Tangible Self-Report Devices: Accuracy and Resolution of Participant Input

Niels van Berkel
nielsvanberkel@cs.aau.dk
Aalborg University
Aalborg, Denmark

Timothy Merritt
merritt@cs.aau.dk
Aalborg University
Aalborg, Denmark

Anders Bruun
bruun@cs.aau.dk
Aalborg University
Aalborg, Denmark

Mikael B. Skov
dubois@cs.aau.dk
Aalborg University
Aalborg, Denmark

ABSTRACT

Tangible input has been explored as a means for participants to self-report experiences while minimising disruption and allowing for discrete data collection. However, the accuracy of these tangible devices has not been studied systematically. We compared six input techniques, including slider, slider with resistance, capacitive touch slider, squeeze, rotary knob, and joystick, to understand their accuracy and resolution profile. Each of these wireless devices was designed in a similar form factor and intended to be operated discretely with one hand. We assessed input accuracy and participant perceptions across devices through a controlled lab study (N = 20), highlighting diverging limits to the accuracy of the input technique and possible explanations for the differences in resolution. Our results indicate that participant accuracy was highest using a slider, and lowest using a squeeze-based input. We discuss the suitability and challenges of discreet tangible self-report techniques, and highlight open research questions for future work.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**; *Mobile devices*; *Empirical studies in ubiquitous and mobile computing*.

KEYWORDS

Self-report, experience sampling method, ecological momentary assessment, accuracy, reliability, tangible user interfaces

ACM Reference Format:
Niels van Berkel, Timothy Merritt, Anders Bruun, and Mikael B. Skov. 2022. Tangible Self-Report Devices: Accuracy and Resolution of Participant Input. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22), February 13–16, 2022, Daejeon, Republic of Korea.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3490149.3501309

TEI '22, February 13–16, 2022, Daejeon, Republic of Korea

ACM ISBN 978-1-4503-9147-4/22/02…$15.00
https://doi.org/10.1145/3490149.3501309

1 INTRODUCTION

In order to study aspects of human experiences that are difficult to capture through automated means, *e.g.* pain levels [31], researchers and practitioners often rely on the use of self-reports. Self-report studies, such as diary studies [8] or experience sampling method (ESM, also known as Ecological Momentary Assessment)-based studies [35], ask participants to (repeatedly) answer a question in which they assess their current state. Such studies are common in Human-Computer Interaction (HCI) as well as the wider academic and professional landscape [64], as they can provide a more detailed understanding of an individual's experience as compared to a one-off survey. In contrast to traditional surveys, ESM studies reduce reliance on participants' ability to accurately recall prior experiences by asking them to reflect on their current experience rather than an event or experience that occurred in the past [16, 35, 64]. Use of the ESM has evolved from analogue pen-and-paper input, to PDAs, and eventually to smartphones as the predominant form of data collection [64]. Recent work shows a need among participants to more discreetly collect self-report data [1, 51]. Furthermore, the process of retrieving one's smartphone, unlocking, and subsequently providing the self-report input has been reported as a barrier to participants providing self-reports [29, 44].

The need for more discreet and convenient input methods has resulted in the development of various bespoke input devices for self-report studies. These include, *inter alia*, a squeezable wearable to record pain levels [1], analogue data collection aimed at Parkinson patients [68], and plush toys with tangible body parts [18]. As these tangible input devices often miss a screen, participants are unable to confirm their input. This could, therefore, easily result in discrepancies between the participant's intended input and the value collected by *e.g.* the researcher or medical expert – potentially resulting in undesired consequences. For example, over- or under-reporting of pain levels can result in an incorrect adjustment of medication. Although a number of different form factors have been explored in the literature, the effect of these tangible devices on the discrepancy between participants' intended and recorded self-reports remains unknown.

To inform the design of future tangible self-report devices, we carried out a lab study (N = 20) in which we compared participant input accuracy across six tangible devices – assessing distinct input techniques (*e.g.*, squeeze, rotate, slide). For each device, participants were asked to input a set of ten randomised target values while holding the device out of sight. By calculating the difference between the recorded input values and the randomised target values, we assessed the devices' accuracy in recording the intended user input. Following prior work from the pain assessment literature [31], we collect input across a high-resolution range (0–100) – in contrast to earlier work on tangible self-report devices. Furthermore, we collected participant preferences towards these tangible devices.

In line with the increasing work in the HCI community towards tangible self-report devices, we contribute a systematic evaluation across distinct input techniques to enhance our understanding of the methodological constraints in the use of tangible self-report devices. Our work highlights implications for research using tangible devices in self-report studies and provides suggestions for future work in this domain.

2 RELATED WORK

The form factor of ESM questionnaires has changed drastically since the initial development of this self-report method. In the initial ESM studies, participants completed pen-and-paper self-reports following the arrival of an alert on their pager [35]. Participants were, therefore, required to carry with them both the questionnaire papers and a beeper at all times. In the early 1990s, researchers adopted the use of Personal Digital Assistants (PDAs), allowing for more advanced questionnaire logic [61] and insights into participant data entry (*e.g.*, response time) [62]. The eventual introduction of the smartphone largely replaced the use of pen-and-paper technology and PDAs [64] due to widespread smartphone adoption among participants as well as the greatly improved sensing and presentation capabilities [41, 49, 64]. Wearable and physical devices have since been identified as a future direction for self-report studies [64].

2.1 Alternative Techniques for Self-Report Data Entry

Starting from the presentation of traditional surveys on a mobile format, various researchers have explored alternative input techniques for a variety of reasons. A common motivation for alternative input techniques is to reduce participant strain. Truong et al. present a smartphone input mechanism which allows users to answer a question while unlocking their phone [63]. Choe et al. introduced a mobile widget for rapid data input without requiring participants to open an application, resulting in a higher response rate [10]. Combining the widget interface with traditional smartphone notifications, Visuri et al. present 'alert dialogues' – an interactive popup that allows for immediate data entry which resulted in high response rates [69]. Hernandez et al. explore three different form factors for self-reports; smartphone, smartwatch, and Google Glass [28]. Their results show that while participants' initial interaction (the time between incoming notification and opening notification) is shortest with a Google Glass and longest with a smartphone, no difference was found in total response time between devices. Youn et al. propose 'WristDial', in which the rotation of the wrist, as measured by a smartwatch, is used to enter a number from 1–10 [76]. Participants are provided with either tactile or speech based feedback, both of which result in high levels of input accuracy. Recently, Yan et al. showed that participants' perception of smartwatch interface features for self-report were affected by the context in which the self-report is completed [75].

Although the HCI literature has focused strongly on reducing participant strain in order to sustain or increase response rates, other motivations are also considered. Our paper is strongly inspired by Adams et al., who present a tangible device for self-reporting pain levels [1]. Their results highlight a preference among chronic pain patients for discreet and convenient input. To this end, Adams et al. develop 'Keppi', which allows users to squeeze a physical device to report their pain level. The authors present three form factors of Keppi; a necklace, a bracelet, and a keychain – all capable of capturing four degrees of input (absence of pain, and a low, medium, and high pain level) [1]. Price et al. present 'Painpad', a numeric input pad to self-report pain values during hospital stays [47]. Painpad showed improved patient compliance as compared to a tablet, and highlights that self-reported pain levels might be more faithful than scores collected by nurses, which were found to be systematically lower. While both aforementioned papers focus on the self-report of pain, their input technique and input range differ significantly; pressure-based across three categorical levels (plus 'no pain' through no pressure) [1] versus haptic-input through a numeric input pad across a 0–10 range [47].

Other solutions are characterised by their low level of technological complexity. Vega et al. assess the use of an analogue booklet for self-assessment by Parkinson patients [68]. In addition to its low costs, Vega et al. praise the solution as "*accessible, frictionless, personalised, portable, low-demand, automatically encoded, straightforward and flexible*" [68]. Tangibles in which the technical elements have been removed or camouflaged often aim to support affective interactions, a third motivator for the design and exploration of alternative self-report techniques. Balaam et al. introduce the 'Subtle Stone' to a high-school classroom; an affective self-report tool which changes colour when squeezed – allowing for emotional communication between students and teacher [2]. Guribye et al. explore a hand-held, tangible stone that allows for data recording through squeezing [23]. Data is recorded and can subsequently be visualised on a companion tablet application. Finally, Duong et al. explore the design of a plush toy augmented with tangible input options to assist in the assessment of children [18]. The goal of these devices is "*not necessarily to minimize the obtrusiveness but to design the technology so that the interaction becomes part of an affective experience*" [23].

The discussed studies highlight the variety of motivations behind, as well as the possibilities enabled by, bespoke tangible devices. In this article, we systematically explore the discrepancy between tangible input as compared to the intended input, motivated by an increasing uptake of tangible self-report devices and the need to collect reliable data.

2.1.1 Tangible Data Input Techniques. Expanding the scope of our related work beyond self-report data collection reveals a plethora of different input techniques and motivates further work on eyes-free interaction. Boem & Troiano present an overview of deformable input techniques [4], identifying five categories which form the basis of deformable interfaces; shape, material, input sensing, I/O mapping, and the use of deformable input. Here we highlight different input sensing techniques, which Boem & Troiano categorise as either 'embedded sensing' or 'external sensing'. Embedded sensing relies on sensors embedded in the tangible device. An example is the 'Skweezee System', which allows for gestural interactions through a squeeze-based interface [67]. Skweezee objects contain conductive filling which changes the resistance of the embedded electrodes, which is subsequently picked up by a classifier to identify the user's gestures. Recent work by Wu et al. presents a pocketed-based textile sensor which combines inductive, capacitive, and resistive sensing to provide both gesture and object detection in the pocket [72]. In terms of gestures (*e.g.*, swiping), the researchers were able to distinguish between low and high pressure.

In addition to the creation of bespoke devices, prior work has considered how to expand the input options on existing devices. For example, Harrison & Hudson explored the use of 'shear', the force tangential to a screen's surface, as an additional touchscreen input technique [24]. Corsten et al. highlight how the force and angle of thumb-input can be used on smartphones to reduce the size of input widgets [15]. Taken the concept of thumb-based input even further, Xu et al. present a miniature fingertip keyboard that allows for eyes-free text entry based on 3D tracking of participant fingers [73]. The backside of smartphones, as well as the case and bezel of smartwatches, have also been explored as an input surface [71, 77], allowing users to interact without obstructing the screen.

Eyes-free interaction removes visual feedback, increasing reliance on users' kinesthetic sense [5]. Movements in eyes-free interaction are kinesthetically identifiable, meaning that the user is aware of the position and movements [43]. The human haptic system is bi-directional, in that it can perceive and act on the environment simultaneously [60]. This tight coupling between motor output (movement and exertion) and sensory input (muscles, joints, and skin) provides us with an implicit closed-loop control. This enables eyes-free interaction that does not require additional feedback.

2.2 Data Quality and Reliability

The reliability of participant responses in ESM studies has seen increased attention in the HCI literature; "*As researchers largely rely on human contributions, ensuring a sufficient level of accuracy in these contributions is essential to produce valid and replicable study results.*" [66]. In an effort to ensure, and potentially increase, the reliability of participant responses, a wide range of contributions have been considered. Xinghui et al. stress the need for reducing participant burden by minimising the disruptive nature of self-report notifications and lowering interaction efforts [74], and introduce five design requirements; 'minimal disruptiveness', 'inclusiveness', 'low-focus', 'intuitiveness', and 'expressivity'. Berrocal et al. suggest to complement self-reports with peers through 'PeerMA', in which a person close to the participant (*e.g.*, spouse) provides 'peer-reports' [3]. PeerMA can help to *e.g.* reduce the social desirability bias, in which participants provide socially desirable answers rather than their honest self-assessment. Gamification, the introduction of game elements in a non-game context, has been shown to increase the quality of participant contributions [65]. Matejka et al. studied the effect of visual appearance of sliders and visual analogue scales on participant input [38]. Their results highlight that decorations, such as for example tick marks at the 25%, 50%, and 75% point of an analogue scale, introduce bias amongst participant responses. Based on these results, Matejka et al. recommend against the use of tick marks along the input axis. Finally, Rabbi et al. present 'ReVibe', a system which asks participants to reflect on multiple moments throughout their day retrospectively, supported by automatically-collected sensor data [48]. The proposed system aims to reduce the interruption of participants throughout the day while providing a memory aid to reduce the effects of recall-bias.

The current study aims to assess the reliability of tangible self-report devices in capturing the intended data input from study participants. In particular, we highlight the discrepancy between the intended and the recorded input value as a potential detriment to the reliability of tangible self-report input.

3 METHOD

To study the accuracy of different input techniques for tangible self-report devices, we first designed six distinct devices. Devices were designed to follow a similar level of finish, shape, and user comfort to allow for a direct comparison between devices. The devices were assessed in a lab-study, in which participants aimed to accurately input numbers as shown on a screen. Following each set of ten, participants assessed the usability of the input device and were able to provide feedback. Here, we discuss the design of the input devices, the study software, as well as the study procedure.

3.1 Design Considerations

We first present our design considerations and introduce the physical components and input techniques of the self-report devices. Following earlier insights which highlight that participants feel a need to report *e.g.* anxiety or pain levels discreetly [1, 51], our first design consideration was to ensure that the devices can record participant input without a direct line-of-sight (*e.g.*, keeping the device in their pocket). To support this, all devices were designed to be wireless, with an onboard microcontroller capable of sending data via WiFi (Adafruit Feather HUZZAH ESP8266). Following the same motivation, our second design requirement was to ensure that all devices can be operated using one hand to support discreet input. Our third design consideration was to ensure consistency in the appearance, fidelity of the physical form, and a uniform level of refinement of the input devices. This was to ensure that none of the devices would stand out as being distinctive in their visual appearance nor that the quality of the construction or other aesthetic concerns would bias the results [53].

Beyond these initial requirements, the five design guidelines for eyes-free interaction identified by Oakley & Park were helpful in focusing our investigation [43]. The devices should provide for *self-monitored input*, meaning that the kinesthetic sense is involved in the input, whether that involves being aware of the position of the body or pressure. The *input should reflect bodily constraints*, in that the motion and stability needed to provide input is compatible with what is comfortable for the body. Thirdly, *minimal interaction models* involve simple mapping between the kinesthetic state and the state of the system, which are much easier for a user to use as opposed to more complex mappings, such as controlling a virtual cursor, which requires additional cognitive demand. Additional feedback – when provided – should be *immediate*, which enables the person to learn the mapping and input quickly. The learning process should be as *fluid* as possible, ensuring that a novice user can quickly become proficient and confident in providing input.

Based on these considerations we created a total of six bespoke input devices for this study. We next discuss their dimensions, orientation, and input characteristics.

3.1.1 Dimensions. All of the six devices are of similar size: 5.5 centimetres (cm) in width, 9.7 cm in length, and 2.2 cm in height (respectively 2.2, 3.8, and 0.9 inches), and weigh between 60-86g. Devices are shown in Figure 2 and Figure 4. We base the size of the devices on human factors literature and studies of single-handed interactions. Karlson *et al.* examined single-handed interactions with various mobile device sizes, including a 'large' device of similar size to our design and found that size was not a factor in the input speed when tapping on a touch screen. However, they also noted that the thumb may not be able to reach all areas of the touch screen [33]. In the selection of the sensors and design of the layout of the devices, we ensured that users could easily grasp the device and that the thumb could reach and provide control over the entire input range. In a small pilot study, we recruited three participants with different hand sizes (EU glove size of 7 (small) to 10 (large)) to gain initial feedback on the physical properties of the input techniques. This helped to confirm that the overall size and shape was appropriate for various hand sizes.

In the same pilot study, we asked the participants to provide feedback on the choice and placement of buttons that could be used to confirm the input value. Two types of momentary switches were explored (push-button and micro switch), as shown in Figure 1. Participants appreciated the audible click from both switches, however, they all preferred the longer, hinged travel of the micro switch. Furthermore, participants commented that they would prefer a larger button cover in order to be able to reach the button easily regardless of their grasp. We developed a larger button cover size and standardised the button placement to the top of the device for ease of reach using the index finger. Prior work by Le et al. shows that the index finger can reach between 10–12 cm on the back of a smartphone [37], indicating that the length of the device appropriately considers the reach of the index finger. Furthermore, we ensure that our devices have two axes of symmetry – enabling the device to be used in an identical manner for left and right-handed participants. The aforementioned momentary switch button is therefore also located at the 'bottom' end of the device.

Figure 1: Various prototypes exploring the type, size, and placement of confirmation buttons. Right photo shows the final design with large button covers mounted to micro switches, fixed at the midpoint of the top and bottom edges (shown here for ANALOGUE STICK).

3.1.2 Orientation. The literature on the effect of slider orientation on input accuracy presents conflicting results. Colley et al. investigate slider-based input on touchscreens and find that a vertical orientation introduced more input error (*i.e.*, offset distortion) [12]. Stephenson & Herman assess the effect of a horizontal and a vertical Visual Analogue Scale (VAS) as used in the Short-Form McGill Pain Questionnaire (SF-MPQ) [56]. Results of their study indicate that a vertical VAS correlated better with the 'present pain intensity' items on the SF-MPQ as compared to the horizontal VAS. Price et al. also present their sliders in a vertical orientation [47]. Given these contradicting results, we follow established practices in pain reporting and make use of a smartphone-like layout with a vertical orientation for both the input mechanism and the tangible device.

3.1.3 Input range. In a 1986 landmark study on reporting pain, Jensen et al. compared six measurements of clinical pain intensity; a visual analogue scale, a 101-point numerical rating scale, an 11-point box scale, a 6-point behavioural rating scale, a 4-point verbal rating scale, and a 5-point verbal rating scale. Their results revealed that the different measures are "*more similar than they are different in terms of the rates of incorrect responding and in terms of construct validity.*" [31]. This has informed *e.g.* the decision by Adams et al. to support four levels of input for their pain self-report prototype [1]. However, Jensen et al. also state that the 101-point numerical rating scale (NRS-101) has an advantage over the other input options due to, *inter alia*, its simplicity in administrating and scoring and its extensive range, and therefore conclude that "*to the degree that a standard measure of pain intensity is needed to facilitate comparison of treatment outcome [...] it appears that the NRS-101 would be a wise choice*" [31]. Although our work is not focused solely on pain assessment, we argue that the consideration to support an extensive input range also applies to other application domains. It is, therefore, of interest to assess the ability of participants to accurately input values on such a high-resolution scale through tangible devices in order to inform future study designs. As such, we design our prototypes to allow for data input from 0 to 100.

3.2 Hardware & Software

Following the aforementioned motivations, we provide an overview of the study's six input devices in Figure 2. For each device, we include motivation from related research and the specific hardware details of the implementation.

a) **BASIC SLIDER**. Slider with a knob which can be moved across the vertical direction. Tangible sliders are widely used in tangible devices [21, 30, 34], and can be operated while the user's attention is focused elsewhere [30]. Features a clearly distinguishable start and end-point at both extreme positions, with a linear mapping of input values and total slider range of 30 mm (1.18 inch). Figure 3 shows the device and input mechanism. Total device weight is 60g, including one 3.7V LiPo battery. We built a voltage dividing circuit to accommodate the micro controller's analogue to digital converter (ADC) range of 0-1V. This served to reduce the 3V supply from the board to 1V, which was then connected to the potentiometer. An identical circuit was used on all devices.

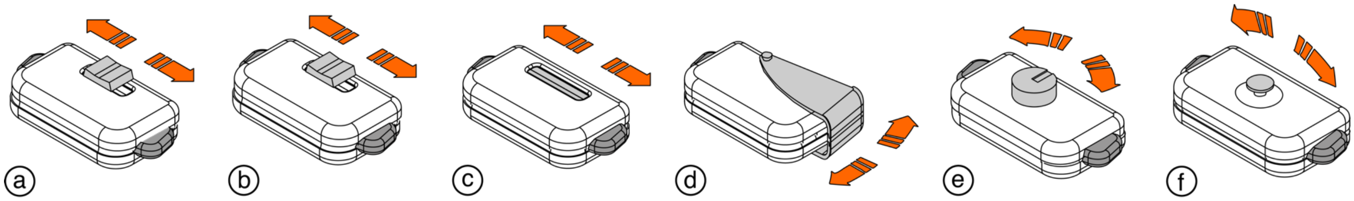

Figure 2: Isometric view of the six devices. Light grey areas indicate the moving part of the sensor, which provides the input value from the participant, orange arrows indicate the direction of movement, confirmation buttons are dark grey. a) Basic slider, b) Resistance slider, c) Capacitive touch slider, d) Pressure grip, e) Rotary knob, f) Analogue stick.

b) **Resistance slider**. Identical to the Basic slider, with the addition of a resistance mechanism which provides increasingly stronger tactile feedback as the input value is increased. Tactile feedback reduces reliance on vision in selection tasks [45], and has been applied in self-report studies [36]. The resistance gradually increases due to a polyurethane elastic band attached to the slider and an internal hook – hiding the mechanism from the participant. The force required to move the slider from the baseline position is approximately 0.69N (70g) to 2.45N (250g) at the maximum value position. The weight of the device is 65g including a 3.7V LiPo battery.

c) **Capacitive touch slider**. Capacitive touch potentiometer, which records touch-input across the vertical direction. Inspired by previous research on eyes-free touch-screen interaction, in which a plastic overlay with an opening served as a tactile guide for one dimensional finger movement [13]. In the Capacitive touch slider, the slider lies flat just beneath the surface of the enclosure, with finger access provided by a rectangular opening. The opening is designed with chamfered edges to provide a smooth and clearly distinguishable start- and end-point. The size of the opening was designed so that the centre-to-centre distance of a fingertip at each extreme would result in a total slider travel similar to the other slider-based devices. The device weighs 71g including two 3.7V LiPo batteries. The capacitive touch potentiometer [55] is a stand-alone device and requires a separate supply of 5V, thus a buck-boost converter was used to boost the voltage from one of the 3.7V batteries. The potentiometer was used in analogue mode, thus functions in the same way as the other potentiometers used in the study.

d) **Pressure grip**. Pressure-based input based on variable resistance of conductive foam, in which the input value increases with an increase in the pressure of the squeeze. Pressure-based input techniques have been explored both as a general input technique [24], as well as for the collection of self-report data [1]. Stewart et al. find that double-sided pressure input (*i.e.*, grasping or squeezing) outperforms single-sided input (*i.e.*, pushing) [57]. We make use of a double-sided pressure input, in which the device's movable part is pressed with the palm. We separated two wires by a small piece of carbon-impregnated conductive polyurethane foam [19], and integrated the composite structure into the handle of the device. The device weighs 86g including one 3.7V LiPo battery. To set the sensitivity of the device to accommodate a comfortable grasp, the device begins to register a change from the minimum value when approximately 2.11N (215g) of grasping force is applied. The value increases with the applied force until reaching the maximum value when approximately 8.83N (900g) of grasping force is applied. Our pilot test helped to identify a suitable range for the necessary grasping force.

e) **Rotary knob**. Physical knob which can be rotated to increase or decrease the input value. Knobs, also called dials, are one of the most prevalent tangible input techniques and praised for their ability to support fine adjustment [34]. The thumb is placed on the side of the knob, with the input value decreasing when the knob is rotated towards the palm and vice versa to increase the value. We affixed physical end stops under the knob, constraining the rotation to a range of 90° to ensure that the full range can be selected without lifting the thumb. Participant input is mapped linearly across this range. The device weighs 70g including a 3.7V LiPo battery.

f) **Analogue stick**. A miniaturised joystick that protrudes vertically and can be pushed with a thumb in two directions. When pushed away, the value increases and when pulled towards the palm, the value decreases. Springs return the protrusion to the centre position (*i.e.*, input value 50 out of 100). Previous work has highlighted the accuracy of the analogue stick for selecting targets and providing precise movement control [50]. Analogue sticks are common on video game controllers [17]. They are also known as a 'thumb joystick', providing linear analogue control over approximately 90°. We locked one of the axes so that the stick's movement is restricted to the axis aligned with the length of the device. The device weighs 65g including one 3.7V LiPo battery.

Characterising the types of grasp involved in typical activities, researchers have developed a repository of grasp types [20]. Research on grasp behaviours during daily human activities suggests approximately 35 grasp types, as well as insights about the unique capabilities and performance measures [52]. A main distinction can be drawn between *power* and *precision* grasps. All of the devices, with the exception of the pressure grip device, involve a precision grasp with the thumb being recruited for fine control of a sensor. The Pressure grip involves a power grasp with thumb adduction, meaning that the thumb moves toward the hand centreline in a pinching motion. As such, it is the squeeze pressure that dictates the input rather than the placement of the thumb on the device.

Figure 3: a) BASIC SLIDER, b) minimum value, c) maximum value, and d) electronic components including microcontroller, battery, potentiometer, voltage divider circuit, and micro switches.

3.2.1 Software implementation. To increase the study's ecological validity we implemented a fully wireless protocol. All devices send data via a WiFi connection, constantly updating the device's current state (input reading and confirmation button). We developed a bespoke Node.js application which continuously reads all incoming values. The application appends these readings to a .CSV file, together with a timestamp, randomly generated participant ID, and current experimental state (*i.e.*, device and target value).

3.3 Procedure

The study took place following lockdown restrictions due to the global Covid-19 pandemic, with lockdown being lifted and most people having returned to their daily jobs. We strictly followed hygiene protocols, instructing participants to wash their hands using the provided materials before and after participation, cleaning the tangible devices following each participant, and ensuring sufficient ventilation. Upon accepting to participate in the study and following the aforementioned hygiene procedure, participants were explained the goal of the study. Seated at a computer running the study software, participants were asked to provide their demographic information as well as their dominant hand. We recruited a total of 20 participants for this study. Our study sample consisted of 14 women and 6 men, with an average age of 27.4 (SD = 3.1). A majority of 18 participants are right-handed, one is left-handed, and one is ambidextrous (preferred to hold the device in their right hand). Handedness of our study sample is approximately in line with population averages [40].

The aforementioned application selected each of the six tangible devices in random order for assessment. Participants were asked to pick up the device, visually inspect it, and explore the mapping of the tangible input with the recorded value. During this exploratory stage, the application displayed a 'live' input value on the screen – updating continuously in accordance with the participant's input. After the participant was ready to proceed, we instructed the participant to place the device in their 'input-hand' and keep it under the table, preventing direct line-of-sight to the device (see Figure 4). This ensured that the data collection represented a real-world usage scenario in which participants provide discreet input. The application randomly selected ten values ranging between 0 and 100, which the participant aimed to input as accurately at possible. No direct feedback was given to the participant, *i.e.* they cannot see the actual value selected. The application automatically proceeded to the next number after the participant confirmed their input. After the participant provided input for all ten values, we assessed the participant's perceived workload using the raw NASA-TLX questionnaire (6 items) [25, 26] and offered participants the opportunity to provide free-text feedback regarding the device. Subsequently, the application randomly selected the next tangible device from the set of remaining devices. This process was repeated until all devices had been completed. The study was finalised with a semi-structured interview in which the participant's considerations of using the tangible in their everyday life were considered.

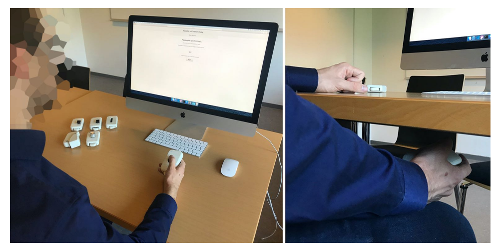

Figure 4: Left: Study setup with participant seated in front of a computer, testing the device while watching the value update on the screen. Right: During the study tasks, the participant holds the device under the table to prevent visual feedback.

4 RESULTS

All of our twenty participants managed to complete the tasks across all devices. Covering six devices, ten tasks per device, and a total of twenty participants, we collected a total of 1200 completed tasks. From the results, we identified 187 tasks which had a completion time of 0 seconds. To ensure the reliability of our analysis, we removed these 187 tasks from further analysis (15.6% of data points) and re-validated the devices. Missing data points were distributed nearly equally between devices, with most tasks missing from ANALOGUE STICK (40 tasks) and least tasks missing from ROTARY KNOB (21), with an average of 31.2 missing tasks per device. We identified the error to be the result of dropped packages between the self-report device and the WiFi router – our validation did not

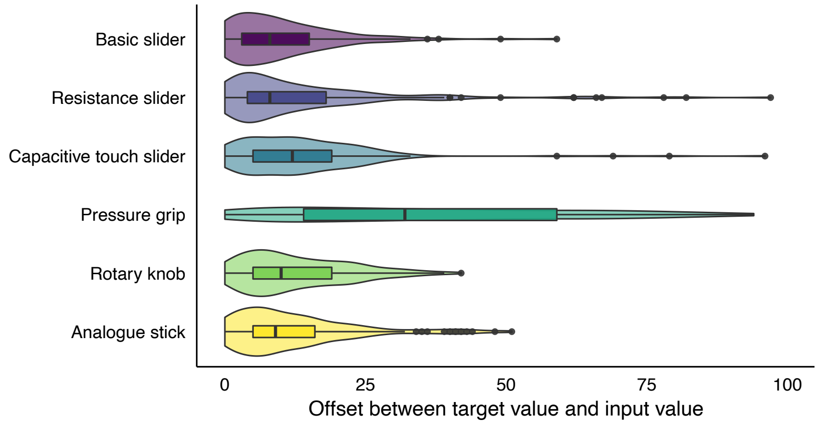

Figure 5: Combined box plot and violin plot highlighting the distribution of the offset between the participants' input value and the requested target value across devices, *i.e.* absolute error values (lower is better).

reveal any other anomalies in the data collection procedure, and we are therefore confident in the reliability of the remaining 1013 data points. We now summarise the results of our study.

We first assess the difference in input time between the six devices. Table 1 shows the mean entry times per device, with the Pressure grip and Rotary knob device resulting in the longest entry times. Given the within-subject nature of the study (*i.e.*, repeated measures) and the non-normality of the data, we use a Friedman test to assess differences. This shows no statistically significant difference in entry times between devices $\chi^2(5) = 9.91$, $p = 0.08$.

	Entry time	**Offset**
Basic slider	9.47 (SD = 14.8)	10.20 (SD = 9.6)
Resistance slider	9.57 (SD = 18.4)	14.10 (SD = 16.4)
Capacitive touch slider	8.40 (SD = 7.1)	13.90 (SD = 12.7)
Pressure grip	12.10 (SD = 25.4)	36.10 (SD = 25.9)
Rotary knob	10.70 (SD = 14.7)	12.80 (SD = 9.3)
Analogue stick	7.80 (SD = 7.9)	12.20 (SD = 10.8)

Table 1: Overview of mean self-report entry time (in seconds) and offset between target and input value (lower is better) per device.

4.1 Input Accuracy

For each self-report entry, we calculated the accuracy offset score by taking the absolute difference between the participant's input value and the target value, *i.e. Offset score* = |*input value - target value*|. The offset score is therefore always positive, with a score of 0 indicating a completely accurate input. We report the mean offset score per device in Table 1, and visualise the distribution of these scores in Figure 5. A Friedman test revealed a statistically significant difference in participant accuracy depending on which input device was used, $\chi^2(5) = 45.75$, $p < 0.001$. Using the Conover pairwise *post-hoc* tests (with Benjamini–Hochberg procedure correction), we find that participant offset is significantly higher (*i.e.*, worse) for the Pressure grip as compared to all other devices (all $p < 0.001$). Further, we find that the Basic slider has a significantly lower offset as compared to all other devices except for the Analogue stick ($p < 0.001$), and that the Analogue stick has a significantly lower offset as compared to the Resistance slider, Capacitive touch slider, and Rotary knob (all $p < 0.05$). A number of trials show extreme outliers (Figure 5), indicating that participants may have erroneously inverted the input axis in their mind on a limited number of trials. Interestingly, these outliers, while limited in occurrence, seem to have occurred solely for the 'slider' devices.

Next, we investigated whether offset in participant accuracy differs in specific input ranges. In Figure 6, we visualise the mean offset in participant accuracy between devices, as grouped in buckets of 10 over the entire input range. The figure highlights a number of interesting differences between devices. The least accurate self-report device, Pressure grip, shows a large offset in input range above 40 – indicating that the large offset for this device is primarily due to inaccuracy when providing higher levels of pressure (mean absolute offset of 36.1). For the Resistance slider (mean offset 14.1), Capacitive touch slider (mean offset 13.9), and Rotary knob (mean offset 12.8), we see a clear pattern that indicates participant input is too high on the lower target values and too low for the higher values. Input for the Basic slider (mean offset 10.2) and Analogue stick (mean offset 12.2) is most consistent over the target range.

4.2 Workload & Preferences

We further evaluated the six input devices according to participants' perceived workload and preferences. We used the NASA-TLX questionnaire [25, 26] and collected user comments following each device trial.

We visualise the self-reported workload of operating each of the six devices in Figure 7. In line with the recorded accuracy levels, we notice that physical demand, required effort, and frustration are clearly higher for the Pressure grip as compared to the other devices. Furthermore, we find that the perceived physical demand, effort, and frustration is also high for the Resistance slider and Analogue stick. Both these devices required the participant to provide continuous pressure to the device while simultaneously

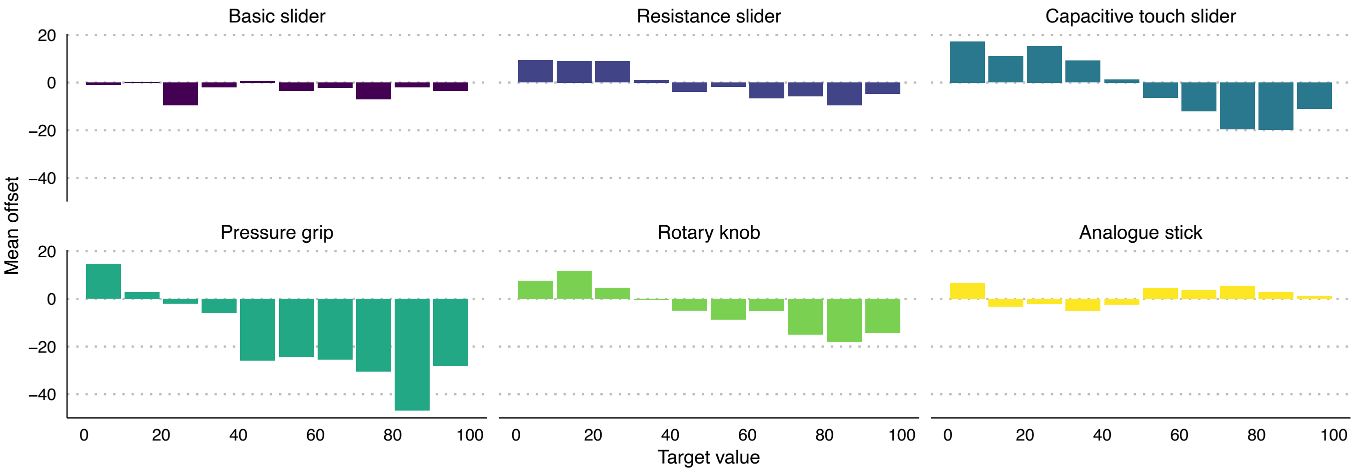

Figure 6: Overview of difference between participant input value and target value across the input range (0–100), grouped in bins of 10.

confirming the input using a different finger, explaining their higher reported load on these NASA-TLX dimensions.

At the end of the study, participants ranked the six devices in order of preference. Devices were rated from 1 to 6, and identical rankings were not allowed – forcing participants to consider their preference between all devices. We summarise the ranking the devices in Figure 8, highlighting the popularity of the Basic slider and Capacitive touch slider as a first-choice option. Furthermore, the ranking shows that the Rotary knob is frequently selected as a second choice, and that the Pressure grip is predominantly considered as the least favourite device among our sample. Rankings for the Resistance slider indicate a disagreement between participants.

4.2.1 User Preferences and Comments. Following each device trial, participants were encouraged to provide as much detail as possible about what they felt could be improved with the device. A total of 17 of the 20 participants provided comments after the device trials, while five participants provided additional comments at the end of the study. Following a process of emergent thematic coding, carried out collaboratively by two of the paper's authors, we categorised participant comments in the themes of *accuracy* and *ergonomics.*

Accuracy. Various comments focused on device accuracy, including characteristics of the devices and strategies for achieving a higher accuracy. Device characteristics most often focused on the haptic feedback inherent to the device. P11 appreciated that the Analogue stick provided a sense for the midpoint, stating, "*I liked the default position of 50 meaning that I had a clear idea of where the top half and bottom half of the numbers would be.*" The midpoint was also mentioned in relation to the Resistance slider. P10 claimed, "*I had a model in my mind that when I suddenly felt resistance, I was halfway.*" Some participants were less favourable, yet also mentioned the perception of the midpoint. P17 noted, "*I liked the resistance however it was still difficult to be accurate with this one without a middle point.*" P17 also missed having a midpoint with the Rotary knob, "*I feel like I was able to be more accurate with this device, however I was missing a little feedback for the 50 (middle point).*"

Various participants explained the strategies they adopted trying to increase their accuracy in the task. The most common strategy involved exploring the limits of the device by moving the sensor between the minimum and maximum values in the range, then subsequently selecting the target value. Another commonly recurring strategy involved relying on the device preserving the previous value and then adjusting to achieve the new target value. P06 stated that the Basic slider, "*...stayed in place, I estimated more relative to my last position.*" P17 expressed similar considerations about

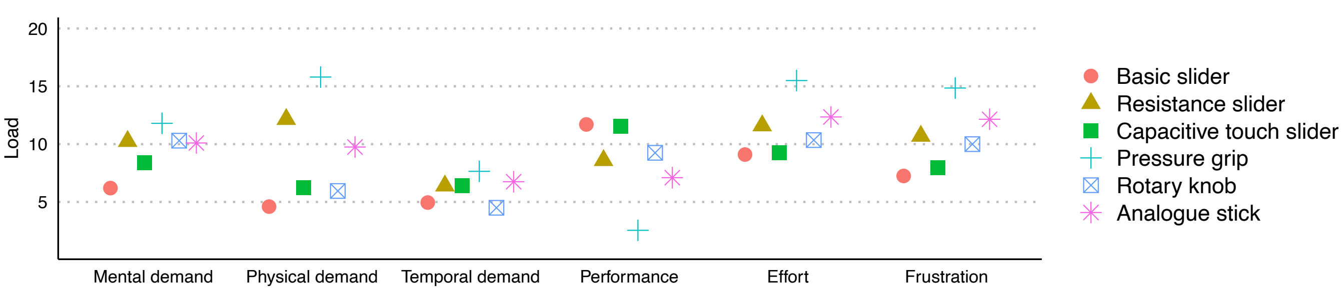

Figure 7: Overview of NASA-TLX scores for each of the six devices across the six dimensions of the NASA-TLX.

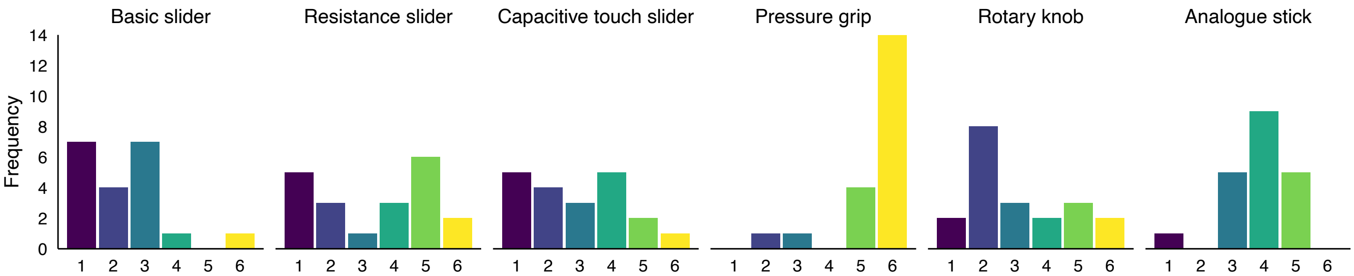

Figure 8: Distribution of participants' ranking of the six devices, a rank of 1 is best.

the Basic slider, "*...I had to find 0 and 100 scale before each input, however I believe that this move can be learned with time.*" Likewise, P02 commented regarding the Capacitive touch slider, "*I developed a technique along the way for figuring out where my finger was placed on the scale.*" P12 gave an indication that they struggled in choosing a strategy when using the Resistance slider stating, "*...I tried different methods of going all the way down or just trying to adjust based on the previous 'judgement', but I never felt secure in my choice.*"

Ergonomics. Several participants provided comments related to the ergonomics of the devices. A few of them mentioned that the Capacitive touch slider felt comfortable to use, for example, P08 appreciated the tactile guidance offered by the window on the Capacitive touch slider, which made it "*Easier to physically feel where the numbers were because my finger could feel the whole scope.*" Some of the devices received a mix of favourable as well as critical comments. For example, P04 noted about the Pressure grip, "*This device felt like I had the most control, as I was able to use my entire hand to both hold and use the device.*" Other participants, however, claimed that the ergonomics of the Pressure grip device could be improved, yet did not provide specific suggestions.

Most of the suggestions for ergonomic improvements focused on the confirmation button, with participants stressing the challenge in simultaneously trying to match the target value and confirming their input. Additional comments focused on refinements to the sensor range and placement. P19 suggested that the Rotary knob could have longer travel in the movement, stating "*If you made the range of motion slightly larger it would be easier to get more accurate values.*" P03 suggested that the Analogue stick would have been, "*[...] more ergonomic from left to right*".

5 DISCUSSION

Despite the wide use of self-report methods by HCI researchers [64], the literature also recognises a number of challenges regarding the accuracy of self-report data. Participants may, for example, aim to give socially acceptable answers [14] or lose motivation to answer accurately over the duration of a study [35, 58]. The increased use of discreet tangible self-report devices introduces another concern atypical of pen-and-paper or smartphone-based self-reports; the possibility for a discrepancy between the intended input and the recorded input. Our results highlight significant differences between the assessed input techniques, with the Pressure grip device resulting in the lowest accuracy and the Basic slider outperforming all devices except for the Analogue stick. Furthermore, we identify similar patterns in offset across multiple devices – indicating a common error among participants to overestimate the required input on lower target values and underestimate the required input on higher target values. Lastly, we compared input accuracy across a number of input ranges and show that, in line with our expectations, participant input is more accurate on less granular scales. Overall, our results present evidence for the fact that, depending on the desired data granularity, input discrepancy can be a serious threat to the reliability of self-report data as collected through tangible input devices.

Before discussing our results and implications in more detail, we wish to highlight that our design criteria – discreet input, one-handed usage, and consistency between the appearance of devices – have directly influenced both the design and operation of the presented devices. We base our motivation of discreet input and one-handed usage on work detailing previous experiences of study participants [1, 51]. Our decision to maintain a consistent form factor between devices follows from the motivation to reduce form factor bias between input techniques. We note that, as a consequence of this design criteria, the presented devices were optimised for comparability rather than their respective input technique. As such, the input devices could be further optimised to support a higher level of usability for the specific input type. For example, the requirement to confirm input with the index finger turned out to be more ergonomic on the devices that did not return the input value to zero upon release (*i.e.*, Basic slider and Rotary knob) as participants were not required to maintain tension. As, to the best of our knowledge, no prior study has evaluated the accuracy of tangible self-report devices, we consider this a viable approach to the validation of scientific instruments already used in the wild (*e.g.*, Keppi [1]), and similar to Hernandez comparison of smartwatches, smartphones, and Google Glass [28]. In order to support the development and assessment of future self-report devices, we provide both the software and hardware blueprints as supplementary material to this paper[1].

5.1 Wearable & Tangible Input Devices

Researchers employing the Experience Sampling Method (ESM), originally asking participants to complete responses on pen-and-paper following an incoming beeper alert [35], have been quick

[1]Please see the supplementary materials.

to adapt the method to make use of novel technologies as they became available, including the PDA and the smartphone. Lately, researchers have looked at the use of wearable and tangible input devices as alternative input methods. Although smartphones are frequently used throughout the day, receiving a questionnaire notification prompt on a smartphone may be perceived as disruptive and requires that the participant unlocks their phone and launches a specific application. This takes additional time away from the participant and may reduce overall response rates [28]. Furthermore, participants may find it difficult to enter data discreetly [1, 51]. Despite these advantages, the use of these alternative input devices also raises new challenges. In a comparison of wearable input devices, for example, Hernandez *et al.* highlighted that "*the limited Watch screen size may potentially influence the use of the full-range of some kinds of response scales*" [28].

Given the importance of accurately capturing the participants' intended self-report values, we studied input accuracy across a range of different input techniques. We stress that while our contribution focuses strongly on the use of tangibles for collecting numeric participant input, previous work has highlighted the importance of considering the affective dimension of self-report devices [2, 18, 23]. For example, Duong *et al.* made use of plush toys to increase the suitability of the self-report devices for hospitalised children. These considerations are out of scope for this work.

5.1.1 Perception of Tactile Feedback. Material choices and physical techniques have been explored to provide additional explicit feedback, *e.g.* flexible devices which provide feedback through the act of deformation [9]. Participant experiences in their use of the six self-report devices revealed wide differences among participants in regards to tactile qualities of the physical form, movement, and resistance. Previous research suggests that eyes-free interaction provides implicit closed-loop control, yet additional haptic feedback can enhance the sense of control and accuracy [5, 43]. In this study we introduced straightforward material-based control resistance feedback in the form of kinematic constraints (*e.g.*, constrained slider movement in a linear path of 30mm) and material based passive haptic feedback (*e.g.*, increase in elastic resistance felt with the Resistance slider), compression of the foam in the Pressure grip device, and the spring that returns the Analogue stick to the centre position).

Through the comments of our participants, we identified that they developed a number of strategies to improve their accuracy based on these material-based feedback mechanisms. The most widely applied strategy consisted of moving the input mechanism between the lowest and highest point to obtain a sense of scale, and subsequently adjusting the input value while using the lowest or highest point as a baseline (*e.g.*, trying to 'subtract' 11 from 100). Similarly, participants commented positively on the additional feedback point provided by the Analogue stick (midpoint). This highlights the value of clear start- and end-points, something that was not provided by the Pressure grip. Surprisingly, participants also described their use of material properties as a feedback mechanism in ways we did not originally envision. For example, the elastic resistance provided by the Resistance slider was perceived by a few of our participants as providing a midpoint indication. In reality, however, the resistance was smooth, linear, and did not expressly indicate a midpoint. A lack of sufficient feedback mechanisms may, therefore, result in participants incorrectly assuming that they identified a reliable support mechanism – indicating the presence of a conceptual discrepancy among users [7].

5.1.2 Tangible Self-Report in the Wild. The presented research contributes to the use of tangible self-report devices. While we focused on the assessment of participant performance in the lab, our work led us to reflect on a number of open questions for applying this technology *in situ*. First, the portability of self-report devices is critical when participants have to carry the device on them throughout their daily life. Previous work has therefore explored the use of wearables [1, 28] or provided multiple situated self-report devices placed strategically throughout the participant's house [44]. Our devices were designed with a limited degree of 'pocketability' in mind – allowing the device to be stored inside a handbag or the pocket of a sweater.

Second, collecting accidental input is a serious concern when dealing with 'always on' tangible devices. While the ability to provide immediate input, without *e.g.* opening an application on a smartphone, lowers the barrier for input, it also increases the possibility for accidental data recordings. Prior work has considered the use of a mechanical slider to easily lock/unlock the possibility of data collection [70].

Finally, any mobile technology will face challenges related to battery life. Our devices were equipped with 3.7V batteries. These batteries provided a limited battery life, were unable to report their remaining battery value, and had to be recharged during the study. Previous work by Vega *et al.* estimated that their battery would suffice for four weeks, sufficient to support the median study duration of two weeks in experience sampling studies [64]. However, the physical dimensions of their battery case (57.75 cm^3) would increase the difficulty of fitting the battery within a tangible device. The ability for participants to recharge devices using a standard USB connector would, therefore, most likely be a requirement for longitudinal deployments.

5.2 Implications for Tangible Self-Report Research

Based on our gathered insights, we outline a number of implications for future work on tangible self-report studies. We stress that our study was limited to a lab evaluation in order to systematically study input accuracy across a wide range of input techniques. However, a further assessment of our implications in situ are required.

- **Haptic feedback to support kinesthetic sense**. Based on the participants' comments, we believe that feedback mechanisms – such as the 'middle point' provided by the neutral position of the Analogue stick – are useful to participants in assessing the input range. Future work may therefore explore additional feedback mechanisms and modalities, *e.g.* through vibrotactile feedback, force feedback, or non-speech audio to augment the kinesthetic sense [43]. Such feedback mechanisms have been considered in other contexts such as in-car interactions, in which users do not always have a direct line of sight to the input controls [42].

- **Precision / power grip trade-off**. Our results clearly identify the Pressure grip as being the least accurate of all self-report devices. While prior work, *e.g.*, Keppi [1], have made use of pressure-based input at a lower resolution, our results highlight that participants were unable to produce reliable input at higher target values (*i.e.*, above 40 – see Figure 6). The Pressure grip device required users to provide input through a so-called 'power grip,' in which a person provides pressure between the palm of the hand and the fingers. Clarkson describes the power grip as being high force and low precision [11], which explains participant inaccuracy as the required pressure increases. While our results point to the inaccuracy of the power grip, squeeze input based on manipulation between the fingers (*e.g.*, thumb and index finger) is potentially able to provide a higher level of input accuracy. It is worthwhile to note that some participant populations may have less fine motor control, *e.g.* stroke patients may not be able to control precise movements, yet could still provide some input through a power grip.
- **Control resistance / confirmation**. While we originally believed that adding resistance to the device (Resistance slider and Pressure grip) would provide additional feedback to participants and act as a 'spring return' [46], our results highlight that this did not support participants in providing more accurate input. Participants' comments highlighted that some found it difficult to ensure that their input remained stationary while simultaneously pressing a separate confirmation button with their index finger. Another option can be to employ input techniques that make use of control resistance that do not require participants to 'hold' the position but remain in place while the user confirms their input (as used in the Basic slider and Rotary knob).
- **Match device features to individual preferences**. The resistance and 'return spring' effect present in the Analogue stick, Resistance slider and Pressure grip received mixed responses from the participants. Some participants found this passive haptic feedback helpful and enjoyable, while others expressed dislike and frustration. Similarly, comments varied in terms of support for the position of the confirmation buttons, even though in the pilot study participants with a wide range of hand sizes supported the placement of the final design. Therefore, we suggest that individual elements of tangible self-report devices, such as the placement of a confirmation button, should be customisable to participants' preferences. Ignoring user preferences or providing undesirable features has been shown to lead to rejection by the user [32]. While outside of the scope of our current study, we hypothesise that a mismatch between participant preference and the actual design of a tangible self-report device may result in reduced accuracy in self-report data.
- **Systematic input offset**. Finally, we note that our analysis shows a similar pattern of offset among a number of the study's devices; on lower target values, participants typically provide a higher input value, whereas on higher target values they provide a lower input value (Figure 6). While this offset pattern requires further investigation across other studies, a systematic offset could be compensated for with a straightforward adjustment to the linear assessment. Offset correction has previously proven useful in improving pointing accuracy for projected targets in real and virtual environments [39], as well as correcting smartphone input [6].

5.2.1 Future work. Our findings also highlight opportunities for future work. Recent research has explored digital control of explicit haptic feedback to enhance accuracy and improve the experience of interactions. Adding vibrotactile cues have been shown to increase accuracy in handle position in a sports setting [54]. Work on flexible smartphones explored the uses of a haptuator to provide additional haptic and auditory feedback beyond the material properties of the device to increase user control [59]. Such feedback elements can be incorporated without requiring the user to hold on to the input mechanism while confirming the input. Virtual springs and detents have been explored to enhance grasp tasks [22], and to support users to more quickly navigate in CAD environments using a tangible controller [27]. Incorporating such active haptic feedback signals could potentially increase participant input accuracy while further supporting the goal of discrete input.

5.3 Limitations

We identify a number of limitations in our work. First, we presented a limited number and variety of devices to participants, whereas an almost endless combination of input types could be considered. We base our selection on the related work on tangible (self-report) devices [1, 4, 43], our three design considerations (can be operated without line of sight, with one-hand, and a consistent appearance between devices), and availability of materials, in order to cover an essential set of input types. Second, the devices used (see Figures 3 and 4) are less polished than some of the devices presented in earlier work – which typically focused on a single device. As the visual appearance is highly similar between devices, we do not expect this to have had a notable impact on the accuracy of results. Third, our work does not consider uncertainty-based input, as discussed *inter alia* by Greis et al. [21]. While this type of input is not very common, a number of application areas would benefit from participants being able to submit a self-report that allows for uncertainty. Finally, given the lab-based nature of our study, we are unable to comment on the impact of long-term deployments on input discrepancy. As tangible self-report devices typically do not provide any direct feedback as to the data that was entered, it is impossible for participants to learn to adjust their input during a longitudinal deployment. We, therefore, expect any discrepancies between the intended and recorded input to maintain or aggravate over the study duration. Future work should explore the input accuracy of tangible self-report devices in the wild.

6 CONCLUSION

In this paper, we study the accuracy of participants in operating tangible self-report devices across a range of input techniques. Our work is motivated by the increasing use of tangible devices in self-report studies, following the need for participants to enter their self-report data more discreetly and without the distractions a smartphone may provide. We designed six devices to explore a range of input techniques. The ability of participants to use these devices to match a randomised number was measured through a lab study, in which participants were asked to input numbers while holding the device outside of their line of sight. Our results reveal that participant accuracy was highest for the Basic slider and lowest for the Pressure grip. We furthermore find that participants input values are too high on the lower end of the input spectrum, whereas their input value is typically too low towards the higher end of the input spectrum. Based on participant feedback and our quantitative analysis, we identify implications for future tangible self-report research concerning control resistance, haptic feedback, and input technique. Our work highlights the importance of evaluating our scientific instruments before deploying them in the wild. By publicly releasing both the software and hardware blueprints of our devices, we hope to encourage future research in this growing area.

REFERENCES

[1] Alexander T. Adams, Elizabeth L. Murnane, Phil Adams, Michael Elfenbein, Pamara F. Chang, Shruti Sannon, Geri Gay, and Tanzeem Choudhury. 2018. Keppi: A Tangible User Interface for Self-Reporting Pain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Article 502, 13 pages. https://doi.org/10.1145/3173574.3174076

[2] Madeline Balaam, Geraldine Fitzpatrick, Judith Good, and Rosemary Luckin. 2010. Exploring Affective Technologies for the Classroom with the Subtle Stone. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, 1623–1632. https://doi.org/10.1145/1753326.1753568

[3] Allan Berrocal, Waldo Concepcion, Stefano De Dominicis, and Katarzyna Wac. 2020. Complementing Human Behavior Assessment By Leveraging Personal Ubiquitous Devices and Social Links: Evaluation of the PeerMA Method. *JMIR mHealth and uHealth* (2020), in press. https://doi.org/10.2196/15947

[4] Alberto Boem and Giovanni Maria Troiano. 2019. Non-Rigid HCI: A Review of Deformable Interfaces and Input. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. 885–906. https://doi.org/10.1145/3322276.3322347

[5] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal 'eyes-Free' Interaction Techniques for Wearable Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. 473–480. https://doi.org/10.1145/642611.642694

[6] Daniel Buschek and Florian Alt. 2015. TouchML: A Machine Learning Toolkit for Modelling Spatial Touch Targeting Behaviour. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. 110–114. https://doi.org/10.1145/2678025.2701381

[7] John M. Carroll and Judith Reitman Olson. 1988. Mental Models in Human-Computer Interaction. In *Handbook of Human-Computer Interaction*, Martin Helander (Ed.). North-Holland, Amsterdam, 45–65. https://doi.org/10.1016/B978-0-444-70536-5.50007-5

[8] Scott Carter and Jennifer Mankoff. 2005. When Participants Do the Capturing: The Role of Media in Diary Studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. 899–908. https://doi.org/10.1145/1054972.1055098

[9] Victor Cheung, Alex Keith Eady, and Audrey Girouard. 2017. Exploring Eyes-Free Interaction with Wrist-Worn Deformable Materials. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17)*. 521–528. https://doi.org/10.1145/3024969.3025087

[10] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: Low-Burden, Self-Monitoring Technology for Capturing and Reflecting on Sleep Behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, 121–132. https://doi.org/10.1145/2750858.2804266

[11] John Clarkson. 2008. Human Capability and Product Design. In *Product Experience*, Hendrik N.J. Schifferstein and Paul Hekkert (Eds.). Elsevier, San Diego, 165–198. https://doi.org/10.1016/B978-008045089-6.50009-5

[12] Ashley Colley, Sven Mayer, and Niels Henze. 2019. Investigating the Effect of Orientation and Visual Style on Touchscreen Slider Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, 1–9. https://doi.org/10.1145/3290605.3300419

[13] Ashley Colley, Lasse Virtanen, Timo Ojala, and Jonna Häkkilä. 2016. Guided Touch Screen: Enhanced Eyes-Free Interaction. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*. 80–86. https://doi.org/10.1145/2914920.2915008

[14] Ross Corkrey and Lynne Parkinson. 2002. A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions. *Behavior Research Methods, Instruments, & Computers* 34, 3 (01 Aug. 2002), 354–363. https://doi.org/10.3758/BF03195463

[15] Christian Corsten, Simon Voelker, Andreas Link, and Jan Borchers. 2018. Use the Force Picker, Luke: Space-Efficient Value Input on Force-Sensitive Mobile Touchscreens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174235

[16] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and Reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, Mihaly Csikszentmihalyi (Ed.). Springer Netherlands, Dordrecht, 35–54. https://doi.org/10.1007/978-94-017-9088-8_3

[17] Alastair H Cummings. 2007. The evolution of game controllers and control schemes and their effect on their games. In *The 17th annual university of southampton multimedia systems conference*, Vol. 21.

[18] Tu Dinh Duong, Ann Blandford, Yvonne Rogers, and Neil Sebire. 2019. Beyond The Smartphone: The Assessment And Reflection Of Wellbeing For Children. In *Adjunct Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (WISH - Workgroup on Interactive Systems in Health)*.

[19] ElfaDistrelec. 2020. RND 60000069 Low Density Conductive Foam, 305 x 305mm RND Lab. https://www.elfadistrelec.dk/da/low-density-conductive-foam-305-305mm-rnd-lab-rnd-600-00069/p/30130230

[20] T. Feix, J. Romero, H. Schmiedmayer, A. M. Dollar, and D. Kragic. 2016. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems* 46, 1 (2016), 66–77.

[21] Miriam Greis, Hyunyoung Kim, Andreas Korge, Céline Coutrix, and Albrecht Schmidt. 2019. Extending Input Space of Tangible Dials and Sliders for Uncertain Input. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '19)*. Association for Computing Machinery, 189–196. https://doi.org/10.1145/3294109.3300985

[22] Sidhant Gupta, Tim Campbell, Jeffrey R. Hightower, and Shwetak N. Patel. 2010. SqueezeBlock: Using Virtual Springs in Mobile Devices for Eyes-Free Interaction. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. 101–104. https://doi.org/10.1145/1866029.1866046

[23] Frode Guribye, Tor Gjøsæter, and Christian Bjartli. 2016. Designing for Tangible Affective Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. Article 30, 10 pages. https://doi.org/10.1145/2971485.2971547

[24] Chris Harrison and Scott Hudson. 2012. Using Shear as a Supplemental Two-Dimensional Input Channel for Rich Touchscreen Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, 3149–3152. https://doi.org/10.1145/2207676.2208730

[25] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[26] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[27] Nikolaj Haulrik, Rasmus M. Petersen, and Timothy Merritt. 2017. CADLens: Haptic Feedback for Navigating in 3D Environments. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems (DIS '17 Companion)*. 127–131. https://doi.org/10.1145/3064857.3079132

[28] Javier Hernandez, Daniel McDuff, Christian Infante, Pattie Maes, Karen Quigley, and Rosalind Picard. 2016. Wearable ESM: Differences in the Experience Sampling Method across Wearable Devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. Association for Computing Machinery, 195–205. https://doi.org/10.1145/2935334.2935340

[29] Stephen Intille, Caitlin Haynes, Dharam Maniar, Aditya Ponnada, and Justin Manjourides. 2016. μEMA: Microinteraction-Based Ecological Momentary Assessment (EMA) Using a Smartwatch. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. 1124–1128. https://doi.org/10.1145/2971648.2971717

[30] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2012. Tangible Remote Controllers for Wall-Size Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, 2865–2874. https://doi.org/10.1145/2207676.2208691

[31] Mark P. Jensen, Paul Karoly, and Sanford Braver. 1986. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 27, 1 (1986), 117–126. https://doi.org/10.1016/0304-3959(86)90228-9

[32] Myounghoon Jeon, Thomas M. Gable, Benjamin K. Davison, Michael A. Nees, Jeff Wilson, and Bruce N. Walker. 2015. Menu Navigation With In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload. *International Journal of Human-Computer Interaction* 31, 1 (2015), 1–16. https://doi.org/10.1080/10447318.2014.925774

[33] Amy K Karlson, Benjamin B Bederson, and J Contreras-Vidal. 2006. Understanding single-handed mobile device interaction. *Handbook of research on user interface design and evaluation for mobile technology* 1 (2006), 86–101.
[34] Hyunyoung Kim, Céline Coutrix, and Anne Roudaut. 2018. KnobSlider: Design of a Shape-Changing UI for Parameter Control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3173574.3173913
[35] Reed Larson and Mihaly Csikszentmihalyi. 2014. *The Experience Sampling Method*. Springer Netherlands, Dordrecht, 21–34. https://doi.org/10.1007/978-94-017-9088-8_2
[36] G. Laurans, P. M. A. Desmet, and P. Hekkert. 2009. The emotion slider: A self-report device for the continuous measurement of emotion. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. 1–6.
[37] Huy Viet Le, Sven Mayer, Patrick Bader, and Niels Henze. 2018. Fingers' Range and Comfortable Area for One-Handed Smartphone Interaction Beyond the Touchscreen. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, 1–12. https://doi.org/10.1145/3173574.3173605
[38] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, 5421–5432. https://doi.org/10.1145/2858036.2858063
[39] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. 1–13. https://doi.org/10.1145/3173574.3174227
[40] Chris McManus. 2002. *Right hand, left hand: The origins of asymmetry in brains, bodies, atoms and cultures*. Harvard University Press.
[41] Geoffrey Miller. 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science* 7, 3 (2012), 221–237. https://doi.org/10.1177/1745691612441215
[42] Alexander Ng and Stephen Brewster. 2017. An Evaluation of Touch and Pressure-Based Scrolling and Haptic Feedback for In-Car Touchscreens. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg, Germany) *(AutomotiveUI '17)*. 11–20. https://doi.org/10.1145/3122986.3122997
[43] Ian Oakley and Jun-Seok Park. 2007. Designing Eyes-Free Interaction. In *Proceedings of the 2nd International Conference on Haptic and Audio Interaction Design (HAID'07)*. Springer-Verlag, Berlin, Heidelberg, 121–132.
[44] Gaurav Paruthi, Shriti Raj, Seungjoo Baek, Chuyao Wang, Chuan-che Huang, Yung-Ju Chang, and Mark W. Newman. 2018. Heed: Exploring the Design of Situated Self-Reporting Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 132 (Sept. 2018), 21 pages. https://doi.org/10.1145/3264942
[45] Jerome Pasquero and Vincent Hayward. 2011. Tactile Feedback Can Assist Vision during Mobile Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, 3277–3280. https://doi.org/10.1145/1978942.1979427
[46] Inc Precision Sales. [n.d.]. J3R "Rocker" Potentiometer. http://www.precisionsales.com/joystick/J3r-rocker-pot.htm
[47] Blaine A. Price, Ryan Kelly, Vikram Mehta, Ciaran McCormick, Hanad Ahmed, and Oliver Pearce. 2018. Feel My Pain: Design and Evaluation of Painpad, a Tangible Device for Supporting Inpatient Self-Logging of Pain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3173574.3173743
[48] Mashfiqui Rabbi, Katherine Li, H. Yanna Yan, Kelly Hall, Predrag Klasnja, and Susan Murphy. 2019. ReVibe: A Context-Assisted Evening Recall Approach to Improve Self-Report Adherence. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 149 (Dec. 2019), 27 pages. https://doi.org/10.1145/3369806
[49] Mika Raento, Antti Oulasvirta, and Nathan Eagle. 2009. Smartphones: An Emerging Tool for Social Scientists. *Sociological Methods & Research* 37, 3 (2009), 426–454. https://doi.org/10.1177/0049124108330005
[50] Adrian Ramcharitar and Robert J. Teather. 2017. A Fitts' Law Evaluation of Video Game Controllers: Thumbstick, Touchpad and Gyrosensor. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. 2860–2866. https://doi.org/10.1145/3027063.3053213
[51] John Rooksby, Alistair Morrison, and Dave Murray-Rust. 2019. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. 1–14. https://doi.org/10.1145/3290605.3300655
[52] Artur Saudabayev, Zhanibek Rysbek, Raykhan Khassenova, and Huseyin Atakan Varol. 2018. Human grasping database for activities of daily living with depth, color and kinematic data streams. *Scientific Data* 5, 1 (May 2018), 180101. https://doi.org/10.1038/sdata.2018.101
[53] Juergen Sauer and Andreas Sonderegger. 2009. The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics* 40, 4 (July 2009), 670–677. https://doi.org/10.1016/j.apergo.2008.06.006
[54] André Schmidt, Mads Kleemann, Timothy Merritt, and Ted Selker. 2015. Tactile Communication in Extreme Contexts: Exploring the Design Space Through Kiteboarding. In *Human-Computer Interaction ‚Äì INTERACT 2015 (Lecture Notes in Computer Science)*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 37–54. https://doi.org/10.1007/978-3-319-22723-8_4
[55] SparkFun. [n.d.]. Touch Potentiometer - PRT-13144. https://www.sparkfun.com/products/13144
[56] Nancy L. Stephenson and JoAnne Herman. 2000. Pain measurement: A comparison using horizontal and vertical visual analogue scales. *Applied Nursing Research* 13, 3 (2000), 157–158. https://doi.org/10.1053/apnr.2000.7658
[57] Craig Stewart, Michael Rohs, Sven Kratz, and Georg Essl. 2010. Characteristics of Pressure-Based Input for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, 801–810. https://doi.org/10.1145/1753326.1753444
[58] Arthur A. Stone, Ronald C. Kessler, and Jennifer A. Haythomthwatte. 1991. Measuring Daily Events and Experiences: Decisions for the Researcher. *Journal of Personality* 59, 3 (1991), 575–607. https://doi.org/10.1111/j.1467-6494.1991.tb00260.x
[59] Paul Strohmeier, Jesse Burstyn, Juan Pablo Carrascal, Vincent Levesque, and Roel Vertegaal. 2016. ReFlex: A Flexible Smartphone with Active Haptic Feedback for Bend Input. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16)*. 185–192. https://doi.org/10.1145/2839462.2839494
[60] Hong Z Tan, Mandayam A Srinivasan, Brian Eberman, and Belinda Cheng. 1994. Human factors for the design of force-reflecting haptic interfaces. *Dynamic Systems and Control* 55, 1 (1994), 353–359.
[61] C. Barr Taylor, Leslie Fried, and Justin Kenardy. 1990. The use of a real-time computer diary for data acquisition and processing. *Behaviour Research and Therapy* 28, 1 (1990), 93–97. https://doi.org/10.1016/0005-7967(90)90061-M
[62] Peter Totterdell and Simon Folkard. 1992. In situ repeated measures of affect and cognitive performance facilitated by use of a hand-held computer. *Behavior Research Methods, Instruments, & Computers* 24, 4 (01 Dec. 1992), 545–553. https://doi.org/10.3758/BF03203603
[63] Khai N. Truong, Thariq Shihipar, and Daniel J. Wigdor. 2014. Slide to X: Unlocking the Potential of Smartphone Unlocking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, 3635–3644. https://doi.org/10.1145/2556288.2557044
[64] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (2017), 93:1–93:40. https://doi.org/10.1145/3123988
[65] Niels van Berkel, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2017. Gamification of Mobile Experience Sampling Improves Data Quality and Quantity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 107 (Sept. 2017), 21 pages. https://doi.org/10.1145/3130972
[66] Niels van Berkel, Jorge Goncalves, Katarzyna Wac, Simo Hosio, and Anna L. Cox. 2020. Human Accuracy in Mobile Data Collection. *International Journal of Human-Computer Studies* 137 (2020), 1–4. https://doi.org/10.1016/j.ijhcs.2020.102396
[67] Karen Vanderloock, Vero Vanden Abeele, Johan A.K. Suykens, and Luc Geurts. 2013. The Skweezee System: Enabling the Design and the Programming of Squeeze Interactions. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. 521–530. https://doi.org/10.1145/2501988.2502033
[68] Julio Vega, Samuel Couth, Ellen Poliakoff, Sonja Kotz, Matthew Sullivan, Caroline Jay, Markel Vigo, and Simon Harper. 2018. Back to Analogue: Self-Reporting for Parkinson's Disease. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Article 74, 13 pages. https://doi.org/10.1145/3173574.3173648
[69] Aku Visuri, Niels van Berkel, Chu Luo, Jorge Goncalves, Denzil Ferreira, and Vassilis Kostakos. 2017. Challenges of Quantified-Self: Encouraging Self-Reported Data Logging during Recurrent Smartphone Usage. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference (HCI '17)*. BCS Learning & Development Ltd., Article 81, 7 pages. https://doi.org/10.14236/ewic/HCI2017.81
[70] Martin Weigel and Jürgen Steimle. 2017. DeformWear: Deformation Input on Tiny Wearable Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 28 (June 2017), 23 pages. https://doi.org/10.1145/3090093
[71] Jacob O. Wobbrock, Brad A. Myers, and Htet Htet Aung. 2008. The performance of hand postures in front- and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies* 66, 12 (2008), 857–875. https://doi.org/10.1016/j.ijhcs.2008.03.004 Mobile human-computer interaction.
[72] Te-Yen Wu, Zheer Xu, Xing-Dong Yang, Steve Hodges, and Teddy Seyed. 2021. Project Tasca: Enabling Touch and Contextual Interactions with a Pocket-Based Textile Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 4, 13 pages. https://doi.org/10.1145/3411764.3445712

[73] Zheer Xu, Weihao Chen, Dongyang Zhao, Jiehui Luo, Te-Yen Wu, Jun Gong, Sicheng Yin, Jialun Zhai, and Xing-Dong Yang. 2020. BiTipText: Bimanual Eyes-Free Text Entry on a Fingertip Keyboard. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376306

[74] Xinghui Yan, Katy Madier, Sun Young Park, and Mark Newman. 2019. Towards Low-Burden In-Situ Self-Reporting: A Design Space Exploration. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion (DIS '19 Companion)*. 337–346. https://doi.org/10.1145/3301019.3323905

[75] Xinghui Yan, Shriti Raj, Bingjian Huang, Sun Young Park, and Mark W. Newman. 2020. Toward Lightweight In-Situ Self-Reporting: An Exploratory Study of Alternative Smartwatch Interface Designs in Context. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 158 (2020), 22 pages. https://doi.org/10.1145/3432212

[76] Eunhye Youn, Sangyoon Lee, Sunbum Kim, Youngbo Aram Shim, Liwei Chan, and Geehyuk Lee. 2021. WristDial: An Eyes-Free Integer-Value Input Method by Quantizing the Wrist Rotation. *International Journal of Human–Computer Interaction* (2021), 1–18. https://doi.org/10.1080/10447318.2021.1898848

[77] Cheng Zhang, Junrui Yang, Caleb Southern, Thad E. Starner, and Gregory D. Abowd. 2016. WatchOut: Extending Interactions on a Smartwatch with Inertial Sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. 136–143. https://doi.org/10.1145/2971763.2971775