# Getting Quizzical About Physical: Observing Experiences with a Tangible Questionnaire

**Sarah Gallacher[1], Connie Golsteijn[1], Lorna Wall[1], Lisa Koeman[1], Sami Andberg[2], Licia Capra[1], Yvonne Rogers[1]**

[1]University College London, London WC1E 6BT, UK
[2]University of Helsinki, Helsinki FI-00014, Finland
{s.gallacher, c.golsteijn, l.wall, lisa.koeman.12, l.capra, y.rogers}@ucl.ac.uk,
sami.andberg@helsinki.fi

## ABSTRACT

Organizers regularly want to understand the experiences of event goers and typically use survey methods, with researchers and clipboards. However, gathering opinions in such ways is difficult to do without disrupting the event goers' experience. In place of clipboard surveys, we developed a quite different form of *tangible questionnaire,* called *VoxBox*, which uses physical interactions to transform feedback giving into a playful and engaging experience that fits much more with the event itself. Here we question if such a device can successfully draw a diverse representation of event attendees to voice relevant opinions during the event. We describe an observational study of VoxBox based on two real-world deployments, and present findings on (1) the experiences VoxBox provides to facilitators and users; and (2) its capabilities as a means for opinion gathering. We conclude by discussing lessons learned, design implications, and the wide potential for tangible questionnaires in other application areas.

### Author Keywords

Public opinion; questionnaires; crowd engagement; playful; in the wild; tangible interaction; Internet of Things

### ACM Classification Keywords

H.5: Information interfaces and presentation (e.g., HCI): H.5.2. User Interfaces; H.5.m. Miscellaneous.

## INTRODUCTION

Gathering public opinion and feedback is a crucial part of decision making and planning for many organizations. Not only does it enable them to change for the better, it also gives a voice to individuals within the community to affect that change. However, gathering public opinion can be challenging, especially when the targeted population is transient, such as at an event. It is desirable to capture opinions in situ while respondents are still in the moment, experiencing the event. Equally, it is desirable to do so without disrupting visitors' event experience, which can happen when they are asked to give their opinion. Current solutions such as online or paper-based questionnaires struggle to do both. The former is typically completed after the event, based on the respondent's memory, and often results in low response rates [9]. The latter involves what can be a disruptive or awkward exchange with a researcher during the event, which many people shy away from [11].

One solution is to design and deploy the opinion-giving activity in such a way that it becomes a part of the event experience itself, as if it was another installation or stall to walk up to and experience. That is, something that people want to approach and engage with on their own terms, resulting in an enjoyable experience that is not at odds with the atmosphere or happenings of the wider event. To achieve this, the VoxBox 'tangible questionnaire' was developed to be attractive, intriguing and playful when approaching it, as shown in Figure 1. VoxBox uses physical interaction to draw people's attention, offer clear and easy interaction mechanisms [4], and provide an engaging experience of giving feedback and answering survey questions. It is designed to operate as an autonomous device in situ at an event, to mitigate known challenges of face-to-face questioning such as awkward encounters, unintentional selection bias [11] and positive, but not truthful, responses [1], while still gathering answers in the moment.

VoxBox was also implemented to be fully IoT (Internet of Things) enabled to support real-time data processing, visualization and device customization. These technical capabilities supported real-time feedback at the event but were abstracted and hidden from the respondents.

Here we present the details and findings of an observational study carried out during two separate field trials of VoxBox. The study provided insights into the *experiences* of using VoxBox and its *capabilities* as a questionnaire system. Regarding the former, this includes looking at how it blended into the overall event experience and whether there were privacy or social embarrassment issues that discouraged people from approaching or interacting with it. Regarding the latter, this includes looking at whether VoxBox can live up to what is expected from a survey


ACM 978-1-4503-3574-4/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2750858.2807529

system, such as reaching a diverse group of respondents and gathering meaningful responses to closed and open questions. Below we describe the technology behind VoxBox, the setup of the observational study and then the findings from its deployment. Based on these, we discuss key lessons learned, followed by design implications for researchers and future developers of such systems. Finally, we discuss the potential of physical questionnaires in other settings and across different application areas.

## BACKGROUND

Technology plays a key role in opinion gathering methods. Online survey platforms are widely used and mobile devices such as tablets are increasingly used in situ instead of paper surveys and clipboards. Additionally, systems have been proposed that can collect feedback and opinions about urban planning [18, 24] or local issues [6, 22] through interacting with technologies such as large screens, simple voting boxes, or tweets and text messages.

### Screen-based Systems

One approach to gathering public opinions has been to use large screens in public spaces. Schroeter et al. [18] developed an application for public displays that facilitated a feedback platform for urban planning, specifically aimed at those who otherwise would not be heard. Whittle et al. [24] installed a system, VoiceYourView, in a library, which showed people's views on public screens throughout the space. While many people freely gave their opinions in both settings, some felt uncomfortable and self-conscious doing so. Hosio et al. [6] presented Ubinion, a service that used public displays to let young people give their opinions on local issues. Feedback was given through photos taken with integrated webcams, alongside annotated thought bubbles or protest signs. This system was found to be successful in getting opinions from young people, who are usually difficult to reach and engage. The Opinionizer [2] was a large display that represented people's opinions as speech bubbles and was used at a book launch party and a welcome event for students. Valkanova et al. [22] developed MyPosition, a large projected display situated in public spaces that asked people to vote on local issues.

These systems have used a variety of mechanisms to interact with the screens, including: texts and tweets [18], social media such as Facebook [6], keyboard input [2], a telephone handset [24], and gesturing in front of the display [22]. Some were found to be more successful than others: Taylor et al. [20], for example, found that users did not like using mobile phones to interact with public displays but preferred to press buttons on a system directly. Similarly, Müller et al. [12] found that mobile phone interaction with public displays did not receive as high uptake as expected, and users often opted for alternative input mechanisms. Using other input modalities, such as gestures and speech, made some people feel self-conscious [22, 24].

Regardless the interaction strategy, a prevalent challenge with using public displays is display and interaction blindness [10, 12]. Studies have shown that most people simply do not see, or pay attention to, large screens in the environment, for example, because they expect them to show adverts. Furthermore, they do not realize that they can interact with the screen, or how to do this. Input and output thus need to be carefully designed to make the possibilities of the system clear to the user [12]. To overcome display blindness, Koeman et al. [8] used a more tangible, but static, approach to visualizing opinion data, using chalk paint on the street, and fluorescent tape on a black canvas. They found that such salient visualizations encouraged people to compare and discuss the data, but that people also expected them to be interactive.

### Tangible Systems

Another approach has been to employ tangible and physical objects as a way of drawing attention to a screen, or to use physical mechanisms through which people can interact with an opinion gathering system. Physical objects can be powerful in drawing people in because they can be novel and at odds with their environment. Displaying objects that provoke curiosity, referred to as 'curiosity objects' [7], can help people to notice an adjacent display that they would otherwise walk past, and thus overcome display blindness [7]. Systems for gathering opinions have used this design principle by placing unfamiliar objects in public spaces to entice interaction, such as simple voting boxes that showed a question and asked users to submit a response by pressing a button [e.g. 8, 20] or posters with physical interactions to support activism within a local community [23]. Commercial versions of such voting boxes are currently seen in airports and shops. The project 'Vote with your feet' [19] enabled users to express their opinions on local issues by stepping on tangible 'yes' or 'no' buttons on the floor. The researchers found that the buttons attracted people's attention (much more so than the screen that displayed the question) and lowered the barrier for participation because they looked easy to use. Mood Squeezer [3] was another lightweight technology intervention that asked people to reflect on their mood by squeezing a colored ball from a box set. The squeezes were mirrored back as an aggregate colorful visualization on a public floor display, with which people engaged at various times throughout the day. Although such systems are cheap to produce and can easily be put in public spaces, they are limited by the extent to which they can gather people's opinions, because they can only ask a few simple questions, and accept a limited set of responses.

However, these systems show that tangibility can draw people in and offer familiar interactions, which makes it clear that a system can be interacted with and can entice people to 'have a go'. Playfulness is a design principle intended to encourage interaction with a system [5, 14]. However, if the interaction is too playful, it can distract from giving meaningful and honest opinions [15, 21]. While this is a recurring challenge with playful systems, Valkanova et al. [22] reported that users of their system

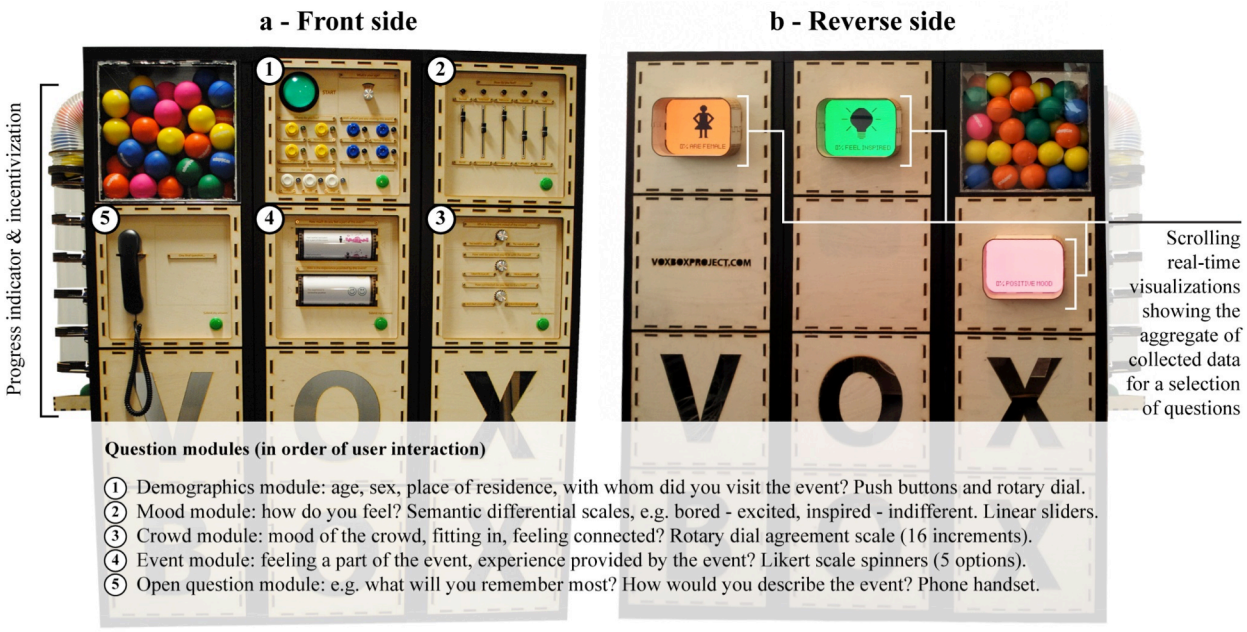


**Figure 1. Front (a) and back (b) views of VoxBox showing the question modules and visualizations**

submitted opinions instead of merely playing with the interactive features of the interface. They also observed few occurrences of off-topic communication. They argue that to achieve a good balance between playful behavior and meaningful engagement, the system needs to make it quick and easy to contribute, e.g. to enter a vote.

Tangible systems thus have great potential for gathering public opinion by being able to draw people in, make it clear that the system can be interacted with and how. The VoxBox was designed to go beyond screens and simple voting boxes by using an array of physical interactions. Its design supports semantically rich questions and enables more detailed opinion gathering while remaining easy to use and accessible to a wide variety of users.

## VOXBOX

The VoxBox, introduced in [4], was designed to create a more enjoyable and engaging experience around opinion giving. This highly visible, interactive, and playful 'tangible questionnaire' lets respondents answer questions through a range of physical interaction mechanisms (buttons, sliders, dials, etc.) on a 1.5 meter high device (Figure 1a). In addition to gathering opinions, VoxBox also visualizes these opinions in real-time on three displays on the reverse side of the system to allow people to compare how their views relate to those of others (Figure 1b). It also incorporates a physical progress bar at the side (Figure 1a). As questions are answered, a rubber ball is dropped one stage at a time through a transparent tube on the side of VoxBox. It is also intended as an incentive for completing the interaction as respondents can keep the ball afterwards.

VoxBox has five separate question modules, labeled 1-5 on Figure 1a that light up (to show which one is currently active) and are answered in sequence. Each question module is a drawer that can be inserted into or removed from the VoxBox frame and hence can be physically moved around and changed as necessary. VoxBox is therefore a modular device enabling customized questionnaires to be compiled for different events. A key challenge is how to develop underlying technologies that can support VoxBox modularity at a digital level. If question modules can be physically changed or moved around, as on VoxBox, then complex re-programming or updating of underlying digital systems is to be avoided.

### VoxBox Technology

The technology behind VoxBox is open source prototyping platforms and web technologies; in this case, Arduino microcontrollers, NodeJS server technology and a MySQL database. Figure 2 provides an architectural overview.

Each question module has its own microcontroller to handle all tangible input and output on that module. In addition VoxBox also contains a *master* microcontroller that controls the order in which the question modules become active for interaction. The master talks to the question module *slaves* over an I2C bus and each slave has a unique ID so it can be individually addressed. Since each question module can be easily removed, moved around or even new ones added, from a digital perspective, VoxBox can also support this level of adaptability without the need to do any reprogramming on any of its microcontrollers. To achieve this, when the master boots, it contacts the backend NodeJS server and asks for an ordered list of IDs for all the question modules that it is attached to. Therefore, if one were to move the question modules around, remove some or add new ones, all they have to do is change the ordered list of

question module IDs in the backend server which could be done through a simple webpage with no need to reprogram anything on the VoxBox itself.

The master is also responsible for all communication with the backend server. During each use of VoxBox the master temporarily stores all the answers from each question module and then sends all the data for that use to the server as one http request. In response the server processes and stores the collected data in a MySQL database. Additionally, the server pushes data to the visualizations (hosted on iPads) via websockets. As such, data captured using physical interactions is processed immediately and available for real-time visualization.

## STUDY METHODS

To study how the system was approached and used by members of the public, VoxBox was deployed 'in-the-wild' twice, at two separate outdoor Fan Parks for The Tour de France (an annual European cycle race) in London. By deploying the system at real events we could gain an understanding of the experience of giving feedback in a realistic event situation, where social context and lack of instructions affect how it is used [17].

The first Fan Park event where VoxBox was deployed was in a city center park, close to the action of the Tour de France cycle race. The event was attended by approximately 20,000 people, visiting throughout the day. We hereby refer to this deployment as *Green Park*. In contrast, the second Fan Park where VoxBox was deployed was in a business area away from the city center and by this time, the cycle race had moved on to France so the action could only be seen on a large screen. This event was not so well attended with approximately 1,000 people visiting throughout the day. We hereby refer to this deployment as *Canary Wharf*. Both Fan Parks were primarily social events, with many groups - mostly families with children.

### VoxBox Set-up

For both in-the-wild deployments VoxBox was set up to question the 'feel-good factor' of the events and questions were developed in collaboration with event organizers. The first four question modules ask closed questions and focused respectively on demographics (using push buttons); the user's current mood (using linear sliders); the user's connection to the crowd (through rotary knobs); and the event (through spinners). The final module uses a phone handset to ask one open question (similar to [24]) out of a set of five. The answer is then spoken into the handset and recorded. No real-time transcription of audio recordings took place during the deployments. Instead, these recordings were analyzed afterwards. Figure 1a shows an overview of the question modules used in both deployments including the questions and answers they supported.

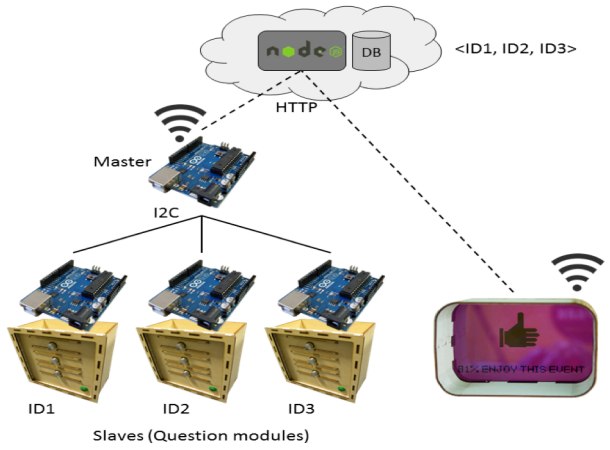


**Figure 2. VoxBox Architecture**

### Deployment Set-up

At both deployments, the set-up consisted of VoxBox, a gazebo, a blackboard with a message inviting people to come and use the device, three video cameras, and a sign advising that we were filming and explaining the research study. Figure 3 shows the set-ups at Green Park (a) and Canary Wharf (b) and how they differed slightly. At Green Park, the set-up was restricted by a stall position where only one side of the device could be seen on approach. This was predetermined by the Fan Park organizers and so was out of our control to adapt it. As such, the set-up resembled a shop front and the data visualizations could not be seen on approach. Additionally, the smaller gazebo gave the impression of restricted stallholder space so people did not naturally come in. At Canary Wharf, the stall position could be approached from both directions, and a larger gazebo created an open space where visitors could explore freely. The VoxBox was deployed for approximately six hours at each deployment.

### Data Collection and Analysis

The video cameras ran continuously throughout both deployments capturing both sides of VoxBox and the crowd. All data captured through VoxBox was uploaded and stored and researchers also performed observations and logged comments from VoxBox users while in situ. At both deployments, a minimum of three researchers were onsite at all times. One acted as the facilitator (or stall holder) and the others were observers. Protocols were defined and followed for consistency, which provided researchers the following guidance to negotiate their roles of facilitators and observers. While in situ at the deployments, the facilitator was instructed to take a hands-off approach before and during user interaction, standing or sitting nearby and only intervening if a person appeared to be having some issue. As such, facilitator-led recruitment or selection of participants was discouraged; instead, it was preferred that participation would be from self-selecting event attendees who would approach VoxBox on their own terms. Observers took up their position on the periphery of the stall and performed observations discretely so members of the public would not realize what they were doing or that they were related to the VoxBox in any way. After users

had finished interacting with VoxBox and had moved away some distance, observers would sometimes initiate a brief discussion about their experience, for example, asking how they felt about giving their opinion in this way. Spontaneous comments made by users were also logged in-situ by the facilitator (again discretely), or afterwards during video analysis.

We used a combination of 'closed' and 'open' coding to analyze video and observation data. A predetermined (closed) coding scheme was used to group and count data to answer questions such as how many people looked at VoxBox, and of those who did, how many approached it and went on to interact with it. Open coding – the clustering of similar instances in the data without a predetermined coding scheme or hypothesis – was further used to analyze such things as how people used VoxBox and the different controls; their comments; answers to the phone's open question; and any issues encountered during interaction.

## FINDINGS

VoxBox logged 163 uses over the two deployment days, with twice as many (n=110) logged at Green Park than at Canary Wharf (n=53). This was largely due to the much bigger crowd attending the first Fan Park compared with the second one. We analyzed how VoxBox was used at both Fan Parks to understand in more detail: (1) *the nature of the experience of using VoxBox*, including what people thought when approaching and answering the survey, how they completed it and if they enjoyed doing so, and whether they had any privacy or social embarrassment concerns; and (2) *the capabilities of VoxBox as a survey method*, including the extent to which it reached a diversity of people and how well it gathered meaningful responses.

### The Experience of Using VoxBox

#### *Self-selection and minimal facilitation*

During both deployment days, the facilitator did not need to approach any members of the public to encourage them to use VoxBox. The intriguing and curious appearance of VoxBox caught users' attention and drew them over, even in busy environments with other activities happening all around. The video data showed that 341 groups/individuals looked at VoxBox over the two days and of those, 70% approached it and watched, either from a distance or up close. Of those who approached, 43% went on to interact with VoxBox and in cases where they did not interact, this was primarily due to it already being in use. As such, respondents self-selected to give opinions on their own terms without feeling pressured or targeted by facilitators. On several occasions people waited for a go, forming a queue at VoxBox to interact and give their opinions.


**Figure 3. The two different set-ups at Green Park (a) and Canary Wharf (b)**

The facilitator only intervened if someone appeared to be having a problem when interacting with VoxBox. This happened typically when they had not realized that they needed to push the large green 'start' button to begin to answer the questions, or the green 'submit' buttons at the bottom of each question module in order to proceed to the next module. Several people missed turning the age dial due to its location on the box and bright sunlight sometimes made it difficult to see what question module was currently active. When this happened, a gesture or a few words to explain what to do next was sufficient to allow people to continue unaided by the facilitator.

#### *Completion of the VoxBox survey*

Completion rate was nearly 100%. Across both deployment days, only one person did not complete the VoxBox survey once they had started. In this case it was because their family had reached the front of the queue at an adjacent stand and they were forced to leave. Our observations show that people were highly motivated to finish each module. Many did not realize that they would receive a rubber ball at the end, via the tube on the side of VoxBox, but were delighted when they did. This suggested that the balls acted more as a surprise gift than the envisioned reward.

Our observations and video analysis also revealed that VoxBox supported individual and group interactions equally well. In 39% of cases, VoxBox was used by one person. The other 61% were groups ranging in size from two to five people. In such group interactions, group members typically crowded around VoxBox, talked through their answers and worked together. One group member often emerged as the lead answerer (especially for the demographic questions), and others took active roles. This included supporting actions such as (1) completing answers that the lead group member had missed; (2) correcting what they felt were incorrect answers; and (3) indicating how to proceed with the VoxBox interaction (e.g. pushing the 'start' or 'submit' buttons). Additionally, some engaged in various kinds of playful actions, including (1) playing with the controls on other inactive question modules while the lead group member was completing the survey (usually, but not exclusively, done by children); and (2) occasionally teasing the lead group member by changing answers and laughing about it before changing them back. Lead group members also actively engaged others in their group by

seeking their views before deciding on an answer. Young children were regularly involved, being asked or guided to move controls and being lifted up when they were too small to reach. Children asked parents to explain the questions, and parents involved young children by explaining it to them in more child-friendly terms: *"what do you think… the sad face or the smiley one?"*

One mother talked about how using VoxBox had had a positive experience on her family: *"It's great… the children want to do it but then it draws the parents in."* As such, individuals did not need to isolate themselves from the group/family they were with when completing VoxBox. Instead, for many groups, the activity of giving opinions became a social event with much discussion and collaboration. However, in some cases there was also a desire from group members to have their own turn to give their own personal views. This behavior was particularly observed of children, who typically are not asked for their views and feedback at such events. Hence, VoxBox was able to provide this group with a voice that was sanctioned in the presence of their parents or other group members they were with.

*Providing an engaging user experience*

The findings showed that almost everyone who approached and answered the VoxBox questions were highly engrossed in the experience. Only one person stopped midway through due to external pressures. VoxBox held their attention for longer than brief glances or 'pokes' at a machine after being lured over by its curious nature. The average interaction time of 3 minutes, 4 seconds per person/group, which is a considerable time to complete four modules of two to five multiple choice questions/rating scales and one open-ended question. Our observations of their behavior while interacting with VoxBox together with their comments confirmed that it was an enjoyable and engaging experience.

When describing their interaction with VoxBox several people talked of becoming completely absorbed in the process, as if in *"the zone"*. Another person described how she felt *"in her own bubble"* and that the event around her "*faded into the background*". In addition, many approving comments were captured while people were using VoxBox, such as *"Oh my goodness this is so cool!"* Video analysis showed that some people also explored the physical answer space by moving sliders up and down, twisting dials and turning spinners before settling on a final response. Furthermore, the different controls used on each question box encouraged a sense of anticipation and continued interaction, prompting comments such as *"I can't wait for this bit",* in reference to the spinners. After interacting with VoxBox many respondents reflected on their experience in a very positive way. For example, one person stated, *"It's something really fun but it is useful and collects data too. It doesn't take too long and it's like a game. If you came up with a <normal> questionnaire I'd run away!"* For many users VoxBox also sparked feelings of nostalgia and fond memories of tangible toys and objects from childhood. When asked to describe VoxBox, answers included *"a Willy Wonka machine", "the controls of the Tardis"* and *"some kind of fairground automata."*

*Privacy concerns when answering in public*

While taking part, there was no evidence that anyone was worried about others watching over their shoulder to see the answers they were entering. On only two occasions, older participants wanted to hide their age from the people they were with. Groups acted as a shield from others waiting for a go or looking on, although that did not seem to bother people. Even when an individual was interacting with VoxBox alone, their body was positioned in front of the current module of questions being answered, acting as a shield. Additionally, the waiting people and passersby stayed at a respectful distance, observing but never invading the current respondent's personal space; adopting the same behaviors as those observed at ATMs. Our observations also didn't show evidence of any social embarrassment or unwillingness to interact with VoxBox once people had approached it and while others looked on. In fact, quite the opposite; as most were unaware of others besides their own group when answering.

*User engagement with the real-time data visualizations*

The real-time visualizations on the back of VoxBox did not engage people to the extent we had envisioned. At Green Park, engagement was unfortunately hampered by a challenging context, where people had to walk into the back of the gazebo to view the visualizations. As such, only 4 people looked at them by walking into the space themselves, while 40 others viewed them when invited to do so by the facilitator, after they had interacted with VoxBox. In contrast, the positioning of VoxBox on a corner site at Canary Wharf enabled people to easily approach it from both sides. Here, despite the smaller crowd attending the event, a total of 29 people viewed the visualizations spontaneously while a further 6 were invited by the facilitator. However, at both deployments engagement was typically short; averaging 35 seconds at Green Park and 20 seconds at Canary Wharf. The visualizations relating to the crowd and the social dynamics of the event evoked surprise and typically drew more attention and comments from those looking at them. In particular, visualizations showing the percentage of people who felt bored, unsafe, or that they didn't 'fit in' with the crowd led to spoken exclamations of the data such as *"5% don't fit in…they don't fit in then go home! Can't believe it."*

**The Capabilities of VoxBox as a Questionnaire**

*Reaching a diverse audience*

VoxBox attracted a diverse user demographic across both deployment days. The observation videos captured people of all ages, from toddlers to the elderly, approaching and using VoxBox. They included individuals, families, groups of friends, tourists, locals, people in wheel chairs and even

while sitting on their bikes. The logged demographics data also shows users from across all selectable genders, ages, group types and home locations. Overall, 41% of users indicated they were female and 46% male with the remainder selecting 'other' or not answering. In terms of age, 50% of users indicated that they were 24 or younger, 41% between 25 and 44, and 9% between 45 and 64. The majority of users came from London or a nearby region (57%), and almost a quarter of users (23%) indicated they were from a different country. Most were with family (61%) or friends (16%), with the rest coming alone or with some other acquaintance.

*Eliciting 'serious' answers*

Although VoxBox was designed to be playful and deployed in a fun vibrant setting, our findings suggest that the majority of respondents took the survey seriously and answered the questions with some reflection. People physically moved the sliders and spinners up and down before finally settling on a level they were happy with. They discussed the questions with others in their group before selecting their answer and even individual users were seen talking aloud to themselves about the current question and what their answer should be. Only one child was observed randomly moving the sliders up and down and pressing buttons without thinking about their answers.

We analyzed the data collected through VoxBox to see if there were differences between the answers from the two Fan Park events. Our own experiences of the events led us to hypothesize that answers collected at Green Park would be more positive than those from Canary Wharf due to much higher attendance figures and a more positive and 'buzzing' atmosphere at Green Park. Twenty times as many people were in attendance at the Green Park event in comparison to the Canary Wharf event, which was held towards the end of the Tour de France when local interest had waned. Even though the organizers had predicted a good attendance at Canary Wharf, many stall owners made a huge loss as a result of the far fewer numbers of people coming along. This led to a much emptier feel on the day with less appearing to be going on. This difference was reflected in the data collected and respondents' answers were generally less positive at Canary Wharf.

| Question | Green Park | | Canary Wharf | | p-values |
|---|---|---|---|---|---|
| | other | +ve | other | +ve | (p < 0.05) |
| **Mood of crowd** | 31 | 79 | 27 | 26 | **0.0076** |
| Fit in | 64 | 46 | 32 | 21 | 0.9228 |
| Connected to crowd | 76 | 34 | 33 | 20 | 0.4903 |
| **Part of event** | 26 | 84 | 43 | 10 | **5.013E-10** |
| **Positivity of event** | 50 | 60 | 43 | 10 | **3.45E-05** |
| **Bored/Excited** | 11 | 99 | 14 | 39 | **0.0127** |
| Unsurprised/Surprised | 32 | 78 | 13 | 40 | 0.672 |
| Unwelcome/Welcome | 32 | 78 | 13 | 40 | 0.672 |
| **Indifferent/Inspired** | 13 | 97 | 19 | 34 | **6.55E-04** |
| Unsafe/Safe | 13 | 97 | 19 | 34 | 0.348 |

**Table 1. VoxBox question results from Green Park and Canary Wharf, collapsed into positive (+ve) and 'other' sets with chi-squared p-values.**

We conducted chi-squared tests on the logged responses for each question from Green Park and Canary Wharf. Since the answer data was mostly scale based (Likert and semantic differential scales), we wanted to highlight the distance between the most common points on the scale for each question (e.g., if a question resulted in points at the higher, more positive end of the scale for Green Park and the lower, more negative end of the scale for Canary Wharf). To do this, we divided each answer scale into two sets where one set contained responses at the highest, most positive points of the scale (+ve) and the other set contained all other responses (other) at lower points of the scale. Table 1. shows the results of the chi-squared tests as well as the counts of high positive (+ve) and 'other' responses for each question from Green Park and Canary Wharf. As is highlighted in bold, significant differences were found between the two regarding mood of the crowd, feeling part of the event, the positivity of the event, whether people felt bored or excited, and whether people felt indifferent or inspired. In all these cases Green Park showed significantly more positive responses than at Canary Wharf. These differences support our own observations of each event as mentioned above, providing clear evidence that VoxBox was able to elicit true responses that reflect the different experiences at the two Fan Parks. (Note that although Unsafe/safe has the same counts as Indifferent/uninspired, responses to the latter question covered a wider range of points on the scale, hence the difference in p value.)

*Handling responses to open and closed questions*

VoxBox asked a series of closed questions followed by a final open question through the Phone handset. Everyone picked it up when it rang and provided a large number of varied responses to the questions posed. Video analysis shows expressions of surprise and delight when people heard the phone ring and a large majority answered the phone with a typical "Hello?" Analysis of the recorded responses also showed that two out of three people answered the open question with a relevant, and often witty, comment. The question "how would you describe this event to a friend?" for example, resulted in lengthy answers, such as *"Looks like it's going to be amazing. We're here early on but it's already packed with lots of people. There's tons of food vendors, lots of flags, big screen TV, it's gonna be great."* Other responses included exclamations such as: "*exciting", "great", "cool", "massive, crazy and fun", "very hot, with lots of food."* Answers to the question about what the organizers could change included: *"I think you don't have to change anything - everything is perfect. Bye, bye!", "maybe some more giant screens - thank you!", "more things for kids"* and *"you could hold it in Australia"*; while what one person would most remember was *"the crowd is really nice."*

In one third of uses, no response was left. This could be because there were a large number of tourists at the Fan Parks and therefore language might have been a barrier. Secondly, when young children answered the phone they

were keen to pick up the handset and say "hello", as they instinctively knew how to do. However, when they heard a strange adult voice talking to them, they often didn't know how to react (particularly the younger children) and handed the phone to their parents who by this time had missed the question being asked.

## DISCUSSION

The findings from the VoxBox study revealed that by making a questionnaire tangible, we were able to turn what is typically perceived to be a disruptive, mundane activity into one that was seen as a 'have a go' activity. Below we discuss lessons learned and outline design implications that consider the benefits and drawbacks of moving to physical interfaces for eliciting information and feedback from people in public places.

### Lessons Learned

*Physical questionnaires benefit users and facilitators*

An aim of VoxBox was to better blend the activity of opinion giving with the event experience, so that opinions could be gathered in situ without disrupting someone's event experience. The findings from our observational study show that this was a successful approach and that something that is normally perceived to be disruptive or a pain to do can be transformed into something that is enjoyable, social and reflective. In particular, the physicality of the VoxBox was successful at drawing people in and enticing them to start interacting, after which almost all of them were captivated to complete all the questions. The experience of giving opinions with VoxBox was engaging, playful, and fun. It did not jar or get in the way of the event experience itself. As such, people did not have to explicitly leave the 'event mode' and enter into 'feedback mode'. Allowing for group interactions aided this transition since none of the group members had to isolate themselves from their family or friends to give feedback, while the others waited. Instead it became a fun family or group activity that included everyone.

From the facilitators' perspective, the experience of gathering opinions through VoxBox was also seen to be positive, as no explicit selection of participants and a minimum amount of mediation was required. This is seen as a benefit for facilitators as approaching a person or group of people for feedback can be a daunting task. By removing this requirement and instead giving passersby the decision to approach or not made for a more pleasant facilitator role. Facilitators could simply watch on as people came up to VoxBox, and did not have to cajole people all the time. Thus, tangible questionnaires can blend opinion giving into the event experience, which is beneficial for both respondents and facilitators. In addition, as autonomous, standalone systems, tangible questionnaires also remove selection bias that can be introduced into the opinion giving process by the facilitators (i.e., only approaching people they think will be likely to answer their questions) [11]. A selection bias may still exist to some degree (i.e., based on those who are more willing to approach and interact with the device). However, our data showed that this was not likely to be the case, judging by those who took part on both days being from a diverse demographic of people.

Although VoxBox was successful at drawing people in to interact and minimized facilitation, it is worth considering how this ability may reduce over time if such tangible questionnaire devices became more familiar. VoxBox did benefit from the novelty effect, and unlike public screens, users had no preconceptions about the device, sparking curiosity and a need to explore further. If such tangible questionnaires do become more ubiquitous in future, developers should also consider initiatives to sustain curiosity and engagement. Longer-term work on public displays could provide inspiration and possible solutions to this challenge.

*Playful engagement can also elicit serious responses*

But how do we know that the responses elicited to the questions posed are genuine? For our context at the Fan Parks, since the lines between event experience and opinion giving were intentionally blurred, passersby might not have entered into the mindset of answering questions and instead simply played with the tangible device. Our observations showed that this was not the case with VoxBox on either deployment day. Far from just playing with it, people got into their 'answering zone', thoughtfully providing answers to the questions in relation to the event. We would argue that physical interaction provided a compelling way for people to consider their responses by letting them explore the answer space (e.g., physically moving sliders up and down) before reaching decisions and submitting their input. These observations are also backed up by quantitative comparisons between the responses gathered at both Fan Parks, which align with the less positive atmosphere experienced at Canary Wharf.

A simple physical interface with clear affordances, direct questions and familiarity with the physical inputs such as buttons, sliders and knobs meant that VoxBox was not abstract or open to appropriation for different uses (such as play). This clarity was further emphasized by basing all interactions on familiar survey concepts such as radio buttons and Likert scales. The telephone handset that was used for the open question is also a very familiar interface that everyone knows what to do with, as was evidenced by the natural reactions to pick it up and say "hello" (even by very young children). By combining physical and familiar interfaces in this way, tangible questionnaires have much potential to provide a fun, enjoyable experience while also capturing useful information for facilitators, visitor experience evaluators and event organizers.

*Privacy or social embarrassment can be mitigated*

VoxBox was deployed at public events in busy spaces with many other people around. No privacy initiatives such as screens were used to obscure interactions with it and as such it was possible to view how people were interacting

and what they were entering. Our observations show that in this particular setup privacy and social embarrassment issues did not arise. In terms of privacy, we propose this was primarily due to the non-serious nature of the event and because the questions posed on VoxBox were not of a sensitive nature and mostly subjective, querying how people felt about different aspects of the event. If the atmosphere of the wider event had been of a more serious nature, e.g., a public demonstration, a playful tangible questionnaire would be at odds with its surroundings and privacy and social embarrassment issues would be much more critical. However, based on our observations, we propose three initiatives that could help reduce such concerns when interacting with a tangible questionnaire.

Firstly, even if the system is a public device, it can still provide a more personal and somewhat private interaction experience, similar to that experienced when at an ATM. By using small, discrete individual interactions it is difficult for others to 'shoulder surf' and see what has been entered. It can be made even more private by having small question and answer text so they cannot be read from a distance. With VoxBox, the user could also naturally obscure the view of their inputs when standing in front of it to interact. Even the vocal telephone interaction did not cause anxieties, we suggest due to the use of a telephone handset, which again made the response giving more discrete and personal. Secondly, using tangible inputs with clear affordances, such as buttons, sliders and dials help people to immediately know how to use the device and reduce worries of making a mistake or not using it correctly. Finally, if a physical questionnaire supports or encourages group interactions, this can also reduce potential concerns about using it in public, as the spotlight does not fall on one person. However, willingness to interact as a group also depends on the application and the sensitivity of the questions and is discussed in more detail below.

### Design Implications

#### *Physical Questionnaire Throughput*

VoxBox was implemented as a single system that allowed for one individual user or one group of users to interact at any time. Thus throughput was influenced by how long each complete interaction took. For application areas that require a higher throughput (e.g., at larger events or where users have less time to interact), future tangible questionnaires could use a design that asks fewer, focused questions or only asks closed questions (which take less time to answer than open ones, such as the telephone). However, this would limit the amount of detail and richness of data collected. Another approach is to distribute the individual question modules throughout a physical space rather than having them all on one device. For example, each of the VoxBox question modules could be an individual device in its own right and, as such, different people could interact with them in parallel, thus reducing waiting times and increasing throughput speed. However, in a distributed system setup, consideration must be given to how to track input from the same user (or user group) across all devices. To support this, mechanisms could be implemented in which the user registers with each distributed part of the system, for example through the use of a smart card, so that all answers are then linked together.

#### *Supporting Group Input*

When developing a tangible questionnaire, careful consideration must be given to whether it is desirable to encourage and support group or solo interactions. There are two key drivers that contribute to this decision. One is the nature of the application and the sensitivity of the questions being asked. If they are of a private or sensitive nature they may naturally lend themselves to being completed by one person in a discrete setup. However, if privacy and sensitive questions are not issues, supporting group input should be considered. The observational study results showed that group input is a positive behavior and worked very well in informal event settings like the Fan Parks. It encouraged discussion and participation by all group members. The second driver is whether or not it is necessary to link input data to an individual. For example, in some contexts such as an exhibition space or gallery, it may be desirable for facilitators to know whether a parent or child within the group entered a specific piece of feedback. This would require each individual group member to register themselves before each point of interaction (e.g., using a smart card); however, these additional registration processes could become complex and would likely affect throughput speed and user experience. As such, the device could be designed to prompt a lead user by asking for one member of the group (e.g., the youngest or the oldest) to answer the questions.

#### *Language Barriers*

Survey question and answer text can sometimes be difficult for children or non-native English speakers to understand. This is a challenge for any opinion gathering method, including online or paper-based surveys. One solution is the use of non-verbal forms of opinion gathering using symbols, images or videos, gestures or sounds. With children specifically, opinion gathering can be challenging [16] and the benefits of using playful interactions such as those above have been discussed in relation to usability, learning, collaboration, and fun [25]. The observational study of VoxBox also highlighted the challenge of gathering open responses from children. We attempted to do this through the telephone, which worked to some degree but could be improved on future systems. While children were keen to answer the phone when it rang, they were often taken aback by the disembodied adult voice on the other end of the line and did not want to respond. An alternative would be to use video footage of a friendly character asking a question, to which the children respond. This would give more context of whom they are talking to.

#### *Visualizations*

At both deployments people only briefly looked at the

VoxBox visualizations and did not extensively discuss the data with others. We propose this was because people were not expecting real-time data to be available as it is not typical to be able to view data that is being collected in situ. However, we suggest that if such data visualizations became more familiar, they could provide a tool for further reflection and discussion, as well as making opinion gathering more reciprocal. This opens up an interesting design space around how to present data to participants in engaging and thought-provoking ways, and in ways that make sense to them. One approach could be to implement the means for participants to explicitly relate their own answers to those of the crowd by showing user data and aggregate crowd data side by side. Another approach could be to think of ways in which to make abstract data more personally meaningful to each user. The challenge is to provide the same positive experience for tangible questionnaire output as that which was achieved for tangible input during the observational study.

### Adaptability and Ubiquitous Potential

During and following the Fan Park deployments, there was a huge amount of interest in VoxBox, and an overwhelmingly positive response. In particular, there was much commercial interest in the system with a number of people asking where they could get a VoxBox and how much it would cost to buy one. Interested parties included market researchers, event organizers, conference organizers, a theatre production company, an exhibition center, a fruit shop, a travel company, and several departments within our own university. Envisioned user groups included children, students, and elderly people, and possible uses varied from gathering immediate feedback on the taste of fruit juice, to understanding the experience of buying a second home, to gathering census data, to obtaining feedback on lectures or conference presentations. One market researcher commented: *"I've been in market research for over 12 years and this is the best questionnaire I've seen so far."* Tangible questionnaires like VoxBox were seen as particularly useful for eliciting feedback from young children, as it is very difficult to gather their feedback using traditional methods.

The breadth of interest shows the potential in a more engaging, playful and accessible approach for gathering public opinion in situ. Event organizers who approached us about VoxBox confirmed that the use of a system such as this is more in keeping with the fun experience they want to create at events compared with more conventional survey methods, and that a 10-15% response rate is typically a good result for them. However with tangible devices there are challenges of customizability and scale, the costs of which must be weighed up against the costs of other opinion gathering alternatives. In terms of the customizability of VoxBox, there is a limit to the number of questions that can be asked within a set number of question modules (which are determined by the size of the system). However, within these constraints customization of VoxBox for different events is relatively simple and cheap, only requiring the switching of detachable question and answer text labels or creating a new question module to slot into the device frame. In terms of building more VoxBoxes to reach more people, physical questionnaires will have up-front costs to build the devices; however, the costs of re-deploying those devices at future events would be minimal compared to paying the wages of facilitators with clipboards every time.

So 'when physical?' and 'when not physical?' We propose that physical opinion gathering systems such as VoxBox are most successful in situations where people are relaxed and not in a hurry, simply because it can take time to meaningfully interact. It is ideal when deployed to blend into the wider event experience as this is less disruptive and helps to contextualize the opinion giving experience and the questions asked. With the growing trends of action research and technology for civic engagement we also see tangible questionnaires as potential tools to facilitate opinion gathering on local issues by councils or local groups. Indeed, our future work is considering how such technologies could become more embedded in everyday environments and brought closer to the general public.

## CONCLUSIONS

VoxBox is a large tangible questionnaire intended to gather opinions from the crowd at events in a more engaging and less disruptive manner. Drawing on findings from two in-the-wild deployments, we presented insights and learnings for researchers and developers of future tangible questionnaires. Firstly, the activity of opinion giving at events should be more a part of the wider event experience to reduce its disruptiveness and make it a more enjoyable thing to do. Tangible questionnaires can achieve this, providing benefits for both respondents and facilitators. Secondly, group interactions during opinion giving should be supported where the application area lends itself to this, and where sensitive or personal questions are not used. Thirdly, even interacting with a large tangible device in a public space can be made to feel like a more personal experience through the use of discrete physical interactions and content that is sized for the interacting individual/group alone. Additionally, by combining discrete physical interactions with familiar form factors (e.g., buttons, sliders, etc.), device re-appropriation (e.g., for play), social embarrassment and anxieties about interacting with a novel device are mitigated. Finally, designers of future tangible questionnaires should consider the requirements of a given application area such as throughput, capabilities of the expected audience (e.g., language barriers) and how opinion giving is reciprocated through visualizations or rewards.

This work highlights the potential for tangible questionnaires that go beyond simple voting possibilities into richer ways of collecting data. Perhaps one day they could be as familiar as ATMs where anyone can use them to give their views in situ at any time, reciprocated by

visualizations of the views of the wider population or perhaps even tangible takeaways [13].


## ACKNOWLEDGEMENTS

This research was funded by the EPSRC and the Intel Collaborative Research Institute: Cities. We would like to thank the Tour de France Fan Park organizers and everyone who participated in the study. Thanks also to Giovanni Quattrone, Alessia Romano and Amber Parish for their support with data and video analysis.


## REFERENCES


1. Blignaut, P.J. Computerized self-administered questionnaires on touchscreen kiosks: do they tell the truth? In *Proc. CHI 2004 Ext. Abstr.*, ACM Press (2004), 1183-1186.
2. Brignull, H. and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. Interact 2003*, Rauterberg, M., Menozzi, M., and Wesson, J., (eds). IOS Press, 2003, 17-24.
3. Gallacher, S., et al. Mood Squeezer: Lightening up the Workplace through Playful and Lightweight Interactions. In *Proc. CSCW 2015*, Vancouver, Canada, 891-902.
4. Golsteijn, C., et al. VoxBox: a Tangible Machine that Gathers Opinions from the Public at Events. In *Proc. TEI 2014*, ACM Press (2014), 201-208.
5. Hespanhol, L. and Tomitsch, M. Designing for collective participation with media installations in public spaces. In *Proc. MAB 2012*, ACM Press (2012), 33-42.
6. Hosio, S., et al. From school food to skate parks in a few clicks: using public displays to bootstrap civic engagement of the young. In *Proc. UbiComp 2012*, Springer-Verlag (2012), 425-442.
7. Houben, S. and Weichel, C. Overcoming interaction blindness through curiosity objects. In *Proc. CHI 2013 Ext. Abstr.*, ACM Press (2013), 1539-1544.
8. Koeman, L., et al. What chalk and tape can tell us: lessons learnt for next generation urban displays. In *Proc. PerDis 2014*, ACM Press (2014), 130-136.
9. Kohut, A., et al. Assessing the Representativeness of Public Opinion Surveys, Available at: http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/ (2012), 1-51 [Accessed September 2014].
10. Marshall, P., et al. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proc. CHI 2011*, ACM Press (2011), 3033-3042.
11. Miller, K.W., Wilder, L.B., Stillman, F.A., and Becker, D.M. The Feasibility of a Street-Intercept Survey Method in an African-American Community. *American Journal of Public Health 87*, 4 (1987), 655-658.
12. Müller, J., et al. Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In *Pervasive Computing*, Tokuda, H., et al., (eds). Springer Berlin Heidelberg, 2009, 1-8.
13. Nissen, B., et al. Volvelles, Domes and Wristbands: Embedding Digital Fabrication within a Visitor's Trajectory of Engagement. In *Proc. DIS 2014*, ACM Press (2014), 825-834.
14. Ojala, T., et al. Multipurpose Interactive Public Displays in the Wild: Three Years Later. *Computer 45*, 5 (2012), 42-49.
15. Peltonen, P., et al. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In *Proc. CHI 2008*, ACM Press (2008), 1285-1294.
16. Read, J. C., and MacFarlane, S. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proc. IDC 2006,* ACM Press (2006), 81-88.
17. Rogers, Y., Yuill, N., and Marshall, P. Contrasting lab-based and in-the-wild studies for evaluating multi-user technologies. In *SAGE Handbook of Technology Research.*, Price, S., Jewitt, C., and Brown, B., (eds). Sage, London, 2013.
18. Schroeter, R. Engaging new digital locals with interactive urban screens to collaboratively improve the city. In *Proc. CSCW 2012*, ACM Press (2012), 227-236.
19. Steinberger, F., Foth, M., and Alt, F. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proc. PerDis 2014*, ACM Press (2014), 44-49.
20. Taylor, N., et al. Viewpoint: empowering communities with situated voting devices. In *Proc. CHI 2012*, ACM Press (2012), 1361-1370.
21. Tomitsch, M., et al. Who cares about the Content? An Analysis of Playful Behavior at a Public Display. In *Proc. PerDis 2014*, ACM Press (2014), 160-165.
22. Valkanova, N., et al. MyPosition: sparking civic discourse by a public interactive poll visualization. In *Proc. CSCW 2014*, ACM Press (2014), 1323-1332.
23. Vlachokyriakos, V., et al. PosterVote: Expanding the Action Repetoire for Local Political Activism. In *Proc. DIS 2014*, ACM Press (2014), 795-804.
24. Whittle, J., et al. VoiceYourView: collecting real-time feedback on the design of public spaces. In *Proc. Ubicomp 2010*, ACM Press (2010), 41-50.
25. Zaman, B., et al. Editorial: the evolving field of tangible interaction for children: the challenge of empirical validation. *Personal and Ubiquitous Computing*, *16*, 4 (2012), 367-378.