

Comparing Paper and Software Tool for Participatory Design from Scratch

Matthias Heintz
University of Leicester
University Road, Leicester,
LE1 7RH, UK
mmh21@leicester.ac.uk

Effie Lai-Chong Law
University of Leicester
University Road, Leicester,
LE1 7RH, UK
elaw@mcs.le.ac.uk

Nervo Verdezoto
University of Leicester
University Road, Leicester,
LE1 7RH, UK
nervo.verdezoto@le.ac.uk

Several paper-based approaches and software tools have been developed to support Participatory Design (PD) in research and practice. However, a formal comparison of paper and software tool is missing in the PD literature. To address this gap we present an empirical study with 28 Informatics students comparing a paper-based approach with our digital tool PDotCapturer. Results show a slight advantage of paper regarding quantitative results and a significant statistical difference in one of the 18 qualities tested (regarding aesthetics [e.g. symmetrical, creative], usability [e.g. user control and freedom, flexibility and efficiency of use], and relevancy [appropriate or not]): designs created with the tool are more 'Pleasant'. Subjectively the participants preferred paper for some of the activities and the tool for others. The results of our comparison show that the tool could be used instead of paper to benefit from some tool-advantages over paper (e.g. digital data gathering and analysis support).

Participatory Design. Paper-based approach. Software tool.

1. INTRODUCTION

Participatory Design (PD, Muller, 1991; Kelly et al., 2006) is a very diverse research area with a variety of settings (e.g. Halskov and Hansen, 2015). From its roots as a Scandinavian movement in the 1960's to democratize the workplace it was adapted in different contexts (Clement and Van den Besselaar, 1993; Greenbaum, 1993; Muller, Wildman, and White, 1993; Gennari and Reddy, 2000; Puri et al., 2004; Porayska-Pomsta et al., 2012).

The goal of participatory design is to empower end-users and other stakeholders by including them in the design process (e.g. Carroll et al., 2000), so that they can actively shape the design according to their needs (Clement and Van den Besselaar, 1993). Diverse approaches have been developed to reach this goal (Sanders, Brandt, and Binder, 2010; Walsh et al., 2013) resulting in many forms of participation. For instance, Sanders, Brandt, and Binder (2010) refer to three forms of participation in which stakeholders can be asked: a) to make things (e.g., by drawing ideas), b) to communicate their needs (e.g., through keeping diaries), and c) to enact to uncover particular needs (e.g., through games, or exercises, etc.).

Paper-based approaches are widely used to enable some of these forms of participation in PD.

Examples of paper usage in participatory design activities are storyboards (Marois et al., 2010), 'Comicboarding' (Moraveji et al., 2007), sticky notes (Carmel, Whitaker, and George, 1993; Druin, 2002; Svanaes and Seland, 2004), 'bags of stuff' (including paper among other materials, Druin et al., 2001) and 'big paper' (Guha et al., 2004).

Although the use of paper to gather PD ideas has a long history and is well established in the PD research area, some researchers have also used software tools in PD activities to support participation. This was done due to several assumed beneficial qualities of software tools over paper (Naghsh and Andy, 2004; Walsh et al., 2012; Heintz et al., 2014; Walsh and Foss 2015). Software tools are often developed to support distributed PD (DPD) activities (Gumm, 2006; Danielsson et al., 2008; Lohmann, Ziegler, and Heim, 2008; Walsh, 2011), where digital data simplifies the exchange between end-users and designers. Nevertheless, software tools could also have advantages over paper in a co-located setting, for example supporting undo and redo, which is more complicated (undo) or even impossible (redo) on paper.

However, the aforementioned software tools did not provide a comparison with the corresponding paper-based counterpart (Naghsh and Andy, 2004; Walsh et al., 2012; Heintz et al., 2014; Walsh and

Foss, 2015). Such a comparison is essential to prove that the software developed for the tool-based method works and that using it results in similar or better designs as compared with using paper. Only then it can be determined and advised in which particular situation each method can be used.

The identified lack of comparison between paper and software/digital tools motivated us to conduct an empirical study which will help to answer the open question, if the quality of ideas gathered with a paper-based and with a tool-based approach are comparable or which differences exist. In this study 28 Informatics students were asked to use the paper-based Layered Elaboration approach and the software tool PDotCapturer to design prototypes and evaluated and compare the two approaches in the process.

2. STATE OF THE ART

One possible outcome of paper-based PD activities in which users/participants are asked to make something (Sanders, Brandt, and Binder, 2010) are prototypes, which are "... physical manifestations of ideas or concepts" (Sanders and Stappers, 2014, p. 9). Besides paper-based approaches (see Introduction), software tools have been developed, to facilitate the creation and redesign of prototypes in PD sessions.

Naghsh and Andy (2004) created the software tool GABBEH to provide a digital version of paper-prototyping. It enabled end-users to annotate website designs and in this way express their re-design ideas and feedback. One of the identified benefits of using a software tool instead of paper was that it could support the evaluation of dynamic behaviour in a prototype. However, it also had some shortcomings for general usage, mainly relying on the DENIM tool (Newman et al., 2003) and specific hardware requirements (graphics tablet as input device) to work. Although developed as a digital version of paper prototyping, GABBEH was not compared with a paper-based approach.

Walsh et al. (2012) designed and developed the software tool DisCo to support intergenerational PD activities. They introduced a tool to help keeping track of PD ideas assigned to different versions of a constantly changing prototype and to support PD in a distributed setting. In Walsh and Foss (2015) a further improved and enhanced version of the system is presented (e.g. supporting audio input). The DisCo online tool is a digital representation of the paper-based Layered Elaboration approach (Walsh et al., 2010), but the two methods to gather PD input from end-users have not been compared.

In the Layered Elaboration approach small groups of end-users and designers work together on an

initial design on paper. The example in Figure 1 shows design ideas (Lim, Stolterman, and Tenenberg, 2008) for a Quiz app. Once the initial design is finished it is then exchanged with one from another group participating in the same PD event. Instead of directly annotating the design, an acetate is put on top of the paper. Annotations and re-design ideas are then drawn and written on the acetate (Figure 2). This has the benefit that the second group does not have the impression to 'destroy' the initial design of the first group.

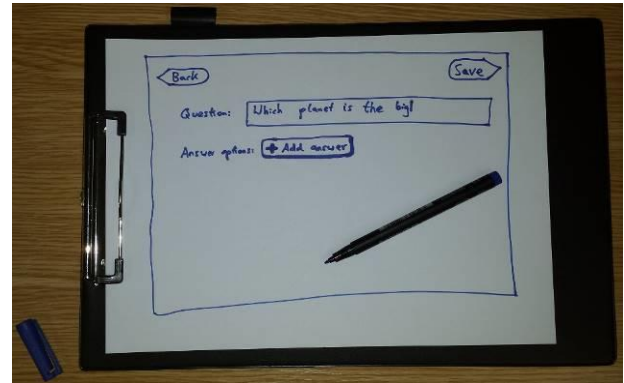


Figure 1: Layered Elaboration: Initial design on paper.

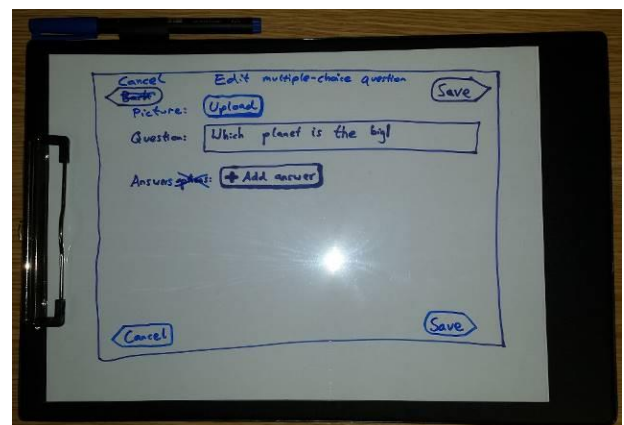


Figure 2: Layered Elaboration: Annotations and re-design suggestions on an acetate on top of the initial design on paper.

Heintz et al. (2014) present the online tool Pdot to address identified shortcomings of paper-based approaches and existing tools in participatory design activities, especially in a distributed setting. However, they only presented an early prototype of their tool, together with preliminary evaluation results. A comparison of paper and tool is still missing.

An example of a comparison of the performance between paper- and tool-based approaches to gather end-user input, in an area closely related to PD, is the evaluation of CanonSketch (Campos and Nunes, 2007), a design tool for User-Centred Design (UCD) activities. Contrary to the PD philosophy this tool does not include actual end-users in the design process. For CanonSketch it

was shown that the tool outperformed paper, for the most of the tested qualities even significantly (Campos and Nunes, 2007).

The aforementioned restrictions (e.g. hardware and software requirements for GABBEH) and unavailability (e.g. DisCo) of existing tools motivated us to develop our own tool (PDotCapturer, based on Pdot) to be able to make a comparison with a paper-based approach.

We selected the Layered Elaboration (Walsh et al., 2010) paper-based PD approach for the comparison due to its versatility to accommodate the needs of a wide range of end-users (from children to adults). Furthermore, DisCo (Walsh et al., 2012) has shown that indeed Layered Elaboration could be transformed into a software tool and used for PD activities. Thus, we implemented our own software tool PDotCapturer (Figure 3) taking into account the affordances provided by the Layered Elaboration approach. Figure 6 illustrates the complete Layered Elaboration process from our empirical study to explore the potential impact of both approaches on the outcome of PD activities. PDotCapturer has been developed in iterative cycles of implementation and evaluation, but the details of implementation are out of the scope of this paper.

3. EMPIRICAL STUDY

3.1. Description of PDotCapturer

After logging in the tool presents a plain white page as the starting point (part of which is shown on the right hand side of Figure 3). The end-user can then click anywhere on the page to start drawing by dragging the mouse (Figure 3 shows that the end-user has drawn a 'start' button as part of the initial design). To specify different design elements, users can create numbered markers positioned next to drawings. Additional information (e.g. longer text which would be complicated to write with the mouse) can be specified in a text input box. In the example presented in Figure 3 the end-user has described the functionality of the 'start' button: "Pressing the start button starts the creation of a new Quiz".

To enable end-users to inspect and annotate the designs of other participants, the PDotCapturer tool offers the feature to show the initial design of others (see second segment from the top in Figure 4). In this case the design of another group of end-users participating in the PD activity is displayed in the design area on the right. End-users can then use the functionality shown in Figure 3 to annotate the design of others instead of working on their own, initial design. The tool creates a 'virtual acetate' for this purpose on which the drawings can be done and the textual comments be located.

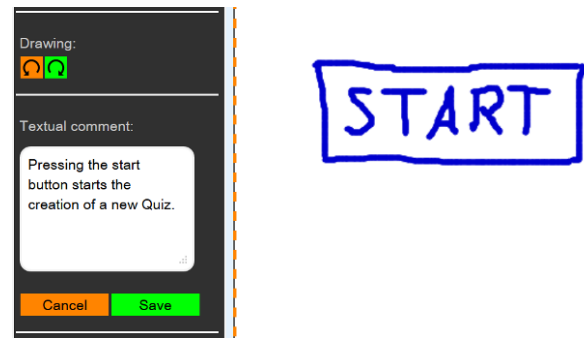


Figure 3: Screenshot of PDotCapturer to create an initial design: Drawing area on the right hand side. Features of the tool on the left hand side (undo and redo for the drawing on top, input for textual comment on the bottom).

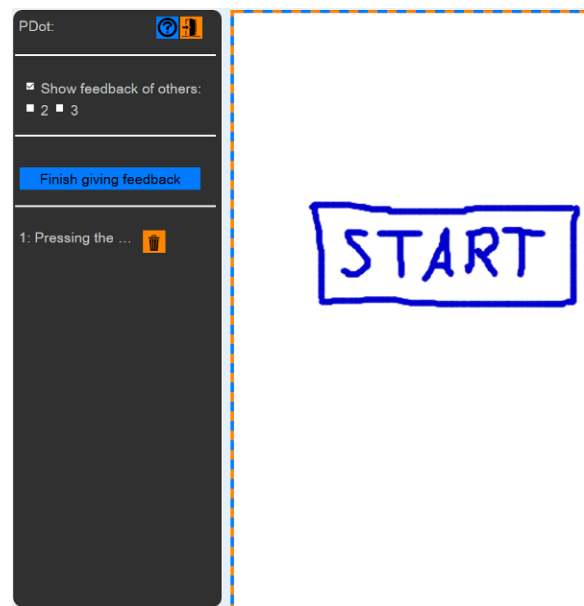


Figure 4: Screenshot of PDotCapturer showing the tool when inspecting designs or feedback of other end-users.

PDotCapturer is implemented using HTML and JavaScript, which means it can be run in any major browser without the need for a plugin. All drawings and annotations created by end-users are stored online in a MySQL database.

3.2 Participants and Procedures

28 Informatics students used Layered Elaboration and PDotCapturer in a cross-comparison study to design: 1) a sleeping personal informatics app home screen, and 2) a customizable dashboard, for their university homepage. These applications were selected because they fit the curriculum, the students have learned about personal informatics applications, and because of the familiarity of the target group with the university homepage. Additionally for both applications it was reasonable to design a single screen. This limitation was necessary for the study due to time constraints regarding the amount of time that could be allocated for creating and annotating each design.

The two-hour session started with a presentation explaining Participatory Design and where it would typically be applied in the software design life-cycle when designing an application from scratch. The Layered Elaboration method and PDotCapturer were described as two options for PD activities. Then the first task, to design a sleeping personal informatics app home screen (showing e.g. sleep last night/this week and alarm time) was presented. The participants were randomly divided into two groups of 14 each (Figure 5), starting with the Layered Elaboration approach (green box on top in Figure 5) or PDotCapturer tool respectively (yellow box on the bottom in Figure 5). The two subgroups were further divided into six teams: four teams of two and two teams of three participants for each approach (grey team boxes in Figure 5) to offer the participants the opportunity to perform the design in small groups, as proposed in the Layered Elaboration approach description. Six teams per approach were created to enable three teams each to discuss and exchange their designs.



Figure 5: Division of participants.

The participants then had 15 minutes time to come up with an initial design, and specify it, either on paper or on a blank white slate in PDotCapturer. Then they explained their design to the other two groups and exchanged designs. They then had 10 minutes time to annotate the initial design of another team, either on acetates or on a virtual layer in PDotCapturer. The process is displayed in Figure 6 and Figure 7. An example result for a sleeping informatics app home screen designed with the Layered Elaboration approach can be seen in Figure 8. An example result for a sleeping informatics app home screen designed with PDotCapturer can be seen in Figure 9.

To avoid order effects the participants were then shown an animated video on ACTA (<https://www.youtube.com/watch?v=gIR5Kf6GvpY>, last accessed 24/03/2017). This video was unrelated to either one of the two design tasks. Together with a short discussion on it afterwards, it

was used to distract the participants to mitigate effects of the first design task on the second.

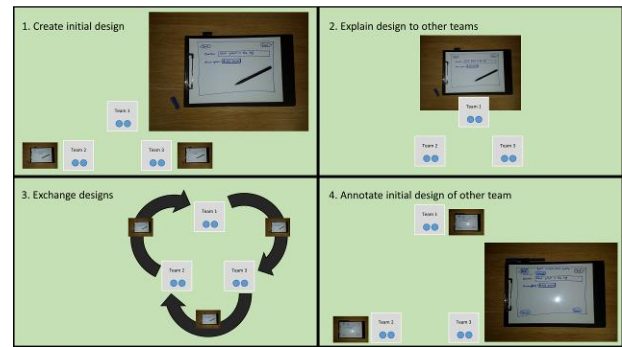


Figure 6: Layered Elaboration approach as performed with paper and acetates (the same process was followed by Team 4, Team 5 and Team 6).

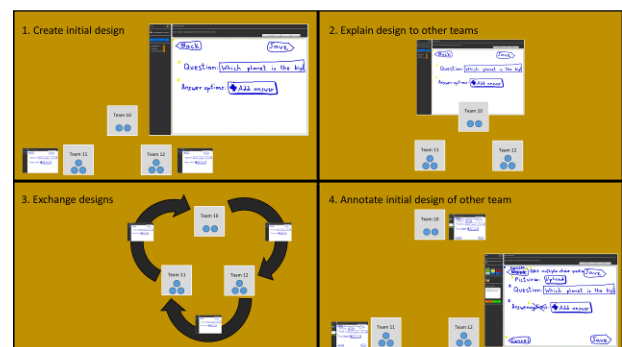


Figure 7: Approach as performed with PDotCapturer (the same process was followed by Team 7, Team 8 and Team 9).

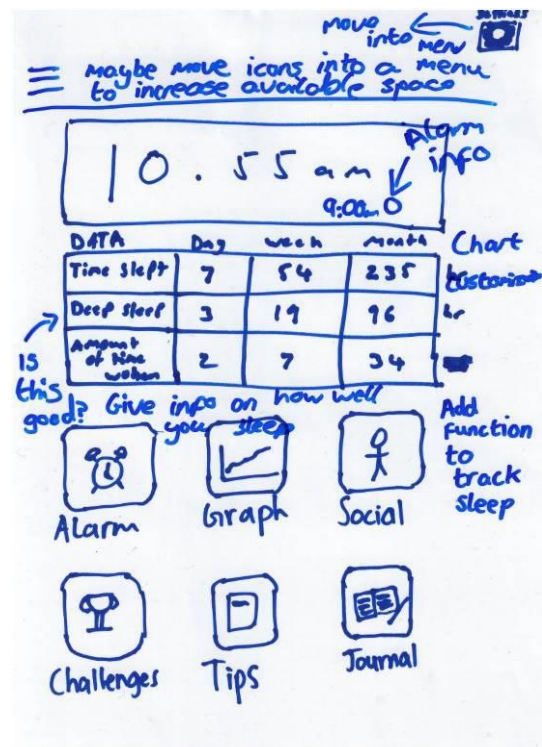


Figure 8: Example of a sleeping informatics app home screen designed with the Layered Elaboration approach.

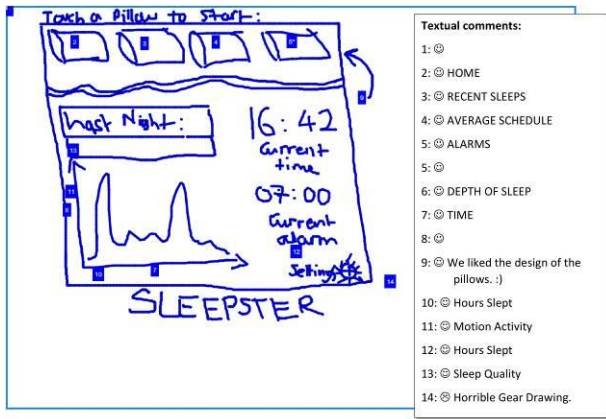


Figure 9: Example of a sleeping informatics app home screen designed with PDotCapturer.

Then the participants swapped the PD method, the ones who used the Layered Elaboration approach before now used PDotCapturer and vice versa. The task this time was to design a university customizable dashboard (showing e.g. announcements, appointments, timetable and an overview of emails). The following procedure of creating an initial design, explaining, exchanging, and annotating it were the same as for the sleeping personal informatics app. In total eleven designs created with the Layered Elaboration approach were handed in by the participants. An example for a university dashboard created with the Layered Elaboration approach can be seen in Figure 10. In total twelve designs created with PDotCapturer were handed in by the participants. An example for a university customizable dashboard created with PDotCapturer can be seen in Figure 11.

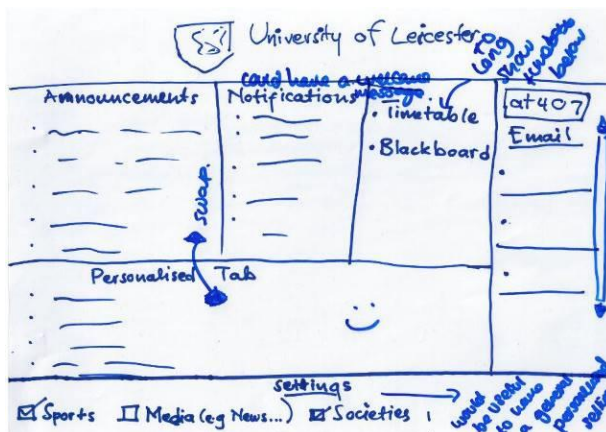


Figure 10: Example of a university customizable dashboard designed with the Layered Elaboration approach.

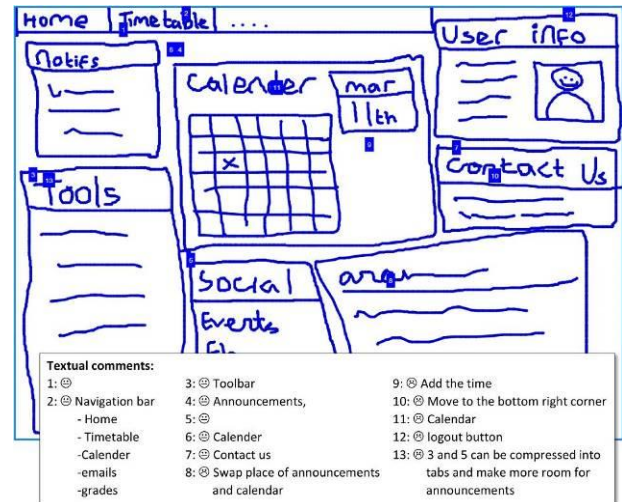


Figure 11: Example of a university customizable dashboard designed with PDotCapturer.

After using both methods the participants filled in a questionnaire. It consisted of a section capturing demographic data (e.g. age and gender) and a section comparing the paper-based approach and PDotCapturer, inspired by the System Usability Scale (SUS; Brooke, 1996), and asking for preference of one or the other regarding different PD idea expression tasks (see Figure 12 for an example statement and Table 8 for a list of the statements) followed by a text field to elaborate on why and for what the participants preferred which method.



Figure 12: Example statement from the questionnaire comparing Layered Elaboration (Paper) with PDotCapturer (PDot).

3.3. Comparison of designs

To compare the designs created with the Layered Elaboration approach and with PDotCapturer each design was rated regarding aesthetics, usability, and relevancy by two HCI specialists. They first did the rating individually and then met to compare and discuss in case of discrepancies.

3.3.1. Aesthetics

To rate the aesthetics of the designs, rating scales for classical and expressive aesthetics were used, developed by Porat and Tractinsky (2012). For classical aesthetics the qualities 'clean', 'pleasant', 'symmetrical' and 'aesthetic' were rated on a 7-point scale ranging from strongly disagree (1) to strongly agree (7). For expressive aesthetics the qualities 'original', 'sophisticated', 'spectacular', and 'creative' were rated using the same 7-point scale.

To make sure the qualities of the designs were not influenced by the two different topics of the design

task the ratings were tested for independence of the design task. A Shapiro-Wilk test for normality showed that the data were not normally distributed for the different rating categories and tasks. Therefore a Mann-Whitney U test was performed. Results showed that no significant differences in aesthetic ratings between the two design tasks were detected (Table 1). This suggested that the aesthetics ratings were independent of the prototype designed (sleeping informatics app home screen or university customizable dashboard).

Table 1: Mann-Whitney U test results for differences in aesthetics qualities between the two design tasks.

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

3.3.2. Usability

For the usability rating Nielsen's 10 Usability Heuristics were applied to the designs (Nielsen, 1994). For each heuristic the raters specified if it was violated or not. In case the heuristic could not be checked based on the prototype the raters could specify 'not applicable' for this heuristic.

Table 2: χ^2 values showing if the usability rating is independent from the design task or not.

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

To test for independence of the usability ratings from the design task, a Chi square test was performed. This test was selected as there are two categorical variables to be compared: heuristic violation (yes, no, not applicable) and task (university customizable dashboard or sleeping informatics app home screen). Three of the heuristic ratings could not be analysed, as they were rated as 'not applicable' for all designs: "Error prevention", "Flexibility and efficiency of use", and "Help users recognize, diagnose, and recover from errors" (see section 3.4.2. below for details). For the other heuristics no significant relation between rating and design task performed was found (Table 2).

3.3.3. Relevancy

To rate the relevancy of the design elements created by the participants the experimenter identified the different elements in each design and indicated them by an orange line separating them from the rest of the design (Figure 13). For each of these elements the raters then specified if it was relevant to the design task (i.e. one would expect to find such an element in the prototype to be designed) or not. An example relevancy rating for one of the designs can be seen on the right hand side in Figure 13.

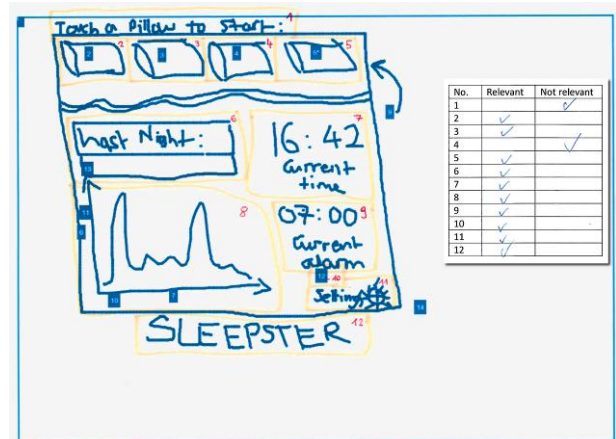


Figure 13: Example of highlighting the twelve design elements identified in the design (left) and relevancy rating (right).

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

3.4. Results

3.4.1. Aesthetics

Table 3 shows the results of the Aesthetic ratings. To compare the two approaches (Layered

Elaboration and PDotCapturer), the same inferential statistics as described above for the design tasks were applied, only this time for the approach used (paper-based and tool-based). Only one item showed a significant statistical difference: "Pleasant" with $U=34.00$, $p<0.05$, Layered Elaboration mean rank=9.09, PDotCapturer mean rank=14.67. Thus the designs created with PDotCapturer (mean=4.42) were rated as more pleasant than the designs created with Layered Elaboration (mean=3.64).

Table 3: Results of the Aesthetic ratings for Layered Elaboration and PDotCapturer (mean and standard deviation (SD) on a 7-point scale).

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

However, as no other results showed a significant difference between the approaches (Table 4) it is suggested that Layered Elaboration and PDotCapturer could be used interchangeably to collect PD ideas. Using a paper- or tool-based approach only slightly (in one out of eight aesthetics aspects) influenced the design outcome regarding aesthetics.

Table 4: Mann-Whitney U test results for differences in aesthetics qualities between the two approaches (LE = Layered Elaboration; PDot = PDotCapturer). * $p<.05$.

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

Even though only one item is statistically significant, two observations can be made when comparing the results for Layered Elaboration and PDotCapturer in Table 3. On the one hand, the classical aesthetics qualities (all means above 3) have been rated higher than the expressive aesthetics qualities (all means below 3) for all designs. On the other hand, PDotCapturer has been rated higher than Layered Elaboration in the

classical aesthetic qualities ('clean', 'pleasant', 'symmetrical', 'aesthetic') while Layered Elaboration has mostly been rated higher than the tool in the expressive aesthetic qualities ([original', being the exception] 'sophisticated', 'spectacular', 'creative').

3.4.2. Usability

None of the usability heuristic rating results (Table 5) showed a significant difference between Layered Elaboration and PDotCapturer in the Chi square tests performed (Table 6).

Table 5: Results of the Usability ratings for Layered Elaboration and PDotCapturer (NV = not violated, V = violated, N/A = not applicable).

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

Table 6: χ^2 values showing if the usability rating is independent from the approach or not.

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

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

3.4.3. Relevancy

The results of the relevancy ratings is shown in Table 7. With the paper-based approach 128 design elements were identified, 45 of which were rated as not relevant for the respective kind of prototype. With PDotCapturer 111 design elements were created by the participants, 27 of which were rated as not relevant. Therefore Layered Elaboration resulted in an average of 7.55 relevant design elements and PDotCapturer in an average of 7 relevant design elements per design. As for the test for independence of task and relevance of screen elements, a Chi square test was performed. The results showed that the amount of relevant to irrelevant design elements was not significantly different between Layered Elaboration and PDotCapturer ($\chi^2(1, n=239)=3.314, p>.05$ (ns)).

Table 7: Results of the Relevancy ratings for Layered Elaboration and PDotCapturer.

Relevancy	Layered Elaboration	PDotCapturer
Not relevant	45	27
Relevant	83	84

3.3.4. Questionnaire

For the data analysis of the statements comparing Layered Elaboration and PDotCapturer (see Figure 12 for an example) the five points on the scale between Paper and PDot were given the values 1 (for the point next to Paper) to 5 (for the point next to PDot). The point in the middle between Paper and PDot was thus assigned the value 3. Therefore

all values of less than three indicated favour of Layered Elaboration (the further away from three, and thus closer to 1, the more), where all values larger than three indicated a preference of PDotCapturer (the bigger, and thus closer to 5, the more).

Table 8: Replies regarding the statements comparing the paper-based and tool-based method on a scale from 1=Layered Elaboration to 5=PDotCapturer.

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

When analysing the comparison of Layered Elaboration and PDotCapturer by the participants, presented in Table 8, it could be seen that PDotCapturer was rated slightly better than Layered Elaboration for three of the statements (mean >3). The first one was providing textual feedback. The participants preferred typing on a keyboard at the PC to handwriting on paper and acetates. This is also supported by the comments specified in the questionnaire, e.g. "Faster and easier to type on pdot and it's condensed into sticky note." The other two statements were the ones based on the PD tasks PDotCapturer was initially designed for (check out an existing design and annotate it). Again, the results from the data analysis were also supported through the comments provided in the questionnaire regarding why and for what which method was preferred: "Evaluating was easier with PDot with the contextual sticky notes." and "Really quick when giving feedback and like, dislike, neutral emotions are useful". Whereas the tool was preferred for textual feedback, non-verbal feedback such as drawings and sketches was clearly preferred to be given using the paper-based approach. This could be explained by the observation that drawing with a mouse on the screen is not common. This was also expressed by the participants in the freeform text fields regarding why they preferred one of the approaches to the other: "easier to draw by pen/pencil" or "Hard to draw on computer". All other ratings were between 2 and 3 and therefore in

(slight) favour of Layered Elaboration over PDotCapturer.

4. DISCUSSION

4.1. Affordances of paper and tool

A concept that can be used to explain paper usage and its comparability with tools is that of 'affordances'. The theory behind affordances, developed by Gibson (1979), implies that people can perceive functionality and possible actions based on the physical properties of objects. "The physical properties of paper ... afford many different human actions, such as ... manipulating ... and in combination with a marking tool, writing on." (Sellen and Harper, 2003, page 17). Although software tools, as digital technologies, have different affordances as paper (Sellen and Harper, 2003, page 18), the results of our comparison showed that the designs created with both approaches were similar. This could be explained with the PC literacy of the participants, which, together with appropriate tool design, might have mimicked the affordance of paper to write on it in the tool. The empty white space on the computer screen, together with a mouse cursor looking like a pen, might have provided a similar affordance to a real pen and empty piece of paper.

However, the different affordances of paper and tool could explain the differences in user perception suggested by the questionnaire results presented. A PC has different affordances from paper: "... because it dynamically displays information, it affords the viewing of moving images" (Sellen and Harper, 2003, page 18), or in our case, interactive prototypes. Although the prototype designs created by drawing in PDotCapturer were not interactive, the presentation on a PC screen might have given the impression of an interactive application more than the drawing on paper and acetates.

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

4.2. Tool- usage

For end-users a tool-based PD sessions could offer several advantages over a paper-based one (as expressed and confirmed by some of the responses in the free text fields in the questionnaire comparing Layered Elaboration and PDotCapturer): When annotating the initial designs of other end-users participating in the same event, there is no issue with illegible handwriting, which can pose a

problem on paper, as most of the text is typed. The understandability of initial designs for other end-users annotating the design could be improved through elaborate textual descriptions and explanations. While these could clutter the interface design on paper and could cause space issues, as they are written inside the design space, they were 'hidden' in small markers not interfering with the prototype design with the tool. Typing was perceived as advantageous over writing by hand by some participants. These benefits of tools from the end-users perspective should motivate more tool-usage in PD.

From the subjective ratings of the participants, given in the questionnaire, it could be seen that they preferred the paper-based approach for the initial design, but PDotCapturer for the following round of annotation activities. To use the tool for this purpose the initial designs could be digitalized, for example by taking a picture using a digital camera or mobile phone. They could then be presented in PDotCapturer for the following rounds of PD activities to further refine the design. PDotCapturer could be enhanced to facilitate and ease this process, e.g. a mobile app could be developed to take a picture of a design on paper and integrate it into the tool. A scenario for this in practice could be applying the different approaches over several workshops: an initial design workshop with paper followed by one or more refinement workshops using PDotCapturer. This would allow the end-users to work with their preferred method for each of the PD design and refinement activities. Like in document management, with paper and digital documents, where "introduction of a new technology can stimulate a synergy between old and new" (Liu and Stork, 2000, page 97), PDotCapturer as a new tool could induce new ways to work with and handle paper.

Besides paper-based methods, there are plenty of other, non-paper-based, non-digital methods in PD, e.g. using foam, clay, or Lego bricks to form a three dimensional prototype (Sanders, Brandt, and Binder, 2010). PDotCapturer could be compared with these approaches as well (e.g. by foreshortening drawings to make them look three-dimensional). However it might be more promising to develop dedicated tools (for example using virtual reality for a digital representation of 3D objects) and conduct similar comparisons between those approaches and tools.

4.3. Tool- support for designers

As well as offering advantages for participants (e.g. typing instead of handwriting) tools could also provide advantages for designers in PD activities. Besides supporting them in distributed PD settings a tool also "opens up new research opportunities with the field of Natural Language Processing and

Machine Learning to develop visualization techniques and automated organization of the ideas" (Walsh and Foss, 2015, page 106). Useful innovative visualisations and automation of tasks following the idea gathering could lead to a better utilisation of PD activities, not through ensuring that end-users are supported in specifying their ideas, but by ensuring designers are supported in considering and appropriately representing these ideas (Read et al., 2016).

5. LIMITATIONS

The main threat of validity regarding the findings presented in this paper is that they are based on a single evaluation. This paper presents a short quantitative analysis comparing paper and tool but the results are indicative rather than definitive. Larger scale studies, including quantitative and qualitative methods, are needed to substantiate our findings.

Designing only single-page applications due to time constraints for the study is a restriction which influenced the results (e.g. some of the usability heuristics could not be rated because the respective graphical user interface elements or dialogs were not part of the designs).

A possible limitation of this research might be that most of the participants in the performed study had a good PC literacy. Results of comparing paper-based PD approaches and software tools might differ if non-PC-experts are participating.

The study presented here was performed with participants that had not used PDotCapturer before. Given the learning of interface and interaction associated with the tool, studies with experienced or even expert PDotCapturer users might have yielded different results.

6. CONCLUSION AND FUTURE WORK

The comparison of Layered Elaboration and PDotCapturer, when designing an application from scratch, showed similarities and differences of the two approaches. However, PDotCapturer could be used instead of the paper-based approach, to enable benefits of tool-usage, as it facilitated the same qualities in the resulting designs.

The presented study comparing paper-based and tool-based PD approaches only found one significant difference: The PDotCapturer tool outperformed Layered Elaboration in one of the eight Aesthetic ratings applied (Pleasant). For the Usability and Relevancy ratings of the designs no significant differences were found. End-users preferred the paper-based Layered Elaboration approach over the PDotCapturer tool for tasks related to creating an initial design but the tool-

over the paper-based approach to explore and annotate the designs of others and to create textual comments.

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

Evaluating paper and tool for a single-page application is a good starting point and gave interesting first insights regarding the comparability of paper and tool in PD. However, more complex applications should be designed with both approaches, and the results compared, to get additional findings from different, more sophisticated design tasks. Larger scale studies, including quantitative and qualitative methods, should be conducted to substantiate our initial findings.

Although PDotCapturer has only been used in a co-located setting for this study (to minimize the number of variables that were not controlled), its design is feasible for using it in a distributed setting. As this is an area where tool-usage could have additional advantages over paper (Naghsh and Andy, 2004; Walsh et al., 2012; Heintz et al., 2014; Walsh and Foss 2015), it would be interesting to compare paper-based approaches with PDotCapturer in a distributed PD setting.

To allow for a general applicability of PDotCapturer in different usage scenarios and environments the technical requirements have been restricted to a minimum (any browser, Internet access, mouse and keyboard as input devices). By loosening these restrictions more technically advanced options could be explored. By requiring touch- or stylus-supporting devices PDotCapturer could offer an input modality closer to the paper-based experience than drawing with a mouse. It would thus be interesting to compare PDotCapturer on a touch-screen device with a paper-based approach.

7. REFERENCES

- Brooke, J. (1996) Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189 (194). 4-7.
- Campos, P. and Nunes, N. (2007) Towards useful and usable interaction design tools: Canonsketch. *Interacting with Computers*, 19 (5-6). 597-613.

- Carmel, E., Whitaker, R.D., and George, J.F. (1993) PD and joint application design: a transatlantic comparison. *Communications of the ACM*, 36 (6). 40-48.
- Carroll, J.M., Chin, G., Rosson, M.B., and Neale, D.C. (2000) The development of cooperation: five years of participatory design in the virtual school. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques, DIS '00*. Brooklyn, 17-19 August. New York: ACM. 239-251.
- Clement, A. and Van den Besselaar, P. (1993) A retrospective look at PD projects. *Communications of the ACM*, 36 (6). 29-37.
- Danielsson, K., Naghsh, A.M., Gumm, D., and Warr, A. (2008) Distributed participatory design. In *CHI'08 extended abstracts on Human factors in computing systems*. Florence, 05-10 April 2008. New York: ACM. 3953-3956.
- Druin, A. (2002) The role of children in the design of new technology. *Behaviour & Information Technology*, 21 (1). 1-25.
- Druin, A., Bederson, B.B., Hourcade, J.P., Sherman, L., Reville, G., Platner, M., and Weng, S. (2001) Designing a digital library for young children. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '01*. Roanoke, 24-28 June 2001. New York: ACM. 398-405.
- Gennari, J.H. and Reddy, M. (2000) Participatory design and an eligibility screening tool. In *Proceedings of the AMIA symposium*. Los Angeles, 04-08 November 2000. American Medical Informatics Association. 290-294.
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin.
- Greenbaum, J. (1993) A design of one's own: towards participatory design in the united states. *Participatory design: Principles and practices*, 1993. 27-37.
- Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., and Farber, A. (2004) Mixing ideas: a new technique for working with young children as design partners. In *Proceedings of the 2004 conference on Interaction design and children: building a community*. Baltimore, 01-03 June 2004. New York: ACM. 35-42.
- Gumm, D.C. (2006) Distributed participatory design: An inherent paradoxon? In *Proceedings of IRIS29*.
- Halskov, K. and Hansen, N. B. (2015) The diversity of participatory design research practice at PDC 2002-2012. *International Journal of Human-Computer Studies*, 74. 81-92.
- Heintz, M., Law, E.L.-C., Govaerts, S., Holzer, A., and Gillet, D. (2014) Pdot: participatory design online tool. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. Toronto, 26 April - 01 May 2014. New York: ACM. 2581-2586.
- Kelly, S.R., Mazzone, E., Horton, M., and Read, J.C. (2006) Bluebells: a design method for child-centred product development. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. Oslo, 14-18 October 2006. New York: ACM. 361-368.
- Lim, Y-K., Stolterman, E., and Tenenber, J. (2008) The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15 (2). 7:1-7:27.
- Liu, Z. and Stork, D.G. (2000) Is paperless really more?. *Communications of the ACM*, 43 (11). 94-97.
- Lohmann, S., Ziegler, J., and Heim, P. (2008) Involving end users in distributed requirements engineering. In *Engineering Interactive Systems*. Springer. 221-228.
- Marois, L., Viallet, J.E., Poirier, F., and Chauvin, C. (2010) Experimenting introductory tools for innovation and participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*. Sydney, 19 November - 03 December 2010. New York: ACM. 259-262.
- Moraveji, N., Li, J., Ding, J., O'Kelley, P., and Woolf, S. (2007) Comicboarding: Using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, 30 April - 03 May 2007. New York: ACM. 1371-1374.
- Muller, M.J. (1991) Pictive-an exploration in participatory design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New Orleans, 27 April - 02 May 1991. New York: ACM. 225-231.
- Muller, M.J., Wildman, D.M., and White, E.A. (1993) Taxonomy of PD practices: A brief practitioner's guide. *Communications of the ACM*, 36 (6). 26-28.
- Naghsh, A.M. and Andy, D. (2004) Gabbeh: a tool to support collaboration in electronic paper prototyping. In *CSCW 2004 the ACM Conference on Computer Supported Cooperative Work*. Chicago, 06-10 November 2004. New York: ACM.
- Newman, M.W., Lin, J., Hong, J.I., and Landay, J.A. (2003) DENIM: An informal web site design tool inspired by observations of practice. *Human-Computer Interaction*, 18 (3). 259-324.

- Nielsen, J. (1994) Heuristic evaluation. *Usability inspection methods*, 17 (1). 25-62.
- Porat, T. and Tractinsky, N. (2012) It's a pleasure buying here: The effects of webstore design on consumers' emotions and attitudes. *Human-Computer Interaction*, 27 (3). 235-276.
- Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M.E., Alcorn, A., Wass, S., Bernadini, S., Avramides, K., Keay-Bright, W., Chen, J., Waller, A., Guldborg, K., Good, J., and Lemon, O. (2012) Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, 16 (2). 117-127.
- Puri, S.K., Byrne, E., Nhampossa, J.L., and Quraishi, Z.B. (2004) Contextuality of participation in is design: A developing country perspective. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1, PDC 04*. Toronto, 27-31 July 2004. New York: ACM. 42-52.
- Read, J.C., Fitton, D., Sim, G., and Horton, M. (2016) How ideas make it through to designs: Process and practice. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction, NordiCHI '16*. Gothenburg, 23-27 October 2016. New York: ACM. 16:1-16:10.
- Sanders, E.B.-N., Brandt, E., and Binder, T. (2010) A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*. Sydney, 19 November - 03 December 2010. New York: ACM. 195-198.
- Sanders, E.B.-N. and Stappers, P.J. (2014) Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign: International Journal of CoCreation in Design and the Arts*, 10 (1). 5-14.
- Sellen, A.J. and Harper, R.H. (2003) *The myth of the paperless office*. MIT press.
- Svanaes, D. and Seland, G. (2004) Putting the users center stage: Role playing and low-prototyping enable end users to design mobile systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '04*, Vienna, 24-29 April 2004. New York: ACM. 479-486.
- Walsh, G. (2011) Distributed participatory design. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. Vancouver, 07-12 May 2011. New York: ACM. 1061-1064.
- Walsh, G., Druin, A., Guha, M.L., Bonsignore, E., Foss, E., Yip, J. C., Golub, E., Clegg, T., Brown, Q., Brewer, R., et al. (2012) Disco: a co-design online tool for asynchronous distributed child and adult design partners. In *Proceedings of the 11th International Conference on Interaction Design and Children*. Bremen, 12-15 June 2012. New York: ACM. 11-19.
- Walsh, G., Druin, A., Guha, M.L., Foss, E., Golub, E., Hatley, L., Bonsignore, E., and Franckel, S. (2010) Layered elaboration: a new technique for co-design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Atlanta, 10-15 April 2010. New York: ACM. 1237-1240.
- Walsh, G. and Foss, E. (2015) A case for intergenerational distributed co-design: The online kidsteam example. In *Proceedings of the 14th International Conference on Interaction Design and Children*. Medford, 21-25 June 2015. New York: ACM. 99-108.
- Walsh, G., Foss, E., Yip, J., and Druin, A. (2013) FACIT PD: a framework for analysis and creation of intergenerational techniques for participatory design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris, 27 April - 02 May 2013. New York: ACM. 2893-2902.