1999; Apache Software 2009). Regular testing helps discover errors early and thus reduces fault localization costs, regular commits help to return to a previous state, and working on one thing at a time makes it easier to commit independent increments—to mention just a few examples. By following these recommendations, programmers can avoid the need for expensive recovery work that easily becomes frustrating.

However, while structure and discipline are certainly useful to get work done, it hardly seems sufficient to be forced to rely on them constantly. On the one hand, it is hard to exert the required discipline when being fascinated by an idea and having the desire to explore it. On the other hand, it is easy to forget to perform recommended practices and it requires much effort to avoid forgetting. This not only takes time but also easily disrupts working on the main task.

The need for structure and discipline is also present when programmers decide to first create a prototype on a separate branch, in order to evaluate a particularly risky idea. One reason is that they might want to reuse the source code and avoid the need to re-implement it. Another reason is that when working on a prototype, it is still likely that changes reveal undesired implications independent of the aspects being evaluated. So, the same rules apply: programmers will still need to recover, and they also need to manually keep recovery cost low, for example, by testing regularly, making meaningful commits, and making only small changes, one at a time.

Additional tool support can help to avoid the constant need for structure and discipline by keeping recovery costs low automatically. We previously presented an IDE extension called CoExist (Steinert et al. 2012), which is implemented in Squeak/Smalltalk. Figure 1 illustrates main concepts of the user interface. CoExist continuously versions the program under development, runs tests in the background, and provides immediate access to intermediate development states. It allows programmers to easily recover from undesired situations, also when they forgot to make the appropriate commit or have failed to run the right set of tests regularly. These features enable programmers to ignore recommended practices. They can try out an idea when it comes to mind, make changes as they think of them, and explore the implications, without having to worry about tedious recovery scenarios and how to prevent them.

We hypothesized that such additional recovery support has an effect on programming performance, in particular on tasks that involve a high degree of uncertainty. We speculate that this is the case for two reasons: (1) making changes directly as one thinks of them supports mental process and is thus more efficient; (2) The constant need for structure and discipline is tiring and contradicts the need for creative thinking.

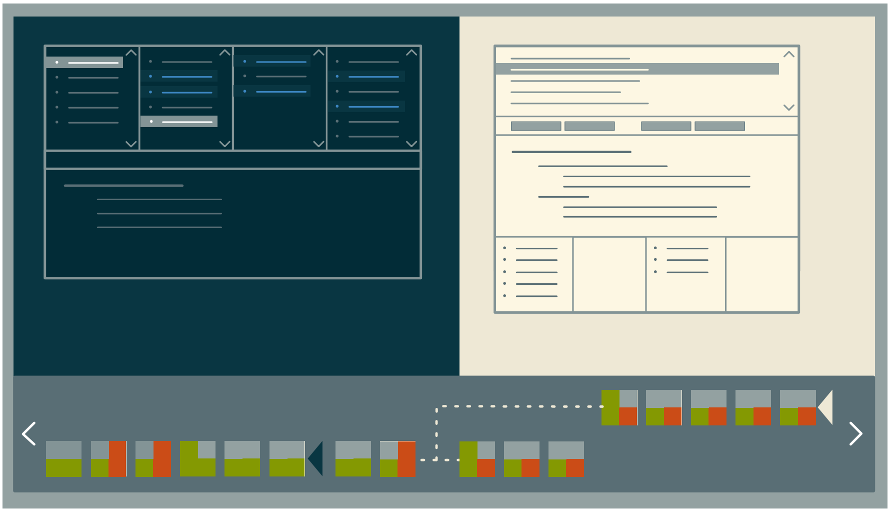


Fig. 1 The CoExist IDE extension featuring continuous versioning, running tests and recording test results in the background, side by side exploring and editing multiple versions

1.1 Why Thinking Is Supported by Doing

Programmers should be encouraged to make changes as they think of them, because it will facilitate inference, understanding, and problem solving, as suggested by research findings in design and cognition (Suwa et al. 1998; Suwa and Tversky 2002).

It avoids mental overload and keeps working memory free. Making the changes instead of conducting what-if reasoning "frees working memory to perform mental calculations on the elements rather than both keeping elements in mind and operating on them" (Suwa and Tversky 2002). Freeing working memory is required because the number of chunks of new information that a human being can keep in mind and process is limited (three to four chunks). Given too many chunks at once, a human being experiences cognitive overload, which impedes learning and problem solving (Bilda and Gero 2007; Farrington 2011).

Making the changes allows for re-interpretation and unexpected discovery. Even if they turn out inappropriate, the changes can trigger new associations. Previously abstract concepts and thoughts will be associated with specific source code elements. When programmers revisit these specific elements, they can see them as something else. They associate abstract concepts with these elements that are different than the original ones. Making the changes brings to mind information from long-term memory that might otherwise not be retrieved (Suwa and Tversky 2002; Kirsh 2010). The particular arrangement can also lead to the discovery of unexpected relations and features (Kirsh 2010; Schon and Wiggins 1992).

1.2 Why the Need for Structure and Discipline Is Tiring and in Contradiction with the Need for Creativity

Psychology distinguishes two modes of thinking: fast thinking and slow thinking, often labeled as System 1 and System 2 (Kahneman 2011). While creativity along with intuition is attributed to System 1 (fast thinking mode), while the analytic approach along with suspicion is attributed to System 2 (slow thinking mode). This implies that creative thinking and analytical reasoning don't go well together. Working on a creative programming task is impeded by the need to reflect about current and planned changes and the need to structure the work ahead. If programmers constantly need to be analytical and careful, it will be difficult for them to be creative at the same time.

Furthermore, the need for a structured and disciplined approach to programming requires self-control, which is a form of exhaustive mental work, as the following quotes from [(Kahneman 2011), chapter "Developing Novel Methods to Assess Long-Term Sustainability of Creative Capacity Building and Applied Creativity"] should illustrate:

- ". . . controlling thoughts and behaviors is one of the tasks of System 2."
- "Too much concern about how well one is doing in a task sometimes disrupts performance by loading short-term memory with pointless anxious thoughts. . . . self-control requires attention and effort."
- "an effort of will or self-control is tiring; if you have to force yourself to do something, you are less willing or less able to exert self-control when the next challenge comes around."

These findings give reason to believe that additional recovery support such as CoExist is preferable over a manual method-based approach. We hypothesize that CoExist improves the performance of programmers in explorative tasks. In the remainder of this article, after first describing CoExist, we report on an experiment conducted to empirically examine our hypothesis.

2 Background: The Coexist IDE Extensions

The basis of CoExist takes care of preserving potentially valuable information. It continuously performs commits in the background. Every change to the code base leads to a new version one can go back to. It thus gives users an impression of development versions to *co-exist*. To make the user aware of this background versioning and to allow for selecting previous versions, we have added a version bar (timeline) to the user interface of the programming environment (Fig. 1).

By continuously preserving intermediate development states, CoExist enables programmers to go back to a previous development state and to start over as shown in Fig. 2. Starting over from a former development state will implicitly create a new

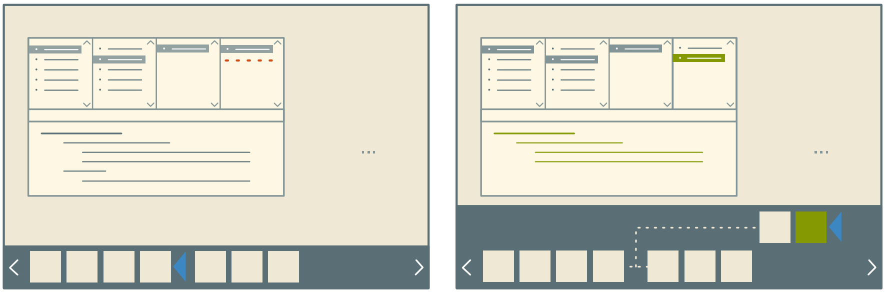
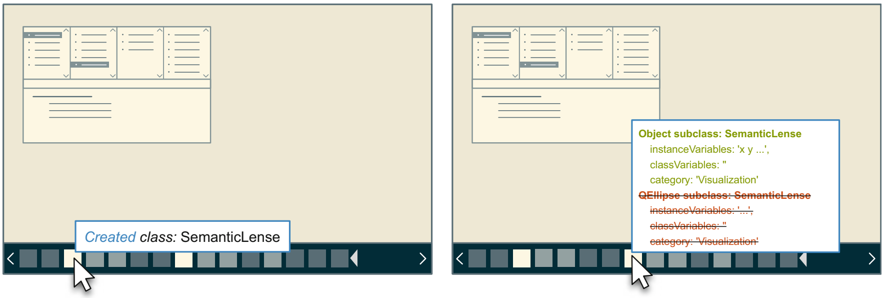


Fig. 2 The (*blue*) *triangle marks* the current position in the history—the version that is currently active. When a programmer goes back to a previous version (*left*), and then continues working, the new changes will appear on a new branch that is implicitly created (*right*)



Created class: SemanticLense

Object subclass: SemanticLense
instanceVariables: 'x y ...',
classVariables: ''
category: 'Visualization'
~~**QEllipse subclass: SemanticLense**~~
~~instanceVariables: '...',~~
~~classVariables: ''~~
~~category: 'Visualization'~~

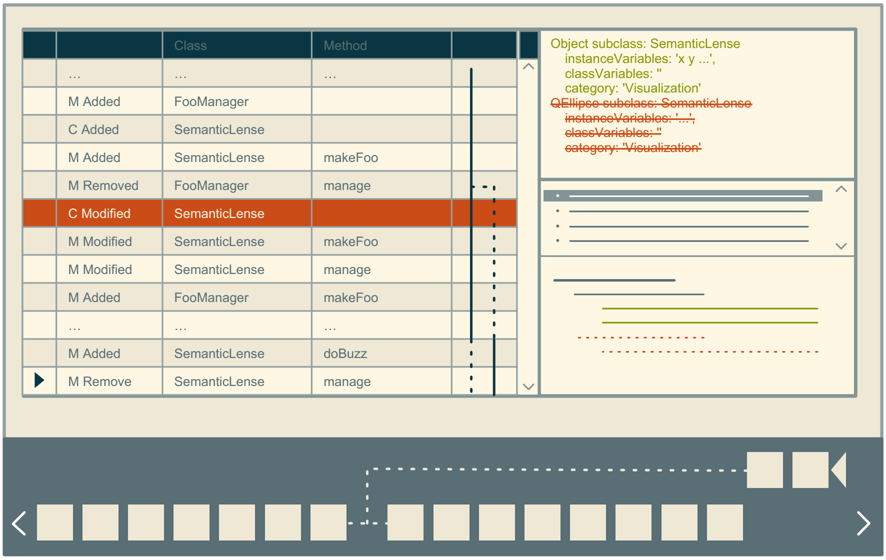
Fig. 3 Hovering shows which source code element has been changed (*left*). In addition, holding shift shows the total difference to the previous version (*right*)

branch of versions. This preserves the changes that are withdrawn, as they might be of use later on.

CoExist provides two mechanisms to support programmers in identifying a previous version of interest. First, it provides the version bar, which will highlight version items that match the currently selected source code element. Hovering the items will display additional information, such as the kind of modification, the affected elements, or the actual change performed (Fig. 3).

Second, programmers can use the version browser to explore information of multiple versions at a glance. The version browser displays basic version information in a table view (Fig. 4), which allows to scan the history fast.

CoExist is meant to close the gap between the undo/redo feature and Version Control Systems such as Git. It is not intended to replace either of them. Furthermore, we acknowledge that conscious and named commits can be useful, but we omitted the possibility of naming or flagging intermediate versions to avoid inducing users to think about it. We also want to explore how far one can go with our approach.



		Class	Method	
	...	...	...	
	M Added	FooManager		
	C Added	SemanticLense		
	M Added	SemanticLense	makeFoo	
	M Removed	FooManager	manage	
	C Modified	SemanticLense		
	M Modified	SemanticLense	makeFoo	
	M Modified	SemanticLense	manage	
	M Added	FooManager	makeFoo	
	...	...	...	
	M Added	SemanticLense	doBuzz	
▶	M Remove	SemanticLense	manage	

Object subclass: SemanticLense
instanceVariables: 'x y ...',
classVariables: ''
category: 'Visualization'
~~QEllipse subclass: SemanticLense~~
~~instanceVariables: '...',~~
~~classVariables: ''~~
~~category: 'Visualization'~~

Fig. 4 The version browser provides a tabular view on change history. Selecting a *row* shows corresponding differences in the panes on the *right*

CoExist also allows continuously running analysis programs for newly created versions. As a default, it runs test cases to automatically assess the quality of the change made. The test result for a version is presented in the corresponding item of the version bar (left of Fig. 5). This makes the effect of each change regarding test quality visible. The user can also run other analyses such as performance measurements. CoExist provides full access to version objects and offers a programming interface to run code in the context of a particular version. So, whenever programmers become interested in the impact of their changes, they can easily analyze it in various respects. This allows programmers to ignore these aspects of programming at other times.

Users of CoExist can explore the source code of a previous version and compare it to the current one. They can open a previous version in a separate working environment as shown on the right in Fig. 5, which is useful, when, for example, the programmer suddenly become curious about how certain parts of the source code looked previously or how certain effects were achieved. It is also possible to run and debug programs in the additional working environment. In doing so, CoExist is capable of efficiently recovering knowledge from previous versions, which avoids the need for a precise understanding of every detail before making any changes.

With CoExist, programmers can change source code without worrying about the possibility of making an error. This is because they can rely on tools that will help with whatever their explorations turn up. They no longer have to follow certain best practices in order to keep recovery costs low.

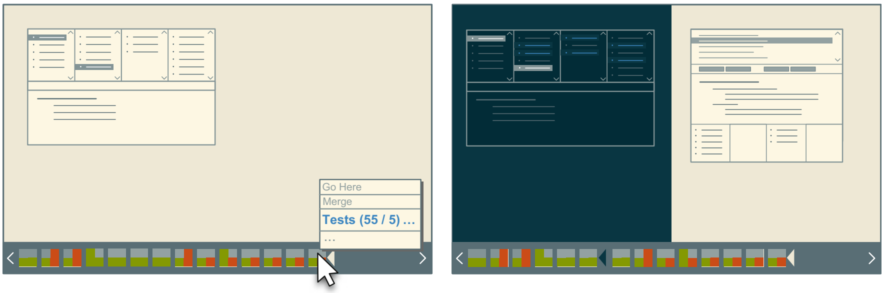


Fig. 5 The items in the version bar are now a visualization of the results of the tests that have been run in the background (*left*). A second inner environment allows the user to explore a previous version next to the current one (*right*)

3 Method

3.1 *Study Design*

Figure 6 illustrates the experimental setup. Participants have been assigned to either of two groups, the control group or the experimental group. Members of the control group used the regular development tools for both tasks. Members of the experimental group used the regular tools only for task 1, and could additionally rely on CoExist for task 2.

We kept participants unaware of what condition they had been assigned to. However, on day 2, participants in the experimental group could guess that they were receiving special treatment because they were introduced to a new tool and could make use of it. At the same time, participants in the control group were unaware about the experimental treatment. They did not know that the participants in the experimental group were provided with CoExist.

The setup resulted in two scores for every participant, which allowed testing for statistically significant differences between task 1 and task 2 as well as between the control and the experimental group. It also allows to test for an interaction effect of the two factors, which is the indicator of whether CoExist affects programming performance.

3.2 *Materials and Task*

On both days the task has been to improve the source code of relatively small computer games. More specifically, participants were requested to study the source code, to detect design flaws in general and issues of unnecessary complexity in particular, and to improve the source code as much as possible in the given time

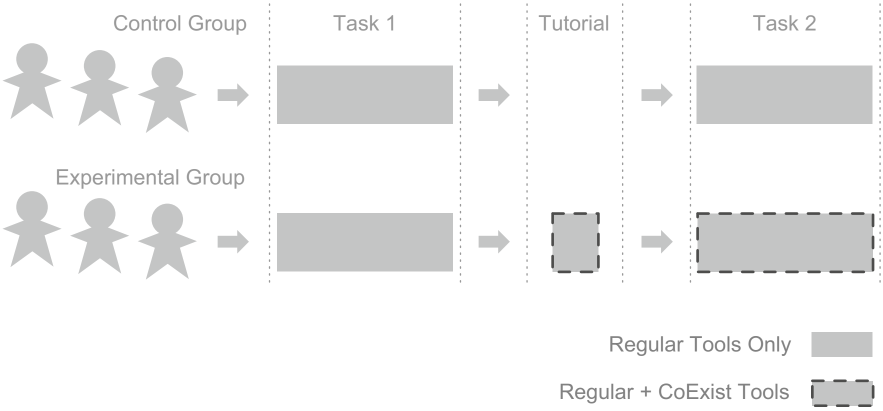


Fig. 6 Our experiment setup to compare performance in program design activities

frame of 2 h. The games needed to function properly at the end of the task. To help participants better understand the task, we provided descriptions of possible improvements such as the following:

- Extract methods to shorten and simplify overly long and complicated methods, and to ensure statements have a similar level of abstraction
- Replace conditional branching by polymorphism
- Detect and remove unnecessary conditions or parameters

Participants should imagine that they co-authored the code and now have time to improve it in order to make future development tasks easier. Also, participants were asked to describe their improvements and to help the imaginary team members better understand them. (Most participants followed this instruction by regularly writing commit messages).

On day 1, participants worked on a game called *LaserGame*, and on day 2 they worked on a gamed called *MarbleMania*. Screenshots of both games are shown in Fig. 7. For the LaserGame (on the left), the user has to place mirrors in the field so that the laser is redirected properly to destroy the wall that blocks the way to the gate to the next level. For MarbleMania (on the right), the user has to switch neighboring marbles to create one or more sequences of at least three equally colored marbles, which will then be destroyed, and gravity will slide down marbles from above.

Both games were developed by students in one of our undergraduate courses. The two selected games function properly and provide a simple but nevertheless fun game play. Accordingly, only a little time is required to get familiar with the functionality. Furthermore, for each of the two games, there is significant room for improvement concerning the source code (because they were created by young undergrads who were about to learn what elegant source code is). Furthermore, both games come with a set of tests cases, which also have been developed by the respective students. However, while the offered test cases are useful, they were not sufficient. Manual testing of the games was necessary.

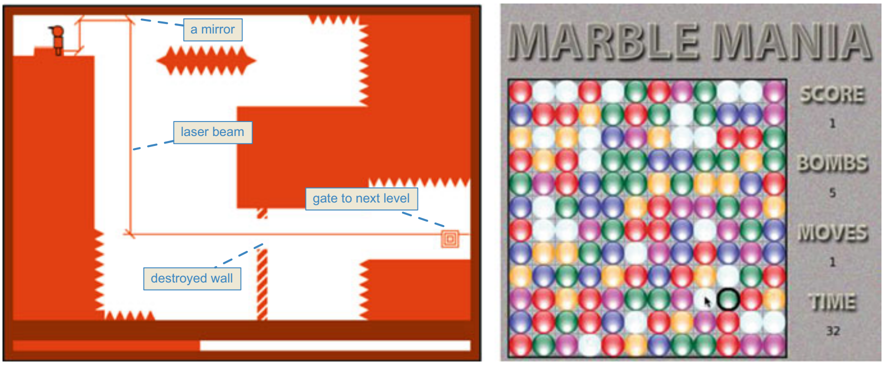


Fig. 7 Screenshots of the games whose source code was improved in the experiment: LaserGame (*left*) and MarbleMania (*right*)

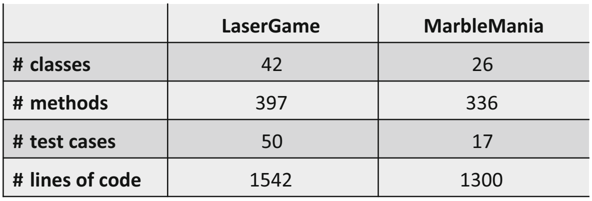
While the numbers shown in Fig. 8 indicate that both games are of similar size, the code base of the LaserGame is easier to understand. The authors of MarbleMania placed a great deal of emphasis on the observer pattern and built in many indirections, which impedes understanding the control flow.

3.3 Participants

We recruited 24 participants, mainly through email lists of previous lectures and projects. Of the 24 participants, 3 were bachelor students who had completed their fourth semester, 6 were bachelor students who had completed their sixth semester (nearly graduated), 13 were master student who had at least completed their eighth semester, and 2 were Ph.D. students. The average age was 23 with a standard deviation of 2. For approximately 5 h of work, each participant received a voucher worth 60 euros for books on programming-related topics. Of the 24 participants, the results of 2 were dropped which is discussed in the results Sect 4.

Prospective participants needed to have experience in using Squeak/Smalltalk and must had passed their fourth semester. By this time students will have typically attended two of our lectures, in which they use Squeak/Smalltalk for project work. Also, these two lectures cover software design and software engineering topics. Thus we could ensure that all participants had theoretical and practical lessons in topics such as code smells, idioms, design patterns, refactoring, and other related topics.

We have balanced the amount of previously gained experience with Squeak/Smalltalk among both conditions (stratified random sampling). Most participants have used Squeak/Smalltalk only during the project work in our lectures. But six participants also have been using Squeak/Smalltalk in spare time projects and/or in



	LaserGame	MarbleMania
# classes	42	26
# methods	397	336
# test cases	50	17
# lines of code	1542	1300

Fig. 8 Size indicators for the games used in the study

their student jobs, so that we could assume these participants had noticeably more experience and were more fluent in using the tools.

3.4 Procedure

We always spread the experiment steps over 2 days, so that participants worked on both tasks on two different but subsequent days. On both days, the procedure comprised two major steps: an introduction to the game and a 2-h time period for improving the respective codebase. On day two participants of the experimental group received an additional introduction to the CoExist tools before working on the actual tasks, during which they could rely on CoExist as an additional recovery support.

Both tasks were always scheduled for the same time of the day in order to assure similar working conditions (hours past after waking up, hours already spent for work or studies, . . .). Typically, we scheduled the task assignments after lunch so that for day 2, there was time left to run the CoExist tutorial session upfront before lunchtime. (We had to make an exception for three participants, who only had time during the morning or evening hours. As we could not arrange a similar schedule for these participants concerning the CoExist tutorial followed by a large break, these three participants were automatically assigned to the control group).

Figure 9 illustrates all steps of the experiment. On day 1, participants received a brief recap of IDE shortcuts, which were also written on the whiteboard in the room. The step of *Introduction to < a game >* started with a short explanation of the game play, followed by some time to actually play the game, to understand details, and to get comfortable with it.

3.5 Dependent Measure

To compare the performance of the individual programming sessions, we have operationalized the notion that a programmer can achieve more or less improvements in the given timeframe. We determined performance by *identifying independent increments* among the overall set of made changes, and *quantifying the effort for these increments* by defining sequences of IDE interactions required to

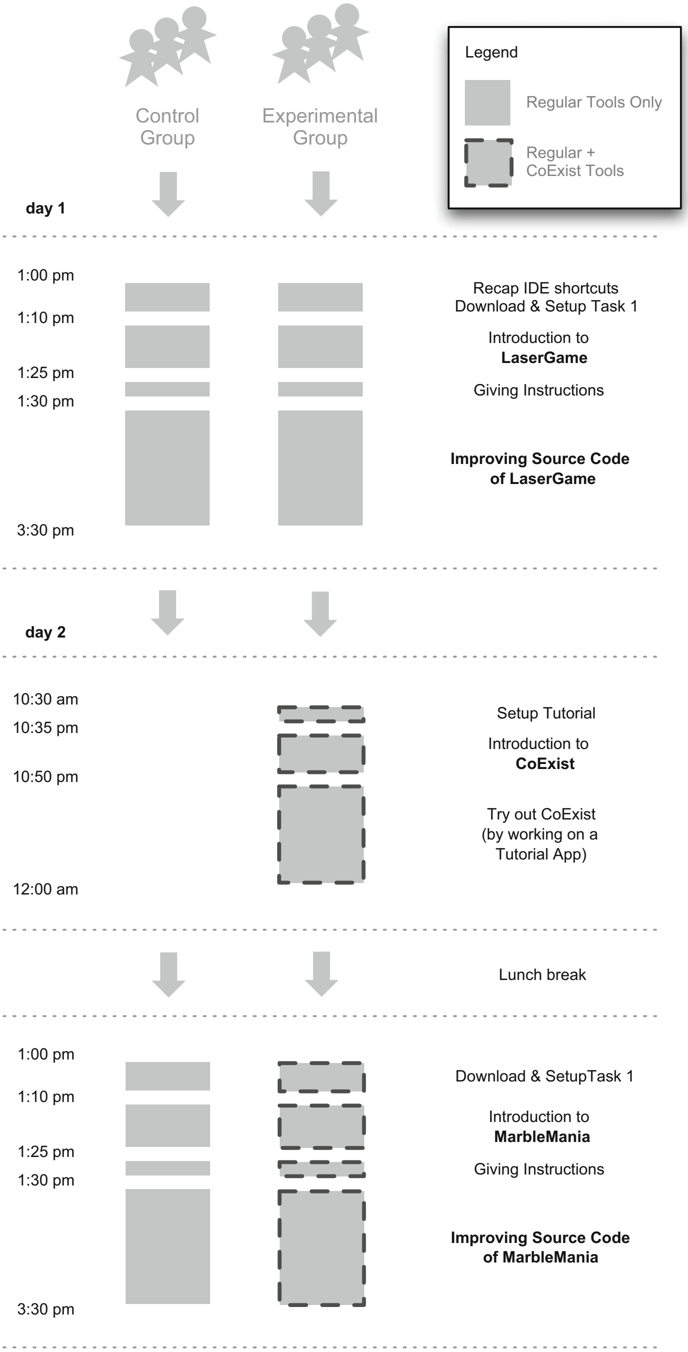


Fig. 9 The experimental procedure for both the control and the experimental group

reproduce them. This gives a measure of how much actual work was done within the 2 h, excluding time that has been spent on activities such as staring into the air or browsing the code base.

3.5.1 Identifying Independent Increments

An independent increment is a set of interconnected changes to the code base that represents a meaningful, coherent improvement such as an *ExtractMethod* refactoring, which is comprised of the changes: (a) adding a new method and (b) replacing statements with a call to the newly created method. Another example for an independent increment is the replacement of code that caches state in an instance variable with code that re-computes the result on every request, or vice versa. Other generic improvements are for example:

- Renaming of an instance variable
- Replace a parameter with a method
- Make use of cascades
- Inline temporary expression
- Replace magic string/number with method

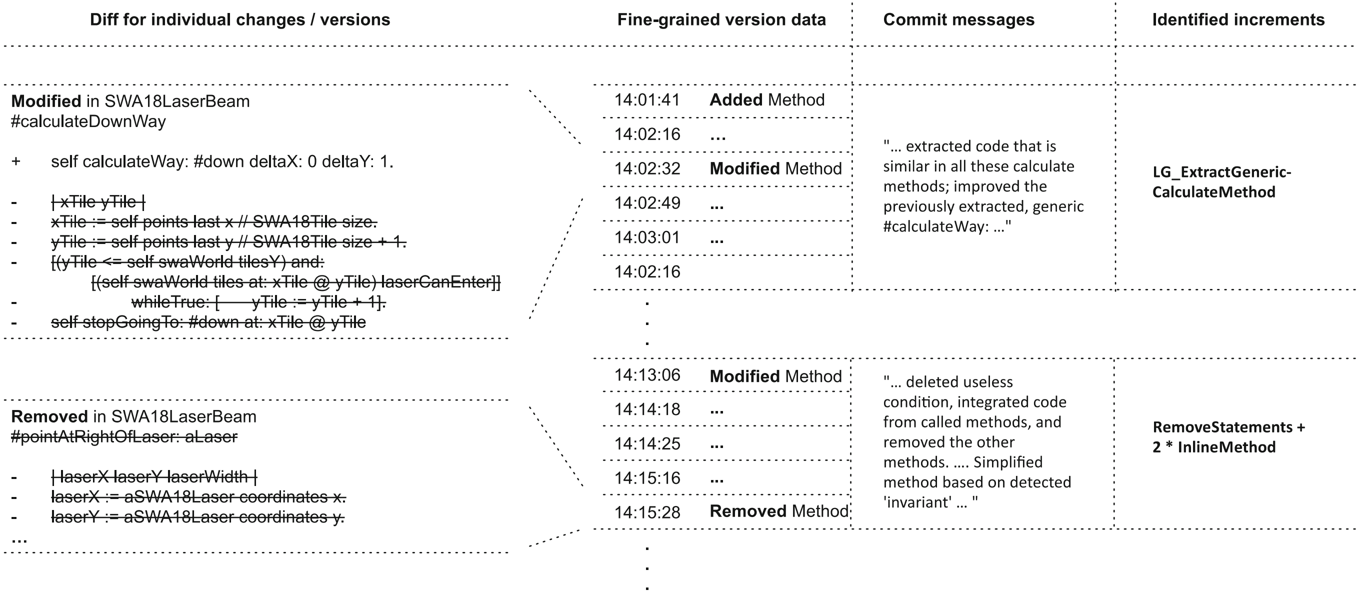
Besides such generic and well-document improvements, an increment can also be specific to a certain application. The following examples are game specific improvements that were identified for the MarbleMania game:

- Replace dictionary that holds information about exchange marbles with instance variables
- Replace "is nil" checks in the Destroyer with null objects (the Destroyer class has the responsibility to "destroy" marbles when, after an exchange, a sequence of three or more marble exists)
- Remove button clicked event handling indirections

For each participant and task, we recorded a fine-grained change history using CoExist's continuous versioning feature. However, the CoExist tools were neither visible nor accessible to the users, except for the experimental group on day 2. We then analyzed these recorded change histories manually to identify the list of independent increments. For each programming session (per programmer and task), the analysis consisted of two steps to gain a corresponding spreadsheet as illustrated in Fig. 10.

First, we extracted the timestamps of all versions and listed them in a column of a spreadsheet. We then grouped these timestamps according to the commits that subjects made during the task, and put the corresponding commit messages in a second column (illustrated in Fig. 10). The commit messages provide context that helps getting an initial understanding of the changes' intent.

Second, we hovered over all version items step by step (compare with Fig. 3) to refine our understanding of the made changes, and put names for identified increments in a third column. Such a coded increment can involve only one actual



Diff for individual changes / versions

Modified in SWA18LaserBeam
#calculateDownWay

+ self calculateWay: #down deltaX: 0 deltaY: 1.

- ~~| xTile yTile |~~
- ~~xTile := self points last x // SWA18Tile size.~~
- ~~yTile := self points last y // SWA18Tile size + 1.~~
- ~~[(yTile <= self swaWorld tilesY) and:~~
~~[(self swaWorld tiles at: xTile @ yTile) laserCanEnter]]~~
- ~~whileTrue: [yTile := yTile + 1].~~
- ~~self stopGoingTo: #down at: xTile @ yTile~~

Removed in SWA18LaserBeam
~~#pointAtRightOfLaser: aLaser~~

- ~~| laserX laserY laserWidth |~~
- ~~laserX := aSWA18Laser coordinates x.~~
- ~~laserY := aSWA18Laser coordinates y.~~
...

Fine-grained version data		Commit messages	Identified increments
14:01:41	**Added** Method	"... extracted code that is similar in all these calculate methods; improved the previously extracted, generic #calculateWay: ..."	**LG_ExtractGeneric-CalculateMethod**
14:02:16	...		
14:02:32	**Modified** Method		
14:02:49	...		
14:03:01	...		
14:02:16			
. . .			
14:13:06	**Modified** Method	"... deleted useless condition, integrated code from called methods, and removed the other methods. .... Simplified method based on detected 'invariant' ... "	**RemoveStatements + 2 * InlineMethod**
14:14:18	...		
14:14:25	...		
14:15:16	...		
14:15:28	**Removed** Method		
. . .			

Fig. 10 Excerpt of a spreadsheet with coded version data

change or consist of many. Sometimes, all the changes made for one commit contribute to one coded improvement. Note that we only coded increments for changes that last until the end of the session. This excludes change sets that were withdrawn later, for example.

3.5.2 Quantifying the Effort for Identified Improvements

We measure the effort for every increment by determining the list of IDE interactions that are required to (re-) produce it. Such interactions are, for example: navigating to a method, selecting code and copying it to clipboard, selecting code and replacing it with the content from the clipboard, inserting symbols. Figure 11 shows two lists of IDE interactions, written down in an executable form (regular Smalltalk code). Executing a script computes a number that represents the effort required to reproduce the described increment.

We determined these scripts by re-implementing every identified increment based on a fresh clean code base, which participants started with. Re-implementing the increments ensured that we had gotten a correct understanding. We always used the direct path to achieve an increment, which might be different than the path made by participants. Thus, we only measured the essential effort and excluded any detours that participants might have made until they eventually knew what they wanted.

For generic increments such as ExtractMethod or InlineMethod, the required effort can vary: extracting a method with five parameters requires more symbols to be inserted than an extracting a method without any parameters. We accounted for such differences by listing the interactions required for an average case. However, for extreme variations (easy or hard), we used special codes such as ExtractMethodForMagicNumber.

The messages used in these scripts call utility methods that are typically composed of more fine-grained interactions. At the end, all descriptions rely on four elementary interactions, which are: `#positionMouse`, `#pressKey`, `#brieflyCheckCode`, and `#insertSymbols: aNumber`. The methods for these elementary interactions increment a counter variable when they are executed. While the former three increment the counter by one, the latter increments the counter by three for every symbol inserted. So we assume that writing a symbol of an average length is three times the effort of pressing a single key. (While this ratio seemed particularly meaningful to us, we also computed the final numbers with a ratio of two and four. The alternative outcomes, however, show a similar result. In particular, a statistical analysis using ANOVA also reveals a significant interaction effect.)

```
CvEval >> #renameClass

  self
    navigateTo: #class;
    requestRefactoringDialog;
    insertSymbols: 1;
    checkSuggestionsAndAccept


CvEval >> #lgReplaceCollectionWithMatrix

  self
    navigateTo: #formWidth... in: #Grid;
    selectAndInsert: 5;
    navigateTo: #at: in: #Grid;
    selectAndInsert: 1;
    navigateTo: #at:put in: #Grid;
    selectAndInsert: 1
```

Fig. 11 The first example represents the list of interactions required for the generic RenameClass refactoring, while the second represents an increment that is specific to the LaserGame

4 Results and Discussion

Figure 12 shows the result scores for each participant and task, the accumulated points for the identified increment. Note that while we recruited 24 participants, we only present and further analyze the scores of 22 participants. One of the two participants had to be dropped because after the session we surprisingly found out that he had already been familiar with the MarbleMania game. He had used the source code of this game for his own research. The other result was dropped because the participant delivered a version for task 2 that did not function properly. Further analysis revealed that this problem could not be easily fixed and that the code already stopped working with a change made after half an hour of work. So we decided to drop this data set.

A 2 × 2 mixed factorial ANOVA was conducted with the task (LaserGame, MarbleMania) as within-groups variable and recovery support (with and without CoExist as additional support) as between-groups variable. Both the Shaphiro-Wilk normality test and Levene's test for homogeneity of variance were not significant ($p > 0.05$), complying with the assumption of the ANOVA test.

The bar plot in Fig. 13 illustrate that the control group scored on average less for the MarbleMania task than for the LaserGame task, while there is a slight increase in the performance of the experimental group. This indicates that improving MarbleMania was the more difficult task, and that the provision of CoExist helped to compensate for the additional difficulty.

Statistical significance tests were conducted from the perspective of null hypothesis significant testing with alpha = 0.05, and effect sizes were estimated using partial eta-squared, $\eta_p{}^2$. The results show a significant interaction effect between the

Fig. 12 Final scores for participants per task

	Task 1 / LaserGame	Task 2 / MarbleMania
Control Group	795	306
	183	62
	783	513
	1031	585
	90	0
	323	460
	1019	278
	394	519
	890	408
	784	480
	611	470
Experimental Group	533	499
	217	479
	1286	1080
	75	420
	726	109
	548	374
	460	338
	195	217
	353	493
	651	1115
	320	771

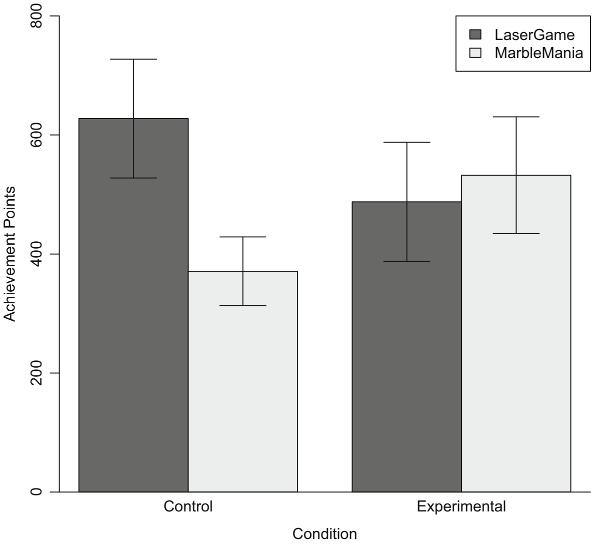


Fig. 13 A bar plot of the study results. Error bars represent the standard error of the mean

effects of task and recovery support on the amount of achievement, $F(1, 20) = 5.49$, $p = 0.03$, ${\eta_p}^2 = 0.22$.

Simple main effects analysis revealed that participants in the control condition (with traditional tool support for both tasks) achieved significantly more for the LaserGame task than for the MarbleMania task, $F(1, 10) = 9.81$, $p = 0.01$, ${\eta_p}^2 = 0.5$, but there were no significant differences for participants in the treatment condition (with CoExist tools), $F(1, 10) = 0.2$, $p = 0.66$, ${\eta_p}^2 = 0.02$.

We performed correlation analyses to illuminate whether the amount of programming experience has an influence on the observed effects. However, there was no correlation between achievements and years of professional education & experience (starting with college education), Pearson's $r(20) = 0.1$, $p = 0.66$. Furthermore, there was no correlation between gains in achievements (difference between points for MarbleMania and points for LaserGame) and years of professional education & experience, Pearson's $r(20) = 0.05$, $p = 0.83$.

The results suggest that the provision of additional recovery support such as CoExist has a positive effect on programming performance in explorative tasks.

5 Limitations

5.1 Order Effects/Counterbalancing

A possible objection to our study design is the lack of counterbalancing the treatment order, as there might be fatigue or learning effects. However, we think that there are complex dependencies between the order of the treatment and the dependent variable. If some participants had received the introduction and the tutorial to CoExist for task 1, which necessarily includes a description of its potential benefits, this would have likely changed how they approach the second task. In particular, they would have been more risk-taking than usual when not having such additional recovery support. So in order to reduce effects of fatigue, we split the study over 2 days. Also, the two tasks were significantly different, rendering each of them interesting and challenging in its own way.

5.2 Construct Validity

Care must be taken not to generalize from our treatment and measure. While we were motivated in this work by discussing recovery support in general, we compared only two levels in our study. Because of this, our results provide only little support that more recovery support is generally better with respect to all these other levels. Additional studies are required to better examine and support the general construct.

Also, the control and experimental group did not only differ in the fact, that one group could rely on CoExist in addition to standard tools for task 2. The members of the experimental group also ran through a tutorial that explains and motivates the CoExist tools. The tutorial or the fact of using a new tool might have contributed to the observed effect.

In addition, there are various social threats to construct validity such as hypothesis guessing or evaluation apprehension that need to be taken into account (Shadish et al. 2002).

5.3 Reliability

We acknowledge the need for further reliability analyses on our measure. Additional studies are required to validate that our construct (the amount of required interactions to reproduce the achieved independent increments) is actually a measure for the amount of work that got done.

We also acknowledge the need for replicating both the coding of change histories, which is the identification of the independent increments, and determining the IDE interactions required for reproduction. Both steps were conducted by only one person, the first author of this article. As the analysis required approximately 2–3 full working weeks, we did not succeed in convincing another researcher to repeat the analysis.

5.4 Internal Validity

While we can observe a correlation between the treatment and the outcome, there might be factors other than the treatment causing or contributing to this effect. As we used a repeated measurement setup, we ruled out single group threats, but need to consider multiple group threats and social threats.

To the best of our knowledge, participants of the control and experimental group are comparable in so far as they experienced the time between both tasks similarly (selection-history threat), that they matured similarly (selection-maturation threat), and learned similarly from Task 1 (selection-testing threat).

However, there is a selection-mortality threat to the validity of our study, because we needed to drop the results of two participants who were both in the control group. But, on the other hand, we had no need to drop any results from the experimental group.

We also need to consider the selection-regression threat, because the average score of both groups is different. So it might be that one of the two groups scored particularly low or high, so that they can only get better or worse respectively. However, the lines in the interaction plot cross. This is an indicator that, besides other possible factors, the treatment is responsible for the observed differences in task 2. The results of the experimental group got better on average, while the results of the control group got worse on average. So, even if one group had a particularly high performance on task 1, the observed differences can hardly just be an artifact of selection-regression.

We dealt with social threats to internal validity, such as compensatory rivalry or resentful demoralization, by blinding participants to the treatment as much as possible.

5.5 External Validity

As we only recruited students for the study, the results are not necessarily representative for the entire population of programmers. However, we conducted correlation analyses to better understand the effect of experience on task performance and gained differences between tasks. The results show that there is no such correlation in the data of our study.

Our study was artificial in the sense that programmers may rarely spend 2 h on improving source code only. It might be more typical that refactoring activities go hand in hand with other coding activities such as implementing new features or fixing bugs.

Furthermore, one might argue that refactoring a previously unknown codebase is also quite untypical. It might be more typical that programmers know a code base and also know their problems that need to get fixed. However, our study design focuses on objectively measuring and comparing programmers' performance.

6 Related Work

We previously presented CoExist and introduced the notion of preserving immediate access to intermediate development states (Steinert et al. 2012). Informal user studies indicated that programmers can identify a previous version of interest within a few seconds and that they appreciate the tools. Continuous versioning, as the basis of CoExist, closes a gap between the undo/redo feature of editors, which works on a more fine-grained level and handles files independently, and Version Control Systems such as Git, which require manual and explicit control. CoExist further builds on early work such as Orwell (Thomas and Johnson 1988) and more recent work such as Delta Debugging (Zeller 1999), Continuous Testing (Saff and Ernst 2003), Changeboxes (Denker et al. 2007), SpyWare (Robbes and Lanza 2007), Replay (Hattori et al. 2011), and Juxtapose (Hartmann et al. 2008).

Continuous testing has been evaluated in a controlled experiment on student developers, showing that this approach helped participants to complete the assignment correctly (Saff and Ernst 2004). CoExist improves on this approach by recording the test results and linking them to the corresponding changes, which allows for analyzing test results only when it is convenient. An empirical evaluation of Replay shows that a fine-grained version history and the possibility to replay changes reduce the time required to complete software evolution analysis tasks (Hattori et al. 2011). In particular, the possibility to replay changes can be considered a meaningful complement to our approach.

Delta Debugging automates the process of testing and refining hypotheses about why a program fails by re-running an automated test and thereby narrowing down the delta that makes the test fail or pass. The dimension on which to narrow down the delta can be the input set provide to the program (Zeller 2002), but also a set of changes between two versions (Zeller 1999). CoExist supports the Delta Debugging approach along the change history well, because it preserves intermediate development states and provides an API to run code on these versions. Also, CoExist records tests results along the history.

Further discussions of related work concerning the technical concepts can be found in our original presentation of the CoExist approach (Steinert et al. 2012).

7 Summary

We have presented an empirical evaluation of the benefits of CoExist over a traditional tool setting on programming performance in explorative tasks. CoExist represents additional recovery support that avoids the need for manually keeping recovery costs low. CoExist continuously versions the source code under development and provides immediate access to intermediate development states and information thereof.

Twenty-two participants ran through a lab study. Using a repeated measurement study, they were requested to improve the design of two games on two consecutive days. The experimental group could additionally rely on CoExist for the second task. Fine-grained change histories were recorded in the background, accumulating approximately 88 h of recorded programming activities. We analyzed the change histories to identify independent increments and determined the required effort for reproducing them. This leads to scores that represent the amount of work achieved within the given time frame. Running an ANOVA test shows a significant interaction effect, $F(1, 20) = 5.49$, $p = 0.03$, ${\eta_p}^2 = 0.22$, which suggests that additional recovery support such as provided with CoExist positively affects the programming performance in explorative tasks.

References

Apache Software Foundation (2009) Subversion best practices. Available http://svn.apache.org/repos/asf/subversion/trunk/doc/user/svn-best-practices.html

Beck K, Andres C (2004) Extreme programming explained: embrace change. Addison-Wesley Longman, Amsterdam

Bilda Z, Gero JS (2007) The impact of working memory limitations on the design process during conceptualization. Des Stud 28(4):343–367

Denker M, Gîrba T, Lienhard A, Nierstrasz O, Renggli L, Zumkehr P (2007) Encapsulating and exploiting change with changeboxes. In: Proceedings of the 2007 international conference on dynamic languages: in conjunction with the 15th international Smalltalk Joint conference 2007, ACM, pp 25–49

Farrington J (2011) Seven plus or minus two. Perform Improv Q 23(4):113–116

Fowler M (1999) Refactoring: improving the design of existing code. Addison-Wesley Professional, Boston, MA

Hartmann B, Yu L, Allison A, Yang Y, Klemmer SR (2008) Design as exploration: creating interface alternatives through parallel authoring and runtime tuning. In: Proceedings of the 21st annual ACM symposium on user interface software and technology, ACM, pp 91–100

Hattori L, D'Ambros M, Lanza M, Lungu M (2011) Software evolution comprehension: replay to the rescue. In: Proceedings of ICPC 2011 I.E. 19th international conference on program comprehension, IEEE, pp 161–170

Kahneman D (2011) Thinking, fast and slow. Farrar, Straus and Giroux, NY

Kirsh D (2010) Thinking with external representations. AI Soc 25(4):441–454

Robbes R, Lanza M (2007) A change-based approach to software evolution. Electron Notes Theor Comput Sci 166:93–109

Saff D, Ernst MD (2003) Reducing wasted development time via continuous testing. In: ISSRE '03: International symposium on software reliability engineering

Saff D, Ernst MD (2004) An experimental evaluation of continuous testing during development. ACM SIGSOFT Softw Eng Notes 29(4):76–85

Schon DA, Wiggins G (1992) Kinds of seeing and their functions in designing. Des Stud 13 (2):135–156

Shadish WR, Cook TD, Campbell DT (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin, Boston, MA

Steinert B, Cassou D, Hirschfeld R (2012) Coexist: overcoming aversion to change. In: Proceedings of the 8th symposium on dynamic languages, DLS '12, ACM, New York, pp 107–118

Suwa M, Tversky B (2002) External representations contribute to the dynamic construction of ideas. In: Diagrammatic representation and inference, vol 2317. Springer, Berlin

Suwa M, Purcell T, Gero J (1998) Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions. Des Stud 19(4):455–483

Thomas D, Johnson K (1988) Orwell—a configuration management system for team programming. In: ACM SIGPLAN notices, vol 23. No. 11, ACM, pp 135–141

Zeller A (1999) Yesterday, my program worked. today, it does not. why? In: Nierstrasz O, Lemoine M (eds) Software engineering—ESEC/FSE '99. Lecture notes in computer science, vol 1687. Springer, Berlin, pp 253–267

Zeller A (2002) Isolating cause-effect chains from computer programs. In: Proceedings of the 10th ACM SIGSOFT symposium on Foundations of software engineering, ACM, pp 1–10

DT@Scrum: Integrating Design Thinking with Software Development Processes

Franziska Häger, Thomas Kowark, Jens Krüger, Christophe Vetterli, Falk Übernickel, and Matthias Uflacker

Abstract Design Thinking has shown its potential for generating innovative, user-centered concepts in various projects at d.schools, in innovation courses like ME310, used by design consultancies like IDEO, and recently even in projects at large companies. However, if Design Thinking activities are not properly integrated with production processes, e.g. software development, handovers become necessary and potentially prevent great ideas from becoming real products.

To reduce the perception of these handovers as acts of "throwing a wild idea over the fence," different integration approaches have been proposed. A seamless integration of Design Thinking into the regular development processes of software development companies, however, is still subject to research.

In this chapter, we present DT@Scrum, a process model that uses the Scrum framework to integrate Design Thinking into software development. Three operation modes, which differ in the ratio between software development and Design Thinking activities, form the foundation of our approach. Development teams chose their respective operation mode after each sprint based on how well the requirements of the product are understood. We present initial applications of our approach in two university courses, and preliminary results of an experiment that tests if and how Design Thinking can benefit from Scrum's planning techniques. The chapter concludes with an outline of future applications of our process model in industry scenarios and experimental validations of further techniques that supplement DT@Scrum.

F. Häger (✉) • T. Kowark • J. Krüger • M. Uflacker
Hasso Plattner Institute for Software Systems Engineering, University of Potsdam, 14482, Potsdam, Germany
e-mail: franziska.haeger@hpi.uni-potsdam.de; thomas.kowark@hpi.uni-potsdam.de; jens.krueger@hpi.uni-potsdam.de; matthias.uflacker@hpi.uni-potsdam.de

C. Vetterli • F. Übernickel
Institute of Information Management, University of St. Gallen, St. Gallen, Switzerland
e-mail: christophe.vetterli@unisg.ch; falk.uebernickel@unisg.ch

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-06823-7_14, © Springer International Publishing Switzerland 2015

1 Introduction

Design Thinking has shown its value as a viable approach for creating innovative, user-centered ideas in projects at d.schools, in innovation courses like ME310, used by Design consultancies like IDEO, and, ever increasingly, during internal projects at major companies. Its core strength is the constant striving for user feedback on prototypes in order to iteratively shape a final solution that provides the maximum benefit for the end user. But good ideas are only half the battle. Turning those ideas into products, may it be physical items, services, or software, requires further efforts that should not be underestimated in their extent. So how do we "bring the prey home" and avoid letting great ideas go to waste?

1.1 Integration Challenges

The key enabler for a transition from idea to product is an effective connection between the idea generation process and product development. Ideally, the two are seamlessly connected, since every piece of information that is lost during handovers reduces the potential for success of the product realization project (Khan and Kajko-Mattsson 2010).

Another factor is the transparency of Design Thinking activities. From a management point of view, it needs to be clear what is being done during Design Thinking projects and how the output can be transformed into a product. From an implementation point of view, developers need to be able to comprehend how ideas have emerged through user research, ideation, prototyping, and testing of prototypes (Katz and Allen 1982). Furthermore, communication between implementation and Design Thinking teams should start early during the projects in order to allow for a realistic assessment of the feasibility of ideas.

The aforementioned challenges might be solved by putting strong regulations on Design Thinking activities. Defining output formats, creating a reporting system for the teams, or extensive planning of all activities throughout the project would come to mind. However, if such bureaucracy hinders the success of relatively straightforward software implementation projects (Beck et al. 2001), what effects could it have on innovation projects? These observations and various ongoing research activities in this area (Vetterli et al. 2012; Lindberg et al. 2012; Hildenbrand and Meyer 2012) show that a balance needs to be found between corporate requirements and creative freedom.

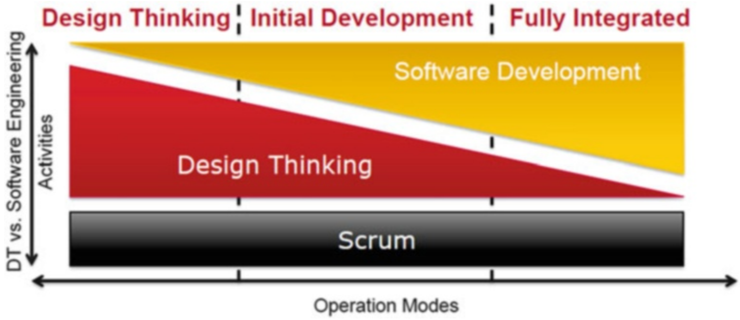


Fig. 1 Integration of Design Thinking into the development process during the different phases of DT@Scrum

1.2 *Outline*

In this chapter, we present DT@Scrum, an approach that combines Design Thinking and Scrum in order to create an agile software development process that can deliver the innovative customer-oriented products and services required by competitive companies.

Scrum provides the overall framework for both development and Design Thinking activities. As presented in Fig. 1, the ratio between the two differs within the three proposed operation modes (*Design Thinking*, *Initial Development*, and *Fully Integrated*). The better the requirements of the product are understood, the more activities are biased towards straightforward implementation tasks. The iterative nature of Scrum allows readjusting the direction of the project and the resulting operation mode in overseeable intervals. A detailed description of the process model along with the included roles, activities, and techniques can be found in Sect. 2.

One of the core techniques of our process model is the so-called Design Planning. This technique adapts Scrum's sprint planning sessions to Design Thinking activities, thereby, potentially allowing for increased structure and transparency of Design Thinking activities. In Sect. 3, we present an experiment that evaluates the effects of Design Planning on the design process and its outcome.

The chapter continues in Sect. 4 with experience reports from two applications of the process model in two university courses. Section "Conclusion and Future Work" summarizes and closes the chapter.

2 DT@Scrum

In our white paper, "Jumpstarting Scrum with Design Thinking," we introduced DT@Scrum, a process framework that aims at seamlessly integrating Design Thinking and Scrum (Vetterli et al. 2013). This section will give a brief introduction to DT@Scrum and our main ideas.

As described in Sect. 1.2, Scrum provides the overall process framework for all activities. This means that teams working with DT@Scrum will use sprints to structure their activities not only during software development but also during design activities, which are often new to team members. In order to let design teams get a feeling for the duration and value of Design Thinking activities, and to enable them to better structure their creative work, DT@Scrum introduces Design Planning. Design Planning adopts planning methods already known from Scrum, e.g. Swim Lane Sizing (Agilepirate 2011) or Planning Poker (Grenning 2002; Cohn 2005), to Design Thinking activities. It includes creating a backlog for design activities, the planning of sprints upfront and an evaluation in a retrospective meeting afterwards.

Additionally, DT@Scrum proposes three different operation modes or phases: the *Design Thinking Mode*, the *Initial Development Mode* and the *Fully Integrated Mode*. The main difference between the phases is the ratio of Design Thinking and development activities. While the *Design Thinking Mode* emphasizes Design Thinking and the *Fully Integrated Mode* focuses on software development, the *Initial Development Mode* aims at balancing the two kinds of activities, thereby allowing the team to gradually move from Design Thinking to software development. With an increasing understanding of the problem and the requirements for a solution, the team decreases Deign Thinking activities and increases software development. The *Design Thinking Mode* explores the problem and the solution space. When the team has formed a product vision that solves the problem, it can start refining the concept in the *Initial Development Mode* by implementing UI concepts, technology tests and first features. After the product vision has been refined and tested with regards to feasibility, viability, and desirability, the team can move forward to the *Fully Integrated Mode* in which the product vision is gradually developed until the software system is fully implemented. Depending on the team's activities, different techniques and roles are needed in each operation mode.

2.1 *Design Thinking Mode*

The *Design Thinking Mode* depicted in Fig. 2 uses Design Thinking techniques to explore the projects' problem statement and the solution space. During this mode, the project team will refine the problem and develop a product vision. The development of low-resolution prototypes, a set of basic User Stories and a clear product vision are the main outputs.

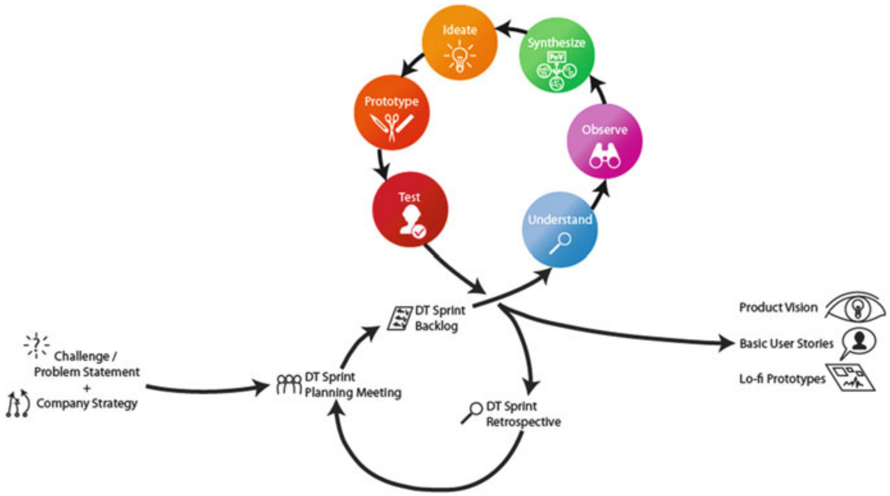


Fig. 2 Overview of the Design Thinking Mode

2.1.1 Prerequisites

The following Prerequisites should be present before starting sprints in the *Design Thinking Mode*:

- Company strategy and a problem statement
- Access to potential users and other stakeholders
- Design Thinking training for the team members

2.1.2 Activities

The activities during this mode follow a basic Design Thinking process, but use Scrum as a process framework. The team starts with a general problem statement and an initial *Understand* phase that helps in collecting information about the project, its goals, constraints, and environment. During the following *Observe* phase, the team gets acquainted with the problem domain, investigates existing solutions, and interviews and observes users and stakeholders. All the information gathered during the first two phases is *synthesized* into the team's *Point of View* on the problem. Based on this *Point of View,* the design team *ideates* aspects of a possible solution. The generated ideas will then be *prototyped* in a rapid fashion that focuses on transporting the main idea instead of creating beautiful artifacts. Each prototype will undergo *testing* with target users. The information gained by *testing* the ideas needs to be *synthesized* again. Depending on the outcome of this *Synthesis* phase, the team will start a consecutive iteration in which it will move on with further *ideation* to refine the idea or, go back to *Understand* and *Observe* phases to answer open questions and investigate new aspects of the problem.

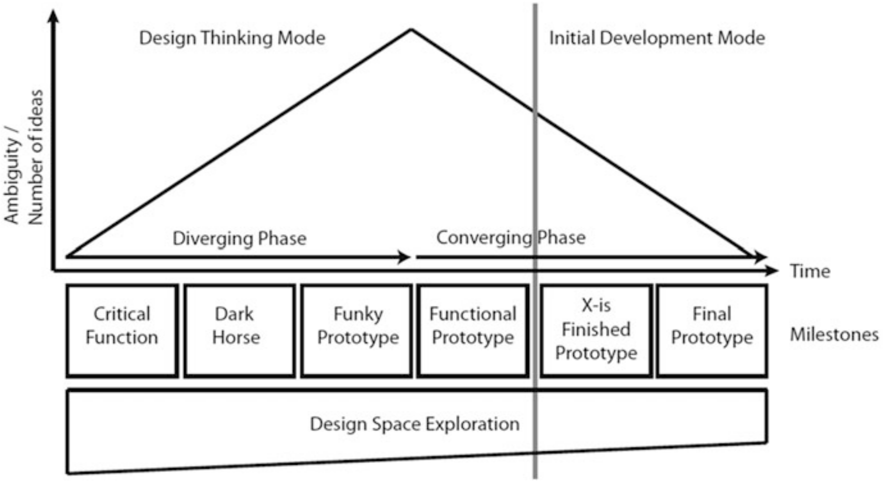


Fig. 3 Milestone concept during the Design Thinking and the initial Development Mode

In order to further structure this operation mode, the milestone concept that is being applied in the global ME310 projects or Embedded Design Thinking (Vetterli et al. 2012) can be adapted by the design team. The following Fig. 3 visualizes a possible distribution of the Milestones to the *Design Thinking Mode* and the *Initial Development Mode*.

2.1.3 Techniques

Design Thinking knows several techniques that help to understand the project environment, the stakeholders, the users, and the design space: 360° research makes it possible to quickly become well-versed on a topic, while user observations and interviews enable the team to understand the user needs and pains. Extreme users (d.school Stanford 2010) can help the design team to get a different perspective on the challenge. Stakeholder maps (Freeman 2010; Thinking.designismakingsense) enable the team to grasp who is involved in the topic. When team members work on different tasks or different activities, short Stand Up Meetings (Yip 2011) help to keep everybody up to date. If it is necessary for other team members to get a deeper understanding of what was achieved or to prepare a synthesis after research interviews, observations, or user testing, storytelling (d. school Stanford 2010) is a potential technique that can be used. Afterwards, different synthesis techniques, like clustering or creating a Persona, a Point of View Madlib, or a 2-by-2 Matrix, can help to discover or convey insights and findings (d.school Stanford 2010). When the team has found its current point of view on the challenge, brainstorming possible solutions generates ideas, which can then be prototyped. During this phase, prototyping is used to understand the users and the challenge, as well as to quickly validate ideas and possible solutions.

Therefore, rough prototypes that are fast and easy to build work best for this purpose. These include cardboard or paper prototypes of, for example, hardware components, sketches of user interfaces, or even role plays of a situation. Testing the prototypes with actual users is essential to understand flaws of the current solution and discover further user needs and pains. Testing can be done by observing users while they are trying everything out and then interviewing him afterwards.

2.1.4 Roles

The *Design Thinking Team* is responsible for planning and executing the Design Thinking sprints. A design team will usually consist of three to six people from different areas of expertise as needed for the software under development, e.g. accounting, sales people, UI designers, developers and consultants.

The main task for the *(Potential) User* is to provide input on the topic and his problems and to give feedback on ideas, prototypes, and the direction of the project. Potential Users will be interviewed and observed by the design team. In the beginning a broad range of users will be interviewed, but after the team reaches a decision about a target user or user group, it tries to secure users of that target group for constant testing and feedback cycles.

The *Corporate Liaison/Project Sponsor* has a strong interest in the project as he represents the group that defined the initial challenge. He serves as a contact person for the team. The responsibilities of this team member include providing interview partners and introductory material for the challenge, facilitating communication with other sections of the company to avoid duplication of efforts, enabling synergetic effects between teams, and allowing reuse of existing software. Additionally, he provides feedback on ideas and prototypes in a way that is similar to the users. In a corporate setting, this role can be taken over by a customer representative and/or a manager.

In Design Thinking processes, teams are often supported by *Design Thinking Coaches*. The responsibility of the coach is to introduce useful techniques, moderate discussions, ensure that the team is focused on its task, and to moderate team dynamic issues, such as conflicts or motivational issues. The *Scrum Master* in Scrum projects makes sure that the Scrum team follows the process structure and moderates discussion during planning and reflection meetings. In our merged process these roles could be merged into one: the *Process Master (PM)*. The person in this role would be responsible for the team's adherence to the overall process and moderate team discussions.

2.1.5 Deliverables

This mode generates different low-resolution prototypes as well as one more sophisticated solution prototype. The solution prototype together with insights gathered throughout the *Design Thinking Mode* should generate a clear solution vision and elaborate why all the aspects of the prototype have been designed in a specific form. Additionally, non-functional requirements for the development of the product, and an initial set of high-level User Stories that describe the core functionality of the intended system need to be created. Documentation of the learnings and insights, which led to the functional prototype, should be created, to be able to trace back decisions made within this mode.

2.2 *Initial Development Mode*

The *Initial Development Mode* shown in Fig. 4 focuses on further exploring the product vision created during the previous operation mode. The main goal of the team during this mode is to start implementing, testing, and refining different aspects of the solution. A set of mid to high resolution prototypes, refined user stories, and non-functional as well as technical requirements are the targeted outcomes of this mode.

2.2.1 Prerequisites

In addition to the prerequisites described in Sect. 2.1.1, the following prerequisites should be present before starting sprints in the *Initial Development Mode*:

- Clear product vision
- Initial set of high-level User Stories
- Functional prototype that resulted from the Design Thinking phase
- Pool of low resolution prototypes
- Initial list of non-functional requirements

2.2.2 Activities

The main activity of this mode is to refine the solution prototype and the product vision. This is achieved by identifying features or design aspects of the solution prototype that need further clarification and testing with regard to feasibility. These can then be prototyped in the form of user experience (UX) prototypes, a proof-of-concept feature implementation, an implementation that tests technical feasibility or an implementation that explores possible technologies. All prototypes will be developed within Scrum sprints. Prototypes that provide a user interface should be tested with target users for maximum user satisfaction. The information gained from prototyping and testing can then be used to further refine the implementation in the next sprints and to refine the user stories and add additional non-functional and technical requirements. In addition, the system architecture and the integration concept should be prototyped.

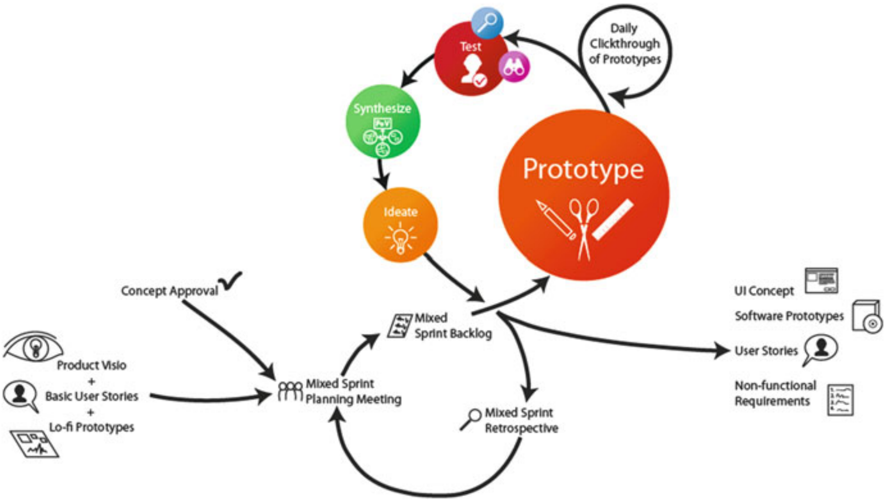


Fig. 4 Overview of the initial Development Mode

2.2.3 Techniques

Core techniques needed during the *Initial Development Mode* include low- and mid-fidelity UX prototyping, user story mapping, and programming.

Simple sketches on paper provide a great low-fidelity UX prototyping tool to create fast UI prototypes. They can be used to test different arrangements of the content, the navigation between different pages or interaction concepts. Paper prototypes can also be "interactive" during user testing if one person in the design team manually changes the UI by drawing additional content or adding and moving pieces of paper around.

Wire framing is a mid-fidelity UX-prototyping technique that uses simple sketches like widgets and controls to build a user interface prototype. Various tools like pidoco[1] or gomockingbird[2] exist, that provide the user with a variety of building blocks to build screens or even clickable prototypes. In cases where a more sophisticated or hi-fidelity UI prototype is required (e.g. to discuss progress with management) tools like Keynotopia,[3] which enable click-able Keynote/PowerPoint UIs or fast HTML prototypes can be used.

User story mapping is a technique that helps teams to understand the functionality of the system under development, identify holes and omissions in a backlog, and plan releases that deliver value to user and business. The User Story Map (Patton 2009) arranges the main activities from left to right in an order that makes

[1] https://pidoco.com/

[2] https://gomockingbird.com/

[3] http://keynotopia.com/

sense, e.g. in a workflow. Task centric User Stories are also arranged from left to right under the activity they belong to. Tasks that can occur in parallel will be placed vertically under one another.

A daily clickthrough of the current prototypes ensures that everyone in the team is up to date on the explored concepts and findings.

2.2.4 Roles

The *Scrum Team* is responsible for the planning and execution of the development sprints. A Scrum team usually consist of eight to ten developers drawn from the design team of the former mode and additional developers from areas of expertise as needed for the software under development, e.g. back end developers, front end developers, database experts, UI developers, etc.

The main task of *(Potential) Users* during this mode is to test the different prototypes developed and give feedback.

The main task of the *Corporate Liaison/Project Sponsor* during this mode is to give feedback on the developed prototypes and the general direction of the project. Additional responsibilities are facilitating communication with other sections of the company and advertise the project progress.

The *Product Owner* is the representative of the customer. He is responsible for filling the backlog with user stories and for prioritizing them. In our combined process model, the product owner can be one of the members of the design team from the previous mode, e.g. a user researcher.

The *Process Master* has the same responsibilities as defined in Sect. 2.1.4.

2.2.5 Deliverables

The deliverables in this mode are mainly the created prototypes and the results from testing them. These include end-user tested UX prototypes that led to further functional and non-functional requirements, and back end spikes to show the feasibility of required functionality and technical requirements. With the insights gained from developing and testing the prototypes, the user stories can be further refined and new user stories can be added. Finally, a clear specification of integration within the company context needs to be created. This includes identifying interdependencies with other, already existing systems or potential for reuse of existing software components in the final implementation.

2.3 *Fully Integrated Mode*

The *Fully Integrated Mode* illustrated in Fig. 5 mainly complies with a Scrum development process, enabling the team to work towards a final product in

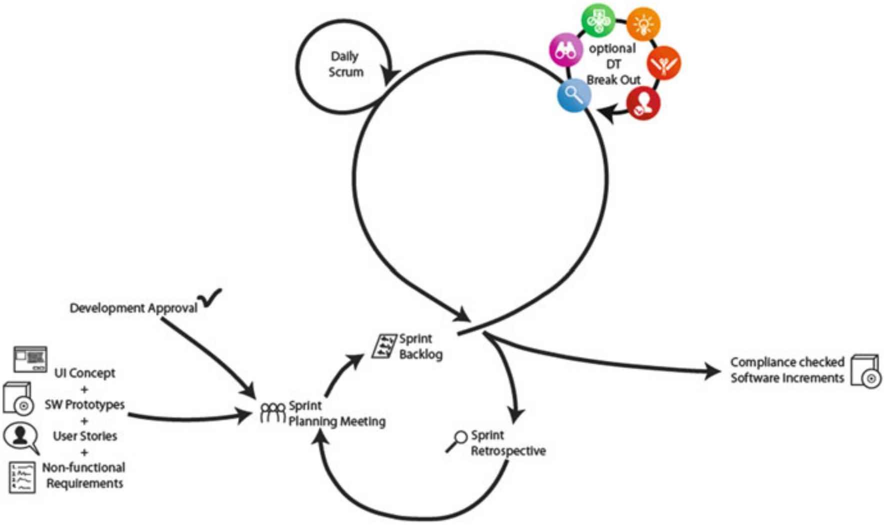


Fig. 5 Overview of the fully Integrated Mode

incremental steps. In case of blockers in the development process, Design Thinking tools will be initiated, hence the Design Thinking application is ad-hoc. A close observation of the development process is needed to quickly react to blockers with the adequate Design Thinking tool.

2.3.1 Prerequisites

In addition to the prerequisites described in Sect. 2.2.1, the following prerequisites should be present before starting sprints in the *Fully Integrated Mode*:

- List of technical requirements
- Prioritized list of detailed user stories
- Set of Proof-of-concept implementations
- Set of UX Prototypes

2.3.2 Activities

The activities during this mode follow a basic software development approach using Scrum as a process framework. The team or teams focus on development of software increments as well as deployment and maintenance concepts. In case features are not defined well enough or problems arise, the team can choose to include short Design Thinking bursts in the activities to refine a feature idea or find solutions to the problem. Thus, Design Thinking in this mode, compared to the other two modes, does not focus on creating insights prior to the software development process. It is rather creates ad-hoc insights and different solutions to overcome some impassable blockers.

2.3.3 Techniques

This mode is completely dedicated to turn the product vision into a fully functioning piece of software. Thus, the entire spectrum of software engineering techniques can and should be used. For example, the practices proposed by Extreme Programming (Beck 2000) are very well suited for Scrum projects. They include test-driven development, continuous integration, and different review techniques to maintain code quality, collective code ownership, and continuous customer testing.

2.3.4 Roles

The responsibilities of the *Scrum Team* during this mode are similar to the preceding mode. They plan and execute the sprints implementing functional software increments. If needed, additional Scrum teams can be added to allow for parallel development.

The main task of *(Potential) Users* during this mode is to test the software increments and give feedback.

The main task of the *Corporate Liaison/Project Sponsor* during this mode is to give feedback on the developed software increments. Additionally, this team member should still facilitate communication with other departments of the company and promote the project progress.

The *Product Owner* has similar responsibilities as described in Sect. 2.1.4.

During this mode the *Process Master (PM)* has the same responsibilities as during the other modes. In this mode it is of special importance that the PM can quickly react if blockers are stopping the development process and provide the team with the right Design Thinking tools to help them.

2.3.5 Deliverables

The *Fully Integrated Mode* focuses on creating tested, working software. Hence, all developments should be potentially shippable by the company. This means that the software adheres to certain product standards and is deployable. The teams should therefore also create, or at least keep in mind, a strategy of how their software can be delivered to the end user. This is rather straightforward in the case of mobile apps, but when developing on-premise software that integrates with existing landscapes the team needs to explicitly reserve time to create a deployment strategy. Finally, developers should not only blindly implement the given stories but be open-minded about potential improvements. Hence, they should also capture their own ideas or suggestions, and, if applicable, transform them into user stories for upcoming sprints.

2.4 Large Scale Projects

Larger software projects, like the development of a complex ERP solution, require a large number of developers possibly split into several development teams. Solutions to solve this problem already exist, for example the Scrum of Scrums or Meta Scrum. In this technique, the individual Daily Scrum of all Teams is followed by a Daily Scrum of Scrums with an ambassador from each team, who will give a progress report from his team and take back important information to his team members. If necessary, this technique can be used on multiple levels. Ambler (2009) or Larman and Vodde (2008, 2010) present examples and case studies on how agile processes can be scaled for large project teams and explain appropriate techniques. We believe that a similar scale up of design teams for the *Design Thinking* and *Initial Development Modes* would not be helpful. Instead, we proposed that a regular design team of four to eight people will work on the project during the *Design Thinking Mode*. During the *Initial Development Mode* the team can split into multiple mixed teams and work on different projects that follow a product idea from the *Design Thinking Mode*. As an alternative, the design team can split into multiple sub teams and work on parts of the product vision created during the *Design Thinking Mode*. The teams or sub teams will then evolve even more fully into development teams, who will perform the sprints in the *Fully Integrated Mode*. Figure 6 illustrates the flow of project teams during the modes.

2.4.1 Summary

In this section we presented DT@Scrum, our initial concept to seamlessly integrate Design Thinking and Scrum. It comprises three operation modes, which provide a different ratio of Design Thinking and software development activities. We presented the general activities of each mode, supporting activities and the involved roles. We want to invite researchers and practitioners to give us feedback on our ideas and try out DT@Scrum. We would gladly support projects that want to try out DT@Scrum, e.g. with training and coaching.

3 Design Planning

As described in Sect. 2, Design Planning is an important concept for the DT@Scrum approach. We want to ensure that these techniques are adaptable to Design Thinking and help to estimate workloads and durations in order to be able

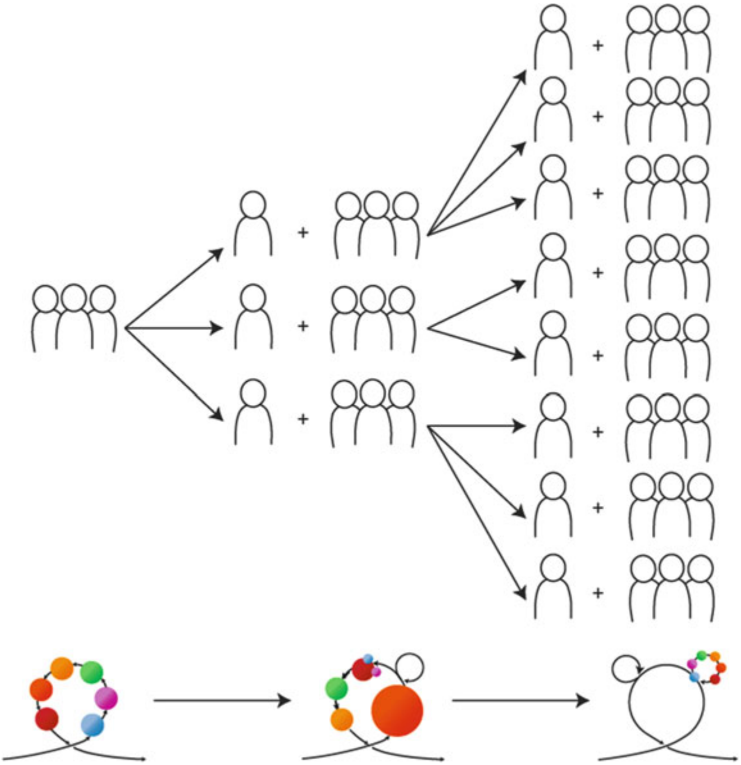


Fig. 6 Team scale up during operation modes

report them to the management but also help the teams to organize themselves. The hypothesis underlying this concept is as follows:

Running design tasks in sprints, estimating and planning them accordingly, and using a regular retrospective can help the team to better understand their process and get a feeling for design tasks.

However, introducing a constraint like planning and following the plan could negatively affect the outcome of the design process as it limits the team's creative freedom. Therefore we aim to answer the following research questions with the help of a 3 h experiment:

- How does planning affect the team's design process?
- How does planning affect the design team's view of their process?
- How does planning affect the outcome of a design task?

3.1 Experiment Setup

The experiment is comprised of two design challenges, each 1 h long, and a series of questionnaires. In one challenge, the team can decide how to use the hour themselves. In the other challenge, the team is required to use some time at the beginning of the hour to collect all tasks they want to do, assess them, and plan the course of the remaining time. The two design challenges are similar in terms of complexity. One challenge asks the students to design the perfect transition from work to free time for a specific user. The other challenge asks the team to design the perfect start into the day for a specific user. The user is available throughout the experiment for interviews and testing sessions. After each challenge the participants are asked to fill out questionnaires asking them to reflect on their process, rate the innovation potential and desirability of the created solution, and rate the value of planning tasks upfront. After the experiment the participants are asked to fill out another questionnaire asking them to compare the two challenges and how they would run a third similar challenge.

3.2 Preliminary Results

We initially ran the experiment with teams of former ME310 projects. In those experiments, we first ran the challenge without asking for a plan. The teams did not decide to plan anything upfront and ran into problems in the second half of the hour realizing they did not have enough time. Most of the time was spent on interviewing the user. This amounted to between 13 and 15 min. Prototyping was rather short and very ad-hoc, it started after minute 51 and took about 3–4 min. Testing, accordingly, started after minute 54 and took 3–5 min, basically the remaining time. Interestingly, one team managed to do a 2 min iteration on their prototype and test it again. In the questionnaires it shows that the teams experienced time pressure and moments of chaos when it was unclear how to proceed. It was mentioned that planning or better time management would make sense. However the teams felt productive and were satisfied with the solution with regards to the available time.

In the first version of the second challenge we requested the teams spend 15 min planning using swimlane sizing as a planning tool. In this test we also used “design the perfect wallet for a specific user” as the design challenge. When setting this challenge, we got the feedback that the wallet exercise was already done several times by participants. It also asks for a specific product instead of addressing a general user need. Thus, the team focused on improving the wallet itself rather than creating the most desirable solution with regards to the user’s needs. Furthermore, 15 min of planning for a 45 min challenge was seen as much too long. The introduction of a new tool was also perceived negatively as it takes several attempts to fully comprehend a new tool.

In the second version of the planned challenge, we simply asked the teams to take some time in the beginning to plan the hour. This allows the team to choose freely how much time to spend on planning and what techniques/tools they want to use to plan. We also changed the challenge to "design the perfect transition from work to free time for a specific user", because it is closer to the unplanned challenge and allows a product or a service as solution.

In these planned challenges, teams placed the greatest importance on interviewing and prototyping by allotting the most time for these activities. They used between 10 and 18 min for interviews and 6–12 min for prototyping. Testing again took 3–4 min, which was the remaining time of the challenge. The team that managed an iteration in the first run decided not to do one in the second run even though there were still 2 min left. In the questionnaire it showed that the teams still experienced time pressure, some of them even more than in the first challenge. This is probably due to the fact that none of the teams used buffer times in their schedule, and thus missed the chance of adopting the schedule while working. It also showed that the process and the steps to take during the challenge were clearer. While overall the teams found planning useful for longer challenges, they also felt that it was too time consuming when 1 h was allotted. They found it good to decide on tasks, but forgot to include buffers. The value of collecting tasks was rated an average of 2 on a 1–5 scale (1 = very good, 5 = very bad). The value of estimating and ordering the tasks was rated an average of 2.5 on a 1–5 scale (1 = very good, 5 = very bad).

Figures 7 and 8 show the timelines for the first and second run from two of the participating teams. Comparing the course of the two challenges, we found that in the first challenge tasks tend to get shorter towards the end of the challenge, probably due to the fact that time was running out. On the second run, with planning, the team chose which tasks needed the most time and the timeline reflects these choices. Another interesting observation that can be seen in the timelines is that the teams tended to do further interviews, ad-hoc during clustering or synthesis during the unplanned challenge. This behavior was decreased in the planned challenge. While the experiment setup allows and encourages the team to ask the user further questions, we believe, that the decline in follow up questions can mean one of two things. Either the teams are more focused on the task they are currently working on or they stop challenging their thoughts and ideas in the planned challenges. This fact should be further observed with other teams in order to evaluate which of the possible explanations is correct.

Further comparing the outcome of both challenges we found that the prototypes in the second challenge were more tangible and self-explanatory, probably due to the longer time taken for prototyping. Figures 9 and 10 show the prototypes from teams A and B for their first and second challenge.

Comparing the ratings for desirability and innovation potential of the solution, teams rated the desirability of the solution higher in the second challenge. A rating of the solutions through the users and our Design Thinking coaches also found the solutions from the second challenge to be more desirable. The innovation potential was rated the same for both challenges by nearly all participants.

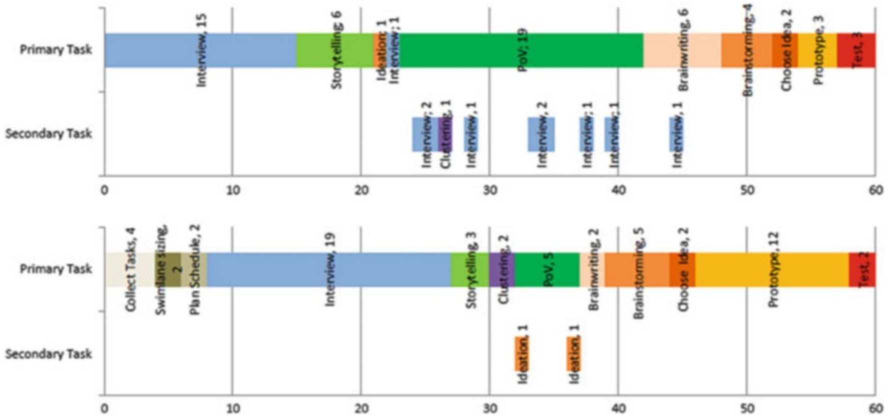


Fig. 7 Comparison of timelines for both challenges—Team A

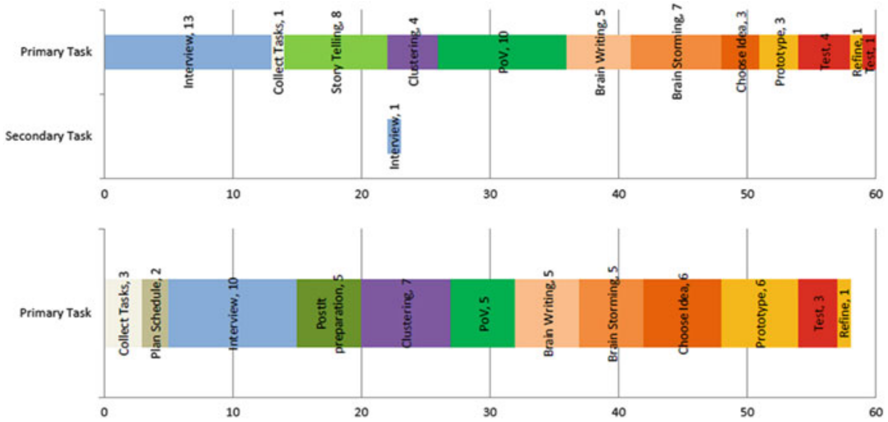


Fig. 8 Comparison of timelines for both challenges—Team B

Comparing ratings for stress, teams found the first challenge to be less stressful. From their explanations we could see that they felt time pressure mainly at the end, when there was no time left for prototyping and testing. While in the second challenge there was time pressure felt throughout as the team tried to keep the schedule. The first challenge was rated to be more successful.

When asked how the participants would run a third similar challenge, all participants wrote that collecting tasks and ordering them helps and, thus, would be included. However, the addition of buffers to the general plan was requested to allow for changes during the challenges.

To sum up our findings, planning created more time for prototyping, which again led to more tangible and self-explanatory prototypes. Keeping the schedule was experienced as stressful by the teams. In a third one hour challenge, teams would

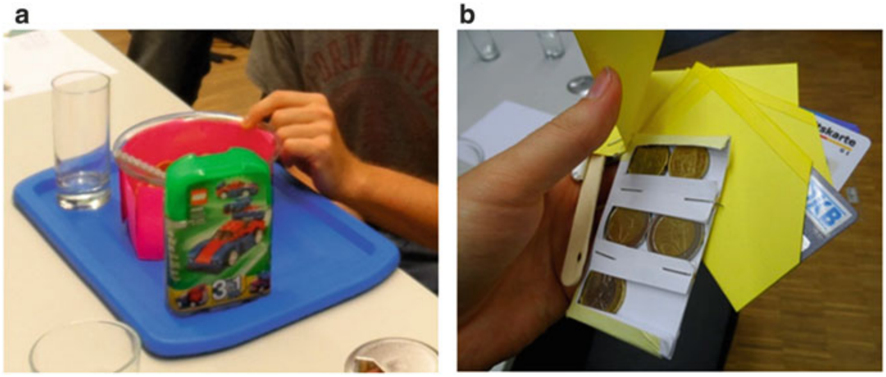


Fig. 9 Prototypes of Team A. (**a**) Prototype from unplanned challenge, (**b**) prototype from planed challenge

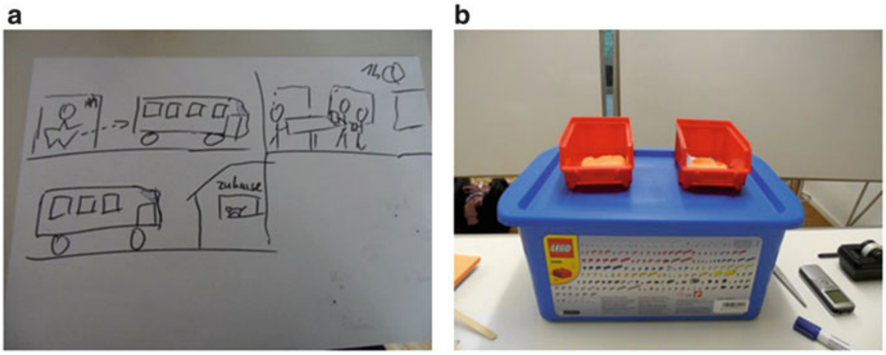


Fig. 10 Prototypes of Team B. (**a**) Prototype from unplanned challenge, (**b**) prototype from planed challenge

make a plan but use buffers. When rating the planned challenge, the teams mentioned that it provided a better overview about the required tasks and helped in comprehending the process. Overall, the preliminary results are in favor of our hypothesis.

3.3 Outlook

From the experience of our first experiments we decided to request buffers when planning. We also decided to randomly switch planning between the first and the second challenge in order to analyze how teams use the second challenge when planning was introduced before beginning. With these changes we aim to conduct the experiment with further teams. In this way, we will have a broader range of teams to analyze and verify the preliminary findings.

To further evaluate planning during design activities in general, we want to observe and interview ME310 and d.school teams that use long and/or short term planning tools (e.g. Kanban Board, day plans) and investigate their motivation and strategies when planning. Further interviews, with d.school teams that do not use planning, will help to reveal problems with existing techniques and obstacles for planning activities in design teams.

Planning for challenges that only take an hour to complete, as done in our experiments, has a bad ratio between planning and actually working. Planning upfront, therefore, "steals" working time in such overseeable settings. On the other hand, if planning is implemented for longer time frames it gets harder to keep an overview of all the necessary steps, make detailed estimates, and foresee problems and changes in the project. Therefore, Scrum suggests planning sprints of 2–4 weeks for software engineering activities. With longer running observation of ME310 teams that use design planning, we want to determine the optimal sprint size for Design Thinking activities. Additionally, these observations could reveal insights on the usability of Design Planning over the course of a project. These insights will then be validated with a quantitative questionnaire. To test Design Planning in a setting with a useful planning to working ratio, we are also investigating the possibilities of a longer running Design Planning experiment, e.g. 1 or 2 day, which would also allow us to introduce additional planning tools.

4 Application in Software Project Courses

In order to test our ideas and gain first-hand knowledge about using DT@Scrum, we started testing it in project based software engineering courses. These courses provide a low consequence environment in which we can easily observe and interview the participants and adapt the process and the used techniques as needed. Additionally, we can ask the participants to test various Design Thinking and Scrum techniques, thus allowing us to identify those that are best suited for software focused projects. In the following we introduce the two courses which we adapted for that purpose, their general setup, the participants, as well as first observations.

4.1 Bachelor Projects

In order to acquire a Bachelor degree in IT Systems Engineering at the Hasso Plattner Institute students need to take part in a bachelor project. The main goal of the bachelor project is to prepare the participants for their work in the software industry and allow them to apply the knowledge and skills learned during their studies.

Over the course of two semesters, teams of four to eight students will solve a real life challenge provided by their project partner, the associated chair, or a company requiring a software solution to their problem. The team will be supervised by a professor and up to three research assistants. The project is composed of two parts, the research phase and the implementation phase. During the first semester the students will work on the project 2 days a week. As projects come from various industries, e.g. healthcare, or automotive, this time is typically used to get to know the industry partner and the challenge. The students learn about the problem domain, learn specific skills needed for the project, and come up with requirements for their software solution. During the second semester, the students work on their projects 4 days a week. This semester is used for the actual implementation of the software solution. During this phase, Scrum is a popular process framework as most bachelor students at HPI already know it from former courses.

In 2013/2014 we are offering two bachelor projects at our chair. The first project (BP1) focuses on managing the life cycle of data, a very technical problem, while the second project (BP2) focuses on the development of tools in the area of computer aided software engineering.

4.1.1 DT@Scrum in Bachelor Projects

As described, bachelor project teams often use Scrum as a process framework during their projects. So far, requirements for the solution are mainly given by the project supervisors or the industry partner, who serves as the product owner. Because the first semester of a bachelor project already aims at understanding the challenge, the environment of the challenge, and collecting requirements for the software solution, this semester is ideal to integrate Design Thinking into the team's processes. Design Thinking is an optional course for bachelor students at HPI, so we cannot assume that all students are familiar with the process and its ideas. Therefore, we introduced an initial Design Thinking workshop 3–4 weeks into the project, which also serves as a first synthesis point for the team.

Before the workshop, both teams focused on researching their topic by reading papers and benchmarking existing solutions to their challenges. Additionally, the team from BP2 was introduced to techniques for performing observations and interviews. They conducted several interviews with software developers employed at the project partner. After this initial research, we held a 1 day Design Thinking workshop with each team, in which we briefly introduced Design Thinking and DT@Scrum. The workshops aimed at bringing the team together, forming a joint understanding of the problem, and building initial prototypes. We introduced personas, brainstorming, UI paper prototyping, and storyboards during these workshops. The next steps for both teams will now be to start prototyping their initial ideas and verifying them with their end users or project supervisors, and to start ensuring the technical feasibility of features and the applicability of the chosen technology with software prototypes. During these steps we will support the teams as Design Thinking coaches and introduce them to further Design Thinking

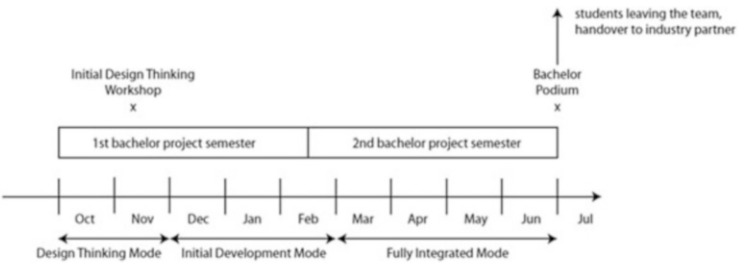

students leaving the team, handover to industry partner
Initial Design Thinking Workshop
Bachelor Podium
1st bachelor project semester
2nd bachelor project semester
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Design Thinking Mode
Initial Development Mode
Fully Integrated Mode

Fig. 11 Timeline of the DT@Scrum during the bachelor projects

techniques suited for their project and progress, such as storytelling, different types of prototypes, or user testing.

Nevertheless, as the projects are focused on producing a functional software system, the *Design Thinking Mode* and the *Initial Development Mode* will take place this first semester, while the second semester is reserved for the *Fully* ***Integrated*** *Mode*—the actual development. The following Fig. 11 illustrates the expected timeline for DT@Scrum in the bachelor projects.

4.1.2 Initial Observations

The Design Thinking workshops provide a good way for the teams to summarize their learnings so far and start moving towards a solution. Brainstorming and prototyping initial ideas help them to form a joint point of view on the project and possible solutions. Participants experience the workshop as a good introduction and a means to get to know each other better. Especially prototyping is experienced as useful, because it helped participants to realize what they usually do not think about/forget when working on software, and because it helps to come up with new ideas along the way.

Beyond the initial workshops, rapid prototyping with paper prototypes or storyboards remains an asset for the teams. It allows them to externalize their ideas, discuss them with the project supervisors, and their external partners. They use an accompanying wiki system to store pictures of all prototypes in order to increase traceability of their ideas and permanently capture the feedback. Based on the prototypes, technical challenges were identified (e.g., prediction of query runtimes for large database systems) and captured as tasks within the ticket system. These challenges are prototypically solved in the second project phase and then combined to create the final prototype in the third phase.

4.2 *Global Team-Based Product Innovation and Engineering*

The course “Global Team-based Product Innovation and Engineering” is a joint course with international universities. It originates in a course called ME310 (Carleton and Leifer 2009), where mechanical engineering students collaborate with students at international partner universities, like the Hasso Plattner Institute, to work on innovation challenges posed by global corporations. Over time, the partner universities have started to cooperate with each other, allowing them to run more than one project in their course, thus creating a large and active network of universities, professors, and research and teaching assistants interested in Design Thinking and its application to various fields of studies, e.g. mechanical engineering, industrial engineering, product design or business administration. Over the course of 9 months, a team of six to eight students from two universities, with three to four students each, work together on the challenges presented by their industry partner.

As depicted in Fig. 12, the 9 months are split into three phases with different goals. In the first phase the team concentrates on understanding the challenge, exploring the problem domain, and researching existing solutions. During this phase the team observes and interviews end users, benchmarks existing solution and analog situations, and creates first low-fi prototypes. During the second phase, the team starts investigating possible solutions with different prototypes. Finally in the third phase the team works towards a final, sophisticated, product-like prototype of their solution. An additional challenge for the teams is managing the dialog between the globally distributed sub teams and their industry partner. All three phases are structured by milestones in the form of weekly meetings and assignments handed out roughly every 2–3 weeks, similar to the milestone concept described in Sect. 2.1.2.

In 2013/2014 we are running the course at our chair with three projects of which two have a challenge that involves software engineering. A total of 12 students on the HPI side work on the projects and are supported by a team of 6 coaches.

4.2.1 DT@Scrum in ME310

As described before, the course setup follows the Design Thinking process, additionally structured by various prototype milestones. The milestones prescribe a form of pulsing by requiring a prototype as deliverable every 2–3 weeks. Additionally the weekly meeting of all teams and coaches provides a sprint-like timeframe that requires reporting of finished and ongoing activities. Reflection sessions, one with the coaches and one team internally, allow the team to recapture the week’s activities. However, planning tools and planning sessions are not required so far. Additionally, the ME310 projects end with a product-like prototype but miss actual productization.

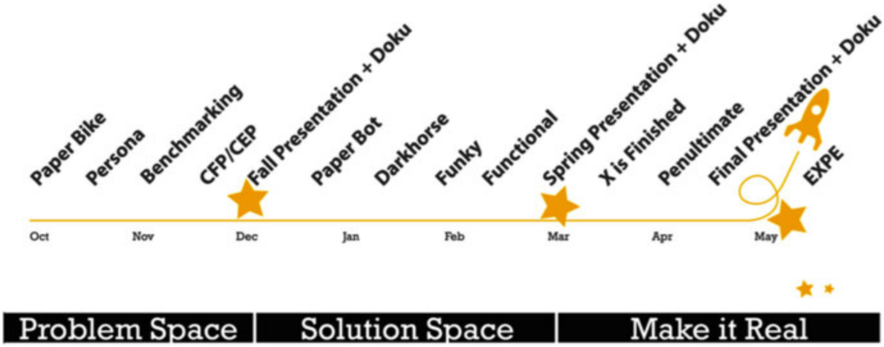


Fig. 12 Timeline of the course global team-based product innovation and engineering

With the ongoing projects we implemented a 2-day Kick-Off workshop that included short experience projects, like the Wallet Project (d.school Stanford), to teach the concept of a Design Thinking project in 1 h. The second workshop day was comprised of a 1-day design challenge to apply the concepts learned on the first day. Throughout the projects we are planning to implement different workshops to introduce or refresh Design Thinking techniques. We will also implement further short experience projects, e.g. the Lego Exercise that teaches the concepts of Scrum in a few hours (HPI 2011). Furthermore, we will introduce coaching sessions by ME310 alumni to gain experiences with knowledge transfer between teams. By applying these different coaching and teaching strategies, we hope to provide our students with a positive learning experience and create valuable coaching guidelines and teaching techniques to enhance our process proposal.

We will also introduce Design Planning to the students of the current projects and let them plan their prototype sprints, to gain experience with the technique over a longer period of time.

4.2.2 Initial Observations

With a first software engineering focused ME310 project, which took place from October 2012 until June 2013, we tried to test some of our DT@Scrum concepts. Since ME310 is a course for mechanical engineering students, it focuses on prototyping and creating physical products, where software artifacts are merely a byproduct, a fact that frequently became a problem in the project. The prototypes of ME310 build on each other and support the refinement and reuse of components. For software prototypes this is harder to achieve. A modular approach to a software system that is integrated later in the project requires decisions on system architecture, interface concepts, technology to use, and so on. A sound decision on fundamental concepts cannot be made early in the project as only vague knowledge has been acquired. On the other hand, if teams start coding too late, the given timeframe is not sufficient to implement the full functionality. Except for wireframing and UI prototypes there seem to be few tools that allow fast and simple software prototyping. Thus the teams frequently struggled to create the requested prototypes and test them in time.

4.2.3 Outlook

Based on these observations we decided to refrain from implementing DT@Scrum with software only projects in our ME310 course. Instead we aim to collect experience with those aspects of DT@Scrum that make sense in the ME310 context and evaluate possibilities for a software-based global Design Thinking course.

With the two ongoing projects that involve software engineering, we will further observe how software engineering activities can be supported in an ME310-like context, evaluating which prototypes make sense and how software prototyping can be better supported. Additionally, we consider the possibility of launching follow-up projects that aim to implement a software solution based on the outcome of the ME310. This would allow us to gain experience with the *Fully Integrated Mode* and test different handover and knowledge transfer concepts.

In the future, we strive to apply this knowledge by setting up a course resembling ME310 with multiple teams that will use Design Thinking to tackle software engineering challenges. Within such a course, henceforth called CS310, we would be able to test DT@Scrum in a suitable context. We could compare different tools and techniques, team setups and coaching strategies by comparing multiple teams that, for example:

- Apply different tools and techniques to the same process steps,
- Experiment with the integration of Design Thinking and other software engineering processes,
- Or test different team setups, e.g. using someone from the design team of the *Design Thinking Mode* as a Product Owner in the following modes.

To allow us to test all three operation modes of DT@Scrum, we plan to setup CS310 as follows. The CS310 design team normally starts off investigating the design and solution space and forming a solution idea. When it comes to implementing functional prototypes the team will be assisted by a team of additional student developers. After a final product prototype has been developed, the additional developers will take over the project and further implement a product-like version. In addition to ensuring the projects run through all desired modes, this setup also ensures that team members will join and leave the project. This will give us an opportunity to test concepts for transition workshops between modes. The following Fig. 13 outlines the timeline for DT@Scrum in such a CS310 project. As can be seen, we plan a setup of three transitional workshops. The Kick-Off Workshop will inform all team members about the project and its goals. The Idea Handover Workshop will help transfer knowledge from the first *Design Thinking Mode* into the *Initial Development Mode* and bring new team members up to date. Finally, the Product Backlog Creation Workshop will transfer the knowledge from

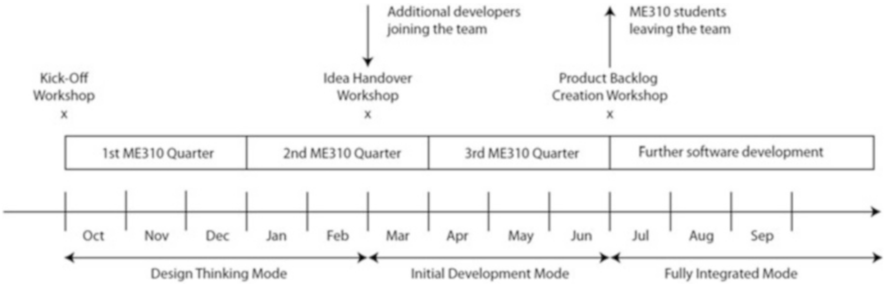


Fig. 13 Timeline of the DT@Scrum during the CS310 projects

leaving team members and from the *Initial Development Mode* to the *Fully Integrated Mode*. It also ensures that all team members help to create the Product Backlog necessary for the Scrum development sprint.

5 Conclusion and Future Work

DT@Scrum is an approach that integrates Design Thinking with Scrum in order to provide a process that seamlessly connects the generation of innovative ideas and their implementation. At the core of DT@Scrum are three operation modes. The *Design Thinking Mode* allows the project team to transition from exploring the problem and possible solutions. The chosen solution is refined and initial coding efforts verify the technical feasibility within the *Initial Development Mode*. Finally, in the *Fully Integrated Mode* the proposed solution is implemented as a product. To achieve its goals, each mode prescribes activities, the roles involved in these activities, and supporting techniques.

Another core element of DT@Scrum is Design Planning, the application of planning and reflection techniques from Scrum to Design Thinking activities. With this concept we hope to achieve a greater transparency of Design Thinking activities for management and team members. We will test the concept and evaluate its effects on design teams and the outcome of design tasks with the help of Design Planning experiments.

In order to gain experience with the implementation of DT@Scrum we partially implemented it in bachelor projects and the ME310 courses at our chair. With the experience and feedback gained in those courses, we plan to create a course that adapts the concept of ME310 to a software engineering focused course, CS310. The course will then serve as a testbed for DT@Scrum, helping us to validate our ideas and improve DT@Scrum.

Apart from the future activities described in Sects. 3 and 4 we want to run additional test projects with partner companies to further evaluate our process model. Such on-site projects at one of our industry partners will help us to test our process with teams in actual enterprise settings, allowing us to identify enterprise specific challenges and opportunities for DT@Scrum. Therefore, we would like to invite you to try out our approach and give us feedback. We would gladly support your efforts by providing workshops, coaching, and teaching materials.

Furthermore, we want to open a discussion with practitioners and researchers on the concepts of DT@Scrum in general, their own ideas and experiences with implementing Design Thinking in a software engineering context, and the possible adoption to different company settings.

References

Agilepirate (2011) http://www.theagilepirate.net (2011) Swimlane sizing—complete and fast backlog estimation. Retrieved from http://theagilepirate.net/archives/109

Ambler SW (2009) The Agile Scaling Model (ASM): adapting Agile Methods for complex environments. Retrieved from ftp://ftp.software.ibm.com/common/ssi/sa/wh/n/raw14204usen/RAW14204USEN.PDF

Beck K (2000) Extreme programming explained: embrace change. Addison-Wesley Longman Publishing, Boston, MA

Beck K, Beedle M, van Bennekum A, Cockburn A, Cuningham W, Fowler M, Grenning J, Highsmith J, Hunt A, Jeffries R, Kern J, Marick B, Martin RC, Mallor S, Shwaber K, Sutherland J (2001) The Agile manifesto. Technical report. The Agile Alliance

Carleton T, Leifer L (2009) Stanford's ME310 course as an evolution of engineering design. In: Proceedings of the 19th CIRP design conference—competitive design. Cranfield, UK

Cohn M (2005) Agile estimating and planning. Prentice Hall International, Upper Saddle River, NJ

d.school Stanford. The Wallet project. Retrieved from https://dschool.stanford.edu/groups/designresources/wiki/4dbb2/The_Wallet_Project.html

d.school Stanford (2010) Bootcamp bootleg. Retrieved from https://dschool.stanford.edu/wp-content/uploads/2011/03/METHODCARDS2010v6.pdf

Freeman RE (2010) Strategic management: a stakeholder approach. Cambridge University Press, Cambridge

Grenning J (2002) Planning Poker or how to avoid analysis paralysis while release planning. Technical report. Retrieved from http://renaissancesoftware.net/files/articles/PlanningPoker-v1.1.pdf

Hasso Plattner Institute (2011) LEGO Scrum exercise. Retrieved from http://www.youtube.com/watch?v=H2NXlDoutcY

Hildenbrand T, Meyer J (2012) Intertwining lean and design thinking: software product development from empathy to shipment. In: Maedche A, Botzenhardt A, Neer L (eds) Software for people, management for professional. Springer, Berlin, pp 217–237

Katz R, Allen TJ (1982) Investigating the not invented here (NIH) syndrome: a look at the performance, tenure, and communication patterns of 50 R&D Project Groups. R&D Manag 12(1):7–20

Khan AS, Kajko-Mattsson M (2010) Core handover problems. In: Proceedings of the 11th international conference on product focused software, PROFES '10, ACM. New York, NY, pp 135–139

Larman C, Vodde B (2008) Practices for scaling lean and agile development: large, multisite, and offshore product development with large-scale scrum. Addison-Wesley Professional, Boston, MA

Larman C, Vodde B (2010) Scaling lean and agile development: thinking and organizational tools for large-scale scrum. Addison-Wesley Professional, Boston, MA

Lindberg T, Köppen E, Rauth I, Meinel C (2012) On the perception, adoption and implementation of design thinking in the IT industry. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Berlin, pp 229–240

Patton J (2009) User story mapping. Retrieved from http://www.agileproductdesign.com/presentations/user_story_mapping/

Thinking.designismakingsense. http://thinking.designismakingsense.de. Stakeholder map. Retrieved from http://thinking.designismakingsense.de/service-design/methoden/stakeholder-map

Vetterli C, Uebernickel F, Brenner W (2012) Initialzündung durch embedded design thinking—ein Fallbeispiel aus der Finanzindustrie. Zeitschrift für Organisationsentwicklung 2:22–31

Vetterli C, Uebernickel F, Brenner W, Haeger F, Kowark T, Krueger J, Mueller J, Plattner H, Stortz B, Sikkha V (2013) Jumpstarting scrum with design thinking. University of St.Gallen, St.Gallen

Yip J (2011) It's not just standing up: patterns for daily stand-up meetings. Retrieved from http://www.martinfowler.com/articles/itsNotJustStandingUp.html