Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

From Parking Meters to Vending Machines: A Study of Usability Issues in Self-Service Technologies

Hamish Henderson, Kazjon Grace, Natalia Gulbransen-Diaz, Brittany Klaassens, Tuck Wah Leong & Martin Tomitsch

To cite this article: Hamish Henderson, Kazjon Grace, Natalia Gulbransen-Diaz, Brittany Klaassens, Tuck Wah Leong & Martin Tomitsch (22 May 2023): From Parking Meters to Vending Machines: A Study of Usability Issues in Self-Service Technologies, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2212228

To link to this article: https://doi.org/10.1080/10447318.2023.2212228

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 22 May 2023.

Submit your article to this journal

Article views: 917

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hihc20

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2212228

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS
Check for updates

From Parking Meters to Vending Machines: A Study of Usability Issues in Self-Service Technologies

Hamish Henderson[a], Kazjon Grace[a], Natalia Gulbransen-Diaz[a], Brittany Klaassens[a], Tuck Wah Leong[b], and Martin Tomitsch[b]

[a]The University of Sydney, Sydney, Australia; [b]The University of Technology Sydney, Sydney, Australia

ABSTRACT
This paper describes a mixed-methods usability study of seven diverse Self-Service Technologies (SSTs). SSTs mediate many of our everyday interactions with individuals, businesses, and government organizations. Parking meters, transport ticket machines, electric vehicle recharge points, and fast-food ordering kiosks are all likely familiar examples of this category of technology, promising convenient access to products and services. Despite their ubiquity, many SSTs suffer from severe usability issues, the nature of which have not been explored to date by the HCI community. This study evaluates the interactions between users and a broad sample of SSTs, details the usability issues that occurred, explores their connections and consequences, and presents a set of design considerations that may lead to their remediation.

KEYWORDS
Self-service technologies; usability; mixed methods; urban interfaces

1. Introduction

Our built environment is a complex amalgamation of built forms, sites, and systems that ultimately demonstrate our alteration of the natural world to suit our needs. While initially that amalgamation consisted mostly of hearths, homes, and public squares, automation has progressively transitioned our built environment into a space strongly mediated by interactive technologies. Self-Service Technologies (SSTs) straddle that boundary between urban infrastructure and Human-Computer Interaction (HCI), exemplifying this pervasive technological presence in the contemporary built environment. SSTs are a broad category, including vending machines, ATMs, parking meters, kiosks, and ticket machines. They provide critical services – payment, wayfinding, retail, etc – for both private companies and the public sector.

Despite the ubiquity of these technologies in our cities and lives, there is a consistent lack of design attention relative to other technologies. SSTs are outside the domain of architects, urban planners, and civil engineers, but they are also not clearly the domain of Interaction Designers, User Interface developers, or other HCI practitioners. Most SSTs lack the glamour of their more novel digital cousins in the modern city (such as automated vehicles and virtual reality devices). However, examples of innovative SSTs do exist: Amazon Go (Polacco & Backes, 2018); Airport Kiosks (Abdelaziz et al., 2010), and Robot Baristas (Sung & Jeon, 2020). These are exceptions outside the reality of most urban dwellers, and by contrast, we expect many of you can vividly recall a negative user experience with at least one SST. Where usability research into SSTs has been conducted, it confirms this *vox populi* perspective: for example, despite the more than eight-decade history of self-service parking meters, interacting with these technologies overwhelmingly evokes feelings of frustration (Henderson & Leong, 2017).

The causes for this neglect are likely systemic and varied: hardware lifecycles, changing payment systems, the cost of updating interfaces, etc. The HCI community has contributed many insights over decades in how we might design more usable and useful technology. SSTs as a class of technology have been underserved by this discipline, perhaps due to their relative mundanity or the complexity of the context in which they reside.

It is also important to note that the lack of attention to SSTs cannot be explained away by their nature as a "transitional technology" – a necessary but brief step on the road to entirely frictionless automation. Some SSTs may be replaceable by mobile or even zero-interaction solutions, but many require physical infrastructure (such as automated bag-drop machines at an airport) or may benefit from a larger display than a mobile device can provide (such as a directory). Some services will surely become frictionless, such as highway toll booths, but it is equally likely that other services currently provided by humans will be replaced with new kinds of SSTs. SSTs as a category, then, are not just unlikely to vanish but may continue to grow. Unless the situation is rectified, their poor usability will continue to affect the daily lives of millions.

CONTACT Hamish Henderson hamish.henderson@sydney.edu.au The University of Sydney, Sydney, Australia

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

As a step towards addressing this paucity of knowledge, this paper describes a usability study of seven SSTs, chosen to represent as diverse a sample of the whole product category as possible. Through a combination of qualitative and quantitative analysis, we present several broad usability concerns, providing rigorous affirmation to the widespread assumption that the user experience of most SSTs is poor. The fractious and frustrating interactions that many of us have had with SSTs are not fundamental to the medium: they could be fixed with the proper application of human-centered design. In order to begin the process of resolving this decades-long lack of attention, we present a set of design recommendations – derived from our broad usability study – for how the SST experience can be re-imagined.

2. Related work

In this section, we describe the history of and recent trends in the field of SSTs, as well as discuss usability evaluation research, with a focus on mixed-methods approaches like those in our study.

2.1. Self-Service technologies

Castro et al. (2010) described self-service as the “process by which consumers engage in all or a portion of the provision of a service or product”. Most self-service technologies are currently delivered via one of four channels: electronic kiosks, the Internet, mobile devices, and the telephone. In this research, we are specifically interested in those self-service technologies physically embedded in urban settings – what Castro et al. call “electronic kiosks”. Implemented by both businesses and governments to reduce the cost of delivering services, they can be found in city streets (parking meters), grocery stores (self-checkout machines), banks (ATMs), and inside shopping centers (information kiosks).

SSTs have a longer and richer history than one might suspect. There is some evidence that the Hero of Alexandria designed a holy-water vending machine around 219 BC, as well as records of tobacco vending machines in taverns in 1600s Britain (Segrave, 2015). The modern history of self-service business models can best be traced back to a 1916 grocery store in Memphis, Tennessee: Piggly Wiggly. Piggly Wiggly proposed a novel idea: consumers would be happy to select their own products in exchange for increased convenience and lower cost (Freeman, 1992). Prior to Piggly Wiggly, employees would select all groceries for each customer. Piggly Wiggly was the first store to feature hand-baskets and customer-accessible aisles, the infrastructure that supports a level of self-service we now take for granted. Because of this technology, Piggly Wiggly was able to offer cheaper prices, and rapidly outcompete other Memphis-area stores. Consumers appeared happy to trade the human interaction of earlier sales modalities for the reduced prices of a self-service model. In the hundred years that have followed, more and more businesses have been building on that lesson from a small grocery store in Tennessee.

This business case quickly spread to other domains, notably the parking meter in 1936 (Grimes, 1947), which was the first mechanical device to enable self-service, and the information kiosk in 1977, the first computer-driven SST (Gómez-Carmona et al., 2018). The widespread adoption of digital computers allowed these technologies to support increasingly varied tasks over the last four decades. As with many product categories, computers did not so much permit an entirely new innovation, but instead accelerate what had come before.

While the commercial opportunity that self-service represented became rapidly apparent, challenges in the consumer experience arose just as quickly. The Piggly Wiggly stores underwent several iterations in response to shortcomings in the implementation, including around aisle width, checkout staffing, and store layout (Freeman, 2019). While the technologies, environment, and urban contexts have evolved rapidly, many of these underlying issues remain unresolved. Modern-day criticism of self-service technology as a paradigm tends to focus on four broad categories: shifting work to consumers, destroying jobs, eliminating consumer choice, and removing the human-to-human experience (Castro et al., 2010). By contrast, criticism of individual SSTs overwhelmingly tends to focus on the poor usability of that specific implementation (Henderson & Leong, 2017; Hoffman, 2000; Turner & Szymkowiak, 2019). A renewed design interest in SSTs could definitely address this latter criticism, but we would propose that a more well-crafted user experience could help address Castro’s broader issues of consumer effort and choice as well.

2.2. Mixed-methods usability evaluation

Where HCI research has been conducted with SSTs, it has tended to focus on specific re-designs (eg, Ekşioğlu, 2016; Kristoffersen & Bratteberg, 2010), leaving the sector-wide usability issues unstudied. In order to address the severe design flaws of the SST as a genre, it will be necessary to understand the common usability issues shared by devices across the sector.

While usability analyses have become a mainstay of both academic and commercial design practice, there has been decades of debate as to how it can be used most effectively (Hartson et al., 2001). Perhaps the most widely applied usability evaluation method is the System Usability Scale (SUS, Brooke, 1996), which can quickly and cost-effectively measure the usability of a system (Peres et al., 2013). While there is still much debate in the HCI community about the effectiveness of the SUS, it’s widespread use supports its continued utility (Bangor et al., 2008).

The emerging consensus is that, where resources permit, a highly effective approach to usability is to triangulate measures of the user experience through the use of convergent, parallel methods (Creswell & Creswell, 2017). Ideally, these should include a mix of qualitative and quantitative measures (Patton, 1999). This mixed methods approach to usability has been used extensively, particularly in electronic health and medicine research (Khairat et al., 2019; Liew et al., 2019). We adopt such an approach in our study, in order to get a richer understanding of the types and causes of usability issues in the SST sector.

3. Methods

We employed a mixed qualitative/quantitative field usability study approach to investigate the usability of various self-service technologies and understand how usability challenges manifest for their users. Users were accompanied around an urban area on a pre-set path, asked to interact with a variety of SSTs, and described their experiences in relation to each SST. Each of these SSTs were pre-existing in live, operational environments and in use by the general public. Data collection comprised of task-based think-aloud (Ericsson & Simon, 1998), user observations (Diaper & Stanton, 2003), SUS (Brooke, 1996) evaluations, and semi-structured interviews. Both the tasks and the interviews were audio- and video-recorded. The verbal data was then transcribed and subjected to an inductive thematic analysis (Clarke et al., 2015). We augmented the thematic analysis with the video recordings in order to more effectively understand the physical context of specific user quotes (Lemke, 2012).

3.1. Participants

Through the researchers' personal networks, potential participants were contacted directly and sent a flyer outlining the study. They were also asked to share this flyer with other individuals they thought may be interested in participating. Further, the flyer was placed on noticeboards throughout a large Australian university in both staff and student areas. To participate, individuals had to be over the age of 18, proficient in the English language and able to articulate their experience of engaging with various SSTs and be comfortable walking to and from different sites near the university campus. Written consent was obtained prior, and the protocol was approved by the University of Sydney's Human Research Ethics Committee.

Of the 30 participants (Table 1), 57% were female ($n = 17$) and 43% were male ($n = 13$). The overall sample was skewed towards a younger demographic with 53% of participants aged 18–24, and another 23% aged 25–34.

3.2. SST selection

We selected our SSTs with the aim of covering the broadest possible range of services and products, within the constraint that we needed participants to be able to reach and interact with them in a single 90-min session. Given that constraint, selected seven SSTs, visible in Figure 1; Parking Meter (PM), Drinks Vending Machine (DVM), Automatic Teller Machine (ATM), Grocery Store Checkout (GSCK), Shopping Centre Directory (SCD), Fast-Food Kiosk (FFK), Train Ticket Machine (TTM).

These machines cover a range of common interaction modalities including touch screens, physical buttons, payment methods, and the vending of physical goods. They also included both unified single-interface devices (eg, the SCD), and devices composed of multiple separate sub-interfaces (eg, the GSCK).

Table 1. Participant demographics.

Variable	Item	Number (%)
Gender	Female	17 (57%)
	Male	13 (43%)
Age group	18–24	16 (53%)
	25–34	7 (23%)
	35–44	3 (10%)
	45–54	2 (6%)
	55–64	1 (3%)
	65 (+)	1 (3%)

3.3. Data collection

Once recruited, participants selected a 2-hr session time that suited them. After acknowledging their consent to participate in the study, participants disclosed their demographic information (age and gender) to the session facilitator. Sessions followed a study protocol to ensure consistency, and progressed as detailed below, with the tasks associated with each SST outlined in Table 2.

1. *General introduction:* brief summation of the research's purpose and a reminder of the structure of the session. The facilitator also introduces participants to the think-aloud method, answering questions if appropriate.
2. *Arrival at SST:* facilitator briefly introduces the SST to the participant, and asks about the participants familiarity and potential past interactions with the device.
3. *Introduction of task(s):* facilitator outlines the task to the participant and provides any supplementary materials required for its completion (eg, a bank card loaded/not loaded with money).
4. *Task(s) completion:* participant interacts with the SST to complete the task(s) while thinking aloud. Audio and visual data is captured via a GoPro Hero 7 hanging around the facilitator's neck. Facilitator concurrently records observations regarding the participant's body language and behavior.
5. *SUS evaluation:* the participant is provided with a paper copy of the SUS scale and then asked to respond to each of the SUS prompts.
6. *Semi-structured interview:* based on their immediate experience, participants were asked to describe the best and worst aspects of the system, as well as general improvements they would make to the design of the device. Additional prompts and inquiries were made at the facilitator's discretion, encouraging the participant to describe any thoughts or feelings they perceived to be relevant to the topic.
7. *Facilitator reflection:* facilitator notes down lingering thoughts and reflections on the session. The facilitator confirms or adjusts field notes where necessary.

Two authors acted as session facilitators and a third author transcribed the audio from all the sessions, and recorded their observations of participant behavior from the captured footage. Overall, the transcribed audio, recorded field notes from three of the researchers and recorded footage was amenable for analysis.

Parking Meter (PM) *Drinks Vending Machine (DVM)* *Automatic Teller Machine (ATM)* *Grocery Store Checkout (GSCK)*

Shopping Centre Directory (SCD) *Fast-Food Kiosk (FFK)* *Train Ticket Machine (TTM)*

Figure 1. The seven SSTs used in this study.

Table 2. SSTs and tasks.

#	Technology	Task(s)
1	Drinks vending machine (DVM)	Using this card [card loaded with money], purchase a bottle of water and tell me how much it is
2	Automatic teller machine (ATM)	Using this card [card loaded with money], withdraw $20, and then deposit this cheque
3	Parking meter (PM)	1. Select 1 hr of parking [card loaded with money], and tell me how much it costs. 2. Using this card [card loaded with no money], purchase the most amount of parking time for $10
4	Grocery store check-out kiosk (GSCK)	Using this card [card loaded with money], purchase one fixed price, packaged product, and one product that requires weighing
5	Shopping centre directory (SCD)	1. Find a route to Just Cuts. 2. Find a route to Aldi using lifts, not escalators. 3. Search for a shoe shop.
6	Fast-food kiosk (FFK)	1. Order a cheeseburger with extra pickles, and pay using this card [card loaded with no money] 2. Order a bottle of water, and pay using this card [card loaded with money]
7	Train ticket machine (TTM)	Using this card (no money), buy a single adult ticket to Lidcombe

3.4. Data analysis

Data from the sessions was analyzed using an open and inductive thematic analysis (Clarke et al., 2015) in a collaborative analysis workshop. Three of the authors participated in these workshops for investigator triangulation, allowing each design researcher to "expand, correct, or check" each other's subjective view of the data (Jenner et al., 2004). The workshop was broadly split into three phases: (1) independent thematic analysis of transcribed data 1st order codes, (2) confirmation of insights with field notes and the collaborative development of 2nd order sub-themes, and (3) identification and explanation of global themes.

3.4.1. Phase 1: Thematic analysis

During the first phase, thematic analysis was performed in NVivo in accordance to Gioia et al. (2013)'s three orders of coding and Corbin and Strauss (2014) notions of axial coding. In order to identify and synthesize common trends in the data and extract meaningful patterns for each SST, the

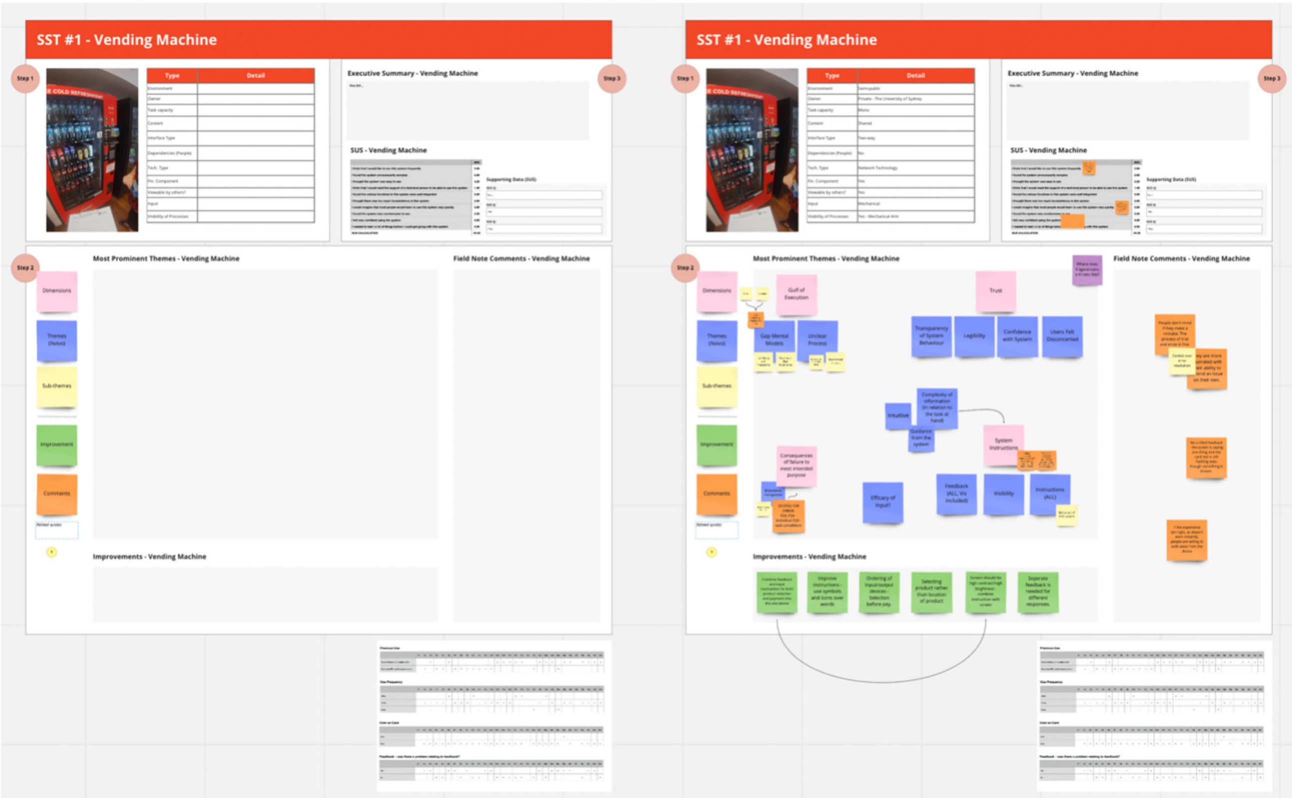

Figure 2. Example structure for Drink Vending Machine SST (Left: blank, Right: partially completed).

researchers independently coded the transcripts relevant to each individual SST in NVivo. Early discussions between the researchers identified preliminary codes such as *performance pressure*, *unexpected outcomes,* and *convenience*, which in turn, established a common perspective through which the researchers continued their analyses. Resulting code reports were collaboratively amalgamated through a process of discussion, deletion, and merging. The remaining 1st order codes were then classified per the participant's terms or similar summative phrases.

3.4.2. Phase 2: Collaborative workshop

The collaborative workshop session spanned three days and was used to complete the second round of thematic analysis and determine if the 2nd order sub-themes could be distilled further into the 3rd order global themes (Gioia et al., 2013). This process of distillation was completed in conjunction with a review of the field notes and a random sample of video recordings for each SST, ensuring that the observable behaviors and emotional expressions of participants was also represented in our interpretations of the data.

A workshop structure (Figure 2) and agenda were prepared prior to the first meeting, ensuring that a consistent approach to analysis was undertaken across all SSTs. The structure firstly required the researchers to identify various characteristics of the technology (eg, the environment in which the technology is situated, owner of the technology, task capacity, content, interface type, dependencies, technology type, etc.). Secondly, SST-specific codes (and their respective quotes when further clarity was required) were added and discussed amongst the researchers. Each researcher's field notes relevant to the selected SST were noted separately to the thematic codes before being incorporated into the analysis where appropriate.

With both 1st order thematic codes and field notes available, the second round of analysis was situated firmly in the theoretical realm, examining and further categorizing the data into 2nd order sub-themes. These sub-themes named various usability challenges experienced by participants engaging with the SSTs and described how the consequences of these challenges may manifest. When the 2nd order sub-themes were established, randomly selected session videos were viewed to confirm the developing data structure had achieved saturation (Corbin & Strauss, 2014). 2nd order sub-themes were also compared against the SSTs SUS score, to understand any correlations between specific question scores and the emergent sub-themes. Following a final discussion between the three design researchers, an executive summary detailing relevant insights was produced. This process of triangulating thematic analysis codes, field notes, and SUS data was repeated for each of the remaining SSTs respectively.

3.4.3. Phase 3: Development of global themes

33 SST-specific themes emerged from our observations and interviews. These themes were then grouped into nine global themes (Table 4).

4. Results

We present the results of our study in two sections: first the quantitative SUS results (4.1) and then the thematic analysis of interview data (4.2).

4.1. Standard Usability Scale (SUS) scores

We integrated the SUS survey into our semi-structured interviews with our participants. All SSTs were used by all 30 participants, with the exception of the ATM (only 29 participants) and the Train Ticket Machine (29 participants). In the case of both of these technologies, the device was out of order at the time of the study. We have ranked the SSTs according their mean score (high to low). Table 3 shows the SUS scores for each SST, and Figure 3 shows the distribution of each SST's results as a box plot.

Our results from the Standard Usability Scale (SUS) component of the evaluation highlighted that these usability issues were not restricted to one type of SST; indeed when using the benchmark "grade scores" (A, B, C, D, and F) for the SUS as defined by Lewis and Sauro (2018), 84% of user responses sit within the D (usually understood as a "minimally acceptable" grade) or F (a "failure" grade) ranges. The utility of Lewis and Sauro (2018) grade scale is supported

Table 3. SUS score results.

#	Technology	Mean	Low	High	SD
1	Train ticket machine	56.57	37.5	70	6.75
2	Shopping centre directory	56.23	40.5	70	8.08
3	Fast-food kiosk	55.52	33	69	8.25
4	Grocery store check-out kiosk	53.62	24	70	12.20
5	ATM	49.91	20.5	67	12.68
6	Drinks vending machine	42.37	12.5	63	14.55
7	Parking meter	29.25	7	55	12.79

Our results from the SUS component of the evaluation highlighted that these usability issues were not restricted to one type of SST. To interpret these SUS results, we utilized a variety of interpretations from the literature. Lewis and Sauro (2018) proposed a "letter grade" system (A, B, C, D, and F) to categorize SUS results. Bangor et al. (2008) additionally mapped SUS Scores to "Acceptability Ranges" and provided adjective ratings of quality. Viewed thusly, 84% of user responses sat across D and F grades (per Lewis and Sauro), and ranged from "not acceptable/poor" to "marginally acceptable/OK" (per Bangor et al). For additional context, contrast these scores with those of everyday products benchmarked in Kortum and Bangor (2013)[1]: in-car GPS units (70.8), Microwaves (86.9), and Google Search (93.4).

Put simply, there was no aspect to the SUS scores that was even marginally redemptive for any of the tested SSTs using any of the published interpretations. Figure 3 depicts the category-wide and substantial SUS failure of SSTs, showing mean, minimum, maximum, and interquartile ranges for each SST.

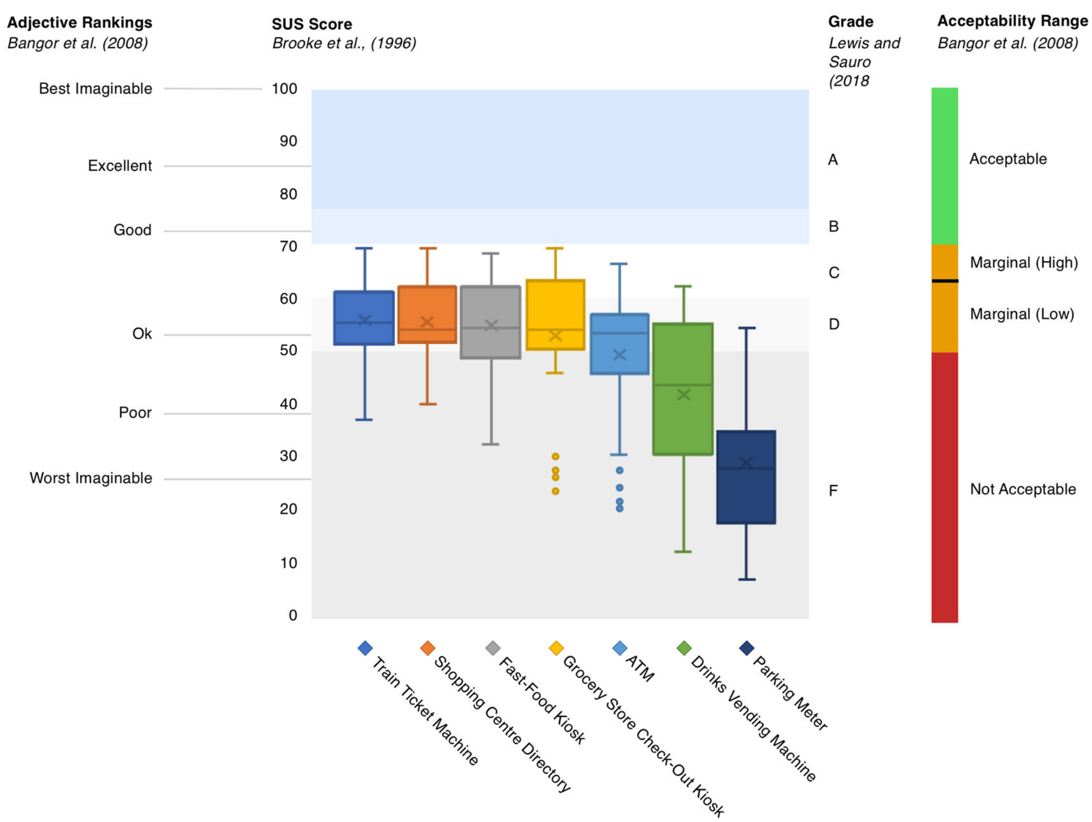

Figure 3. SUS scores per SST, presented alongside rating scales from Lewis and Sauro (2018) and Bangor et al. (2008). SST usability was bad across the board, with several SSTs (the vending machine and parking meter) being exceptionally bad. Not a single rating from any user exceeded a score of 72, considered the bottom-end of "good".

While the scale is only a broad guideline on the acceptability of an interface's usability, the results are utterly conclusive: no single user rated a single SST as a good experience, and on average all experiences were poor or minimally acceptable. Of particular note is the heteroskedasticity of the scores: the means in Table 3 are clearly inversely correlated with the standard deviations. This increase in variance as scores decrease can also be seen in Figure 3, with the SSTs towards the right of the figure having wider ranges of scores. We suspect that this effect is a result of the significant usability issues identified in our interviews: major interruptive issues will affect only a portion of users, resulting in two "groups" of scores: those who were affected, and those who were not. Intuitively, it seems likely that when the usability of a system is very low (as is the case with several of our SSTs), there will be serious issues and failed interactions for a portion of users, rather than a uniform degradation of experience for all, and our data support that.

While the age skew of our participants makes it difficult for us to say these results would generalize to the entire population, the strength of the data makes it unlikely that population-representative results would be at all positive for SST usability. To provide some data towards that assumption, a t-test comparing the 7 (23%) of our participants who were 35+ against the 23 (76%) who were younger showed absolutely no difference in SUS scores ($mean_{\text{old}} = 44.3$, $mean_{\text{young}} = 45.8$, $p = 0.55$). While the statistical power of that comparison is limited, there is no evidence in our data for an age effect on SUS in our context.

4.2. Thematic analysis

33 SST-specific themes emerged from our observations and interviews (Table 4). These themes were then grouped into nine global themes: (1) Clarity of Guidance, (2) Confidence and Trust, (3) Interface Cohesion and Capability, (4) Efficiency and Legibility, (5) Feedback, (6) Recoverability, (7) Social, (8) Assumed Knowledge and (9) Comprehensibility and Accessibility. We explore each of these themes in turn and how they affected the usability of SSTs.

4.2.1. Clarity of guidance

Clarity of Guidance refers to an SSTs ability to give timely and meaningful instructions to the user on how to use the machine. Clear guidance on how to use the machine is essential for a successful and efficient interaction. As SSTs are used infrequently, an SST cannot rely on a user's familiarity to offset and overcome usability issues.

Guidance is not limited to explicit written instruction; it includes the organization and ordering of input and output devices, the affordances of those input and output devices along with visual cues such as symbols and icons. Instructions can be displayed dynamically on a screen, or statically; printed on the device itself, and included labels beside buttons.

The presence of instructions was not apparent or obvious to participants, which led to difficulties initiating the interaction and completing the intended task. *("No I don't think I even noticed the printed labels"—P17)*. Participants often employed a trial-and-error approach to the interaction and often appeared confused in how to proceed to the next step *("It didn't tell me at all what to do, so I just pressed a button."—P29)*.

Given the perceived simplicity of the task, participants expected to be able to complete their task quickly and to not have to spend too much time learning how to use the system. *("The confirmation screen is too complicated. Like just give me the $20, why do I need all of this information."—P15)*. The unexpected complexity as a result of usability issues present in the SSTs led to frustration and confusion ("This selection screen is a very confusing

Table 4. Thematic analysis.

Global theme	SST level theme	SST
Clarity of Guidance	Visibility and prominence of instructions	DVM, ATM
	Quality and usefulness of instructions	ATM, DVM
	Visual clutter and disruption to user flow	FFK
	Initiating the interaction	DVM, PM, SCD
	Inline guidance	DVM
Confidence and Trust	Confidence in system	DVM
	Trust in system	PM
Interface cohesion and capability	Fragmented interaction	DVM
	Perceived capability and integration	ATM
	Fragmented process	GSCK
	Component proximity	TTM
Efficiency and legibility	Speed	ATM
	Compromised input mechanism	PM, FFK
	Poor legibility	PM
	Visual clutter	GSCK
Feedback	Poor feedback	PM, DVM
Recoverability	Control, independence and error recoverability	GSCK
Social	Performance pressure	TTM, FFK
Assumed knowledge	Reliance on previous experience	GSCK
	Reliance on external information	PM
	Assumed knowledge	TTM
Comprehensibility and accessibility	Situational awareness	SCD
	Accessibility	SCD
	Information hierarchy	FFK

interface, I just want a yes or no option."—P12). In some cases, the participant was not even aware that the SST was interactive. In the case of the shopping center directory, a number of participants mistook the device for a passive advertising panel, primarily as a result of an absence of clear purpose messaging and system affordances *("No it wasn't, because it wasn't clear if it was touch screen or another ad."—P15).*

Where instructions did appear, their composition was a factor in their utility to the participant. Verbose, text-based instructions are often ignored by the participant in favor of exploring the device (through trial and error) to figure out how it works. Icons and symbols were in many cases more effective, and more noticeable than text-based instructions. Comprehending icon-based instructions was quicker than text-based instructions. The ordering of instructions was important; participants often required inline, contextual instructions to successfully complete their task.

Where a dynamic screen exists alongside printed instructions, participants tended to focus on the screen first. Usability issues were exacerbated when screen instructions contradicted the printed instructions.

Participants became frustrated when the flow was interrupted by visual clutter, primarily in the form of advertising and up-sell messages *("I think there were some screens that were interruptive. Like the upsell"—P11)*. This additional information increased the participants cognitive load as it was another piece of information to process. These disruptive messages affected participant momentum and caused frustration *("Getting to the actual checkout was challenging as they kept trying to add things to my cart."—P27).*

4.2.2. Confidence and trust

Participants had expectations of how the machine would work and behave when used. When the machine didn't meet these expectations, participants lost confidence in their ability to use the SST effectively and did not trust that the SST would do what they wanted. These feelings were further amplified when there was a monetary component to the interaction *("I think the machine is really trying to rip me off."—P24).*

The deterioration of Confidence and Trust occurred primarily as a result of insufficient feedback, which was either non-existent, inadequate, or unclear *("I would not go back to this machine as it made me feel unsafe and it didn't do what I asked it to."—P13)*. Participants were left feeling confused and wary of whether the SST was behaving as intended. This was further magnified by poor-quality hardware, which led feedback to become illegible in some instances. Poor feedback mechanisms combined with a cumbersome input mechanisms undermined the trust the user had that the system would perform as desired *("The card reader is clunky, I don't trust the machine. It never really works properly. I am sceptical based on past experience"—P16).*

Participants also became concerned when they perceived that they were being asked to pay before making a selection, which was the case with the Vending Machine. In addition to being a reversal of the ordering of a typical transaction, participants did not understand how the machine could charge their card an amount for a product they had not selected which varied in price *("I didn't really want to give them my money before I made a selection."—P12)*. In reality, the SST was not charging their card the full amount but was charging a "pre-authorization" amount. This, however, was not clear to the participants. Some participants even became sceptical and believed that the SST was trying to deceive them of their money.

It became clear that the inability for a participant to reason with the SST left them feeling vulnerable and extremely frustrated *("It was really unclear, and they have made no attempt to make it clear. I feel like I have been tricked"—P29)*. This feeling was exacerbated in cases where the SST was located in an area where no human assistance could be provided.

4.2.3. Interface cohesion and capability

The SST interfaces that were evaluated comprised of a number of different input and output mechanisms. In many cases, these mechanisms lacked cohesion and relationship with one another. This fragmentation led to several usability issues for participants. Fragmentation or perceived fragmentation occurred when the input for an action and feedback for that action appears in multiple locations. This would present difficulties, for example when participants were focused on one screen that until that point had been responsible for providing feedback, only to discover that another screen was now responsible for providing feedback for the current step (such as a payment terminal screen). As a result of the broken feedback sources, participants were not always sure where they needed to focus at a given point in time. This led to confusion and frustration as users did not always know whether they'd successfully completed the current step, or what the next step was *("I was very confused as to where to look each time."—P20).*

The proximity of the input/output devices to one another played an important part in the participants' perception of cohesiveness *("It would have been nice to integrate the eftpos in the screen rather than a separate system. Just to make it flow better."—P16)*. The further apart the input/output devices were from each other, the more difficulty the participant had in seeing the relationship between them. As many of the SSTs are physically bigger than the user, the space between devices could be material. When participants were focussed on a particular input/output device, they often failed to recognise another input/output device necessary to complete their task as it was outside of their immediate field of view *("What is interesting is that when the card was declined, this machine over here made a sound and went red. I didn't even notice its existence. More integration there is probably needed."—P15)*. Participants occasionally had to step back and visually scan the SST to bring all elements into their field of view.

Participants were not always aware of the capability of the input/output devices on the SST. For example, participants when confronted with mechanical buttons and a screen assumed that the screen was not touch-enabled *("I didn't realize this touch screen was an ATM, I am used to the more mechanical ATM."—P12)*. This in turn caused some confusion as the mechanical buttons did not support all the actions that the user was required to make. In one case, the SST had to rely on a message to remind the user that the screen was a touch screen *("What is weird to me is that I don't immediately know if the screen is a touchscreen or I have to use the buttons. Granted it does say "touch screen" at the bottom, but that was not clear to me."—P15)*.

4.2.4. Efficiency and legibility

The type of input/output mechanisms materially affected the usability of the SSTs, specifically the ability for the participant to accurately and efficiently make selections and receive feedback. Some input mechanisms, such as the buttons on the parking meter, proved to be cumbersome and time-consuming for the participant to make their selection. While the buttons did allow the participant to eventually complete the task, other input methods such as a control knob would have allowed them to make the selection faster *("Ok this is frustrating, because I have to press this a bunch of times, and it is reduced by 50 cents each time. The fact that it started at $20, laborious and annoying."—P15)*.

In a number of cases where the most appropriate input/output mechanism was used, the execution was problematic. For example, using a screen to provide feedback on the parking meter was an appropriate mechanism, however, the type of screen used meant that content was often illegible when viewed on a bright, sunny day *("I first looked at the screen for the information, but the sunlight is interfering with the visibility of the screen."—P17)*. Other examples include the use of touch screens that did not support smartphone gestures (swiping, pinching, etc) and glitchy scroll bars on the fast-food kiosk.

The legibility of both dynamic and static instructions and feedback dramatically affected the participants ability to effectively and efficiently complete their assigned tasks. The legibility of instructions and feedback was compromised by poor quality hardware, small font size, placement of display (where line-of-sight was impinged in some way), low-contrast color selection of content and background, and environmental damage to the SST *("It was very hard to read"—P7; "Well the first thing is that I can't even read it, I must be missing something. Surely there has to be a better way."—P25)*.

Issues with legibility were sometimes amplified by the amount of visual clutter present on the interface. This clutter included items deemed nonessential by the participant to completing their task; such as advertising messages. The more information provided by the SST meant more effort on the part of the participant to process, understand and act. This could lead to an increase in the amount of time required to complete a task *("it was verbose and told me things I already know."—P30)*.

4.2.5. Feedback

An SST's ability to provide timely and meaningful feedback to the participant proved decisive in the overall success of the interaction. Where an SST was unable to provide feedback quickly, the participant was often left unsure and confused *("You also don't really know if your machine is registering what you are doing"—P13)*. This delayed the completion of the interaction and created a level of frustration in the user. In some cases when feedback was present, the participant was unsure of the meaning *("it told me something was wrong, but not what I could do to fix it."—P2)*. This was particularly true of audio feedback, where a sound was heard, but the tone was not informative of the state of the system or outcome of the action *("I think the feedback was inconsistent as I didn't know what the beeps mean"—P7)*.

Audio feedback in certain contexts was inhibited by loud environmental noise, particularly in high-traffic areas such as the train station. In these cases, both physical and visual feedback was prioritized. Many of the SSTs did not provide physical feedback given the reliance on touch screens, which led to an over-reliance on visual feedback. In the case of the parking meter, where physical buttons were used, participants expressed frustration at the inability of the button to provide physical feedback. The pressure-sensitive buttons did not move in their container and required an unusual amount of force to activate.

Feedback was a vital component for understanding the success of the current step of the interaction, and encouragement to proceed to a subsequent step. Participants remained uncertain of the success of their actions when using SSTs where feedback was slow, unclear, or not present *("When you press the button it doesn't give you feedback and you have to really look at the screen, and the button is so hard to press."—P16)*. This uncertainty could lead to false assumptions of the system state, followed by decision-making based on faulty information. *("You don't know what you are supposed to do. You can see the time or money change, that gave me some confidence, but there was no confirmation"—P12). ("I am confused because I am unsure if I did it right."—P24)*.

4.2.6. Recoverability

In the event that a participant encountered an error or an unexpected outcome when using an SST, their ability to recover and remedy the issue was an important factor in the overall usability of the device *("It is annoying to have the assistant come over because I know what the error is, so the fact I have to wait for someone to come help is so frustrating. I should be able to fix that problem myself."—P18)*. In cases where the SST was not in a location where human assistance was available, the importance of a user being able to self-recover was increased. This issue was amplified in interactions that involved money, as some participants felt that their access to recourse was diminished in these situations should something go wrong. SSTs that did show that an error had occurred but did not provide information on how to recover were particularly frustrating.

While there appeared to be a desire for human support in the event something went wrong, there was seemingly a contradictory reaction in the case of the supermarket check-out kiosk. Participants became frustrated and occasionally embarrassed when an assistant was required to help them complete their transaction. Participants also became annoyed at how easily the system would self-determine that assistance was required. When the system entered this mode, all inputs were blocked and participants would be forced to wait for assistance—unable to recover themselves.

4.2.7. Social

Rather than being the cause of a usability issue, increased performance anxiety was a consequence of poor usability in an SST. Participants became stressed when using an SST in a busy location where other people could observe them or were waiting to use the device. Some participants felt that his performance pressure in turn make them more likely to make a mistake *("it is the pressure of people queuing. I think it also if you are trying to use this for the first time, I make mistakes if I am under pressure like this, so I am more likely to make mistakes."—P2)*.

Participants did not want to be perceived to be incompetent and were visibly annoyed when the SST made them look like they were at fault *("Feeling observed using the system makes me feel uncomfortable, I don't care what people think but I don't want to be perceived as incompetent."—P15)*

4.2.8. Assumed knowledge

A number of the SSTs were designed in such a way that required prior experience with the technology or familiarity with the broader concepts that appeared in the interaction (such as barcode scanning). As a result, it took some participants longer to complete their task when they were unfamiliar with a stage in the interaction *("without experience I would not know where to weigh my items."—P28)*. While the participants were able to eventually complete their tasks in many cases, some participants flagged that they knew other potential users who would have had difficulty in completing the assigned task. The technical literacy and ability that was expected of some users by an SST could reveal knowledge gaps and as a result, an inability to complete the interaction.

This issue was mitigated somewhat in situations where the SST existed alongside other SSTs being used by other users. This allowed for learning through observation, as participants could watch another user to understand how the system worked.

To increase the chance of an SST interaction being successful, the SST must be capable of providing all the information the user needs to complete the interaction. In the case of the train ticket machine, the SST required the participant to have a knowledge of the different modes of transport (light rail vs heavy rail) available at different destinations *("I also found it confusing where you had to select modes of transport, I don't think everyone would know that Lidcombe is on the train line."—P4)*. In the case of the parking meter, the SST did not show the current time. As a result, users were required to have access to their own device in order to calculate the amount of time they wanted to purchase.

4.2.9. Comprehensibility and accessibility

The extent to which a user was able to comprehend the information presented to them during the course of their interaction directly affected their ability to use the SST to complete their goal. This was particularly evident when evaluating the shopping center directory. This SST provided users a map of the center and walking directions from the SST to their destination. A number of participants found the execution of the map did not assist them in building situational awareness and as a result, found it difficult to articulate the directions provided *("it doesn't help me find the place, it just directs me to it. The wayfinding part could be improved."—P8)*.

Several participants registered concerns about the SSTs capacity to accommodate users of varied physical ability. A key concern was the lack of support for people who were very short, or in a wheelchair. Other concerns included people who were vision impaired, as many of the SSTs relied very heavily on visual communication. While participants did not feel that their physical abilities prevented them from using the SSTs, they perceived that these devices were biased against people with differing physical abilities *("There is no auditory or haptic feedback at all and I think that is needed for accessibility purposes."—P28)*.

An additional barrier to comprehension was the information hierarchy applied to the SST. Several participants raised concerns with the prioritization and categorization of information *("Yes, one area it could be improved is the information could be more in the center of the display. I am conditioned to look there instead of the top."—P17)*. In some cases, they felt that this made it more difficult to understand the system and engage with the information being presented *("The categorization of things doesn't make sense to me, I think organizing it around popularity would be great."—P30)*.

5. Discussion

Our quantitative and qualitative results suggest that most SSTs fall short of accepted usability standards. Furthermore, we observed a negative correlation between the mean and standard deviation of SUS scores, which supports our observation that when something starts to go wrong, it quickly becomes a catastrophic usability issue for a portion of users. Many of the themes observed in our analysis describe interactions each that violate multiple well-established heuristics for interaction design (Nielsen & Molich, 1990; Shneiderman et al., 2016). For example, the lack of clarity of guidance we observed violates Nielsen's heuristics of *visibility of system status* and *availability of help and documentation*, as well as Shneiderman et al's golden rules of *providing informative feedback* and *reducing short-term memory load*. Remediating these issues and designing better SSTs will require – at a minimum – addressing these many and overlapping violations, but there is a lack of a clear path towards doing so.

As a starting point toward that aim, we found that the majority of these usability issues with SSTs are broadly attributable to the over- or under-provision of information to the user: a failure to appropriately balance their cognitive load. Both the scarcity that occurs when insufficient information is provided (eg, when getting started at the vending machine) and the saturation that occurs when there is too much presented at once (eg, when completing checkout at the grocery store) slow users down by forcing them to search for critical information before they can proceed. While many participants were able to recover from these (and other) issues, the experience degraded user trust, a consequence particularly dire in an SSTs context (see Section 4.2.2). We discuss the notions of cognitive load-balancing and rebuilding trust in SSTs here, and provide some recommendations for avoiding the pitfalls our study suggests are all too common in this sector.

5.1. Cognitive load-balancing

All SST interactions require cognition on the part of the user. Exceeding a user's cognitive processing bandwidth will almost guarantee usability issues. Furthermore, by virtue of the context of most SSTs (busy shopping centers and public spaces), that bandwidth is already compromised as users are exposed to a variety of stimuli. Information on how to translate goals into action is required, but in our study, this often happened too fast, overwhelming users, or too slow, leaving them confused. Managing this cognitive bandwidth is critical.

We observed that the lowest-scoring SSTs (drinks vending machine and parking meter) had particularly fragmented interfaces; inputs were made via physical buttons and feedback was delivered via a screen. We interpret that this led to an increase in workload for participants, as it was not always clear where they were required to focus their attention after completing an interaction. This was especially evident in heavily fragmented interfaces, such as when multiple screens were used on the device. Gestalt grouping and proximity principles are understood in the HCI community as an important consideration in the design of usable interfaces. It was particularly evident in the case of the drinks vending machine and the parking meter, that these principles were not observed. The significant physical space between elements of the interface, combined with no discernible logic as to their grouping was the source of many usability issues. By contrast, the highest-scoring SSTs (train ticket machine and shopping center directory) all featured touch screens as the primary input/output device and kept functions in close proximity. We hypothesize that this cohesion helped users to focus on the relevant interface element, particularly when their familiarity with the device was low. This fragmentation may be exacerbated by the retrofitting of componentry to support new functionality, such as contactless payment.

Many SSTs are used infrequently. It is unsurprising therefore that users require guidance to complete their tasks, or that mistakes are common when that guidance is lacking. Earlier generations of consumer technology were rarely sold without comprehensive user manuals, an explicit form of user guidance. Increasingly, however, consumers have the expectation that they should not be required to read large amounts of information: they expect to be able to discover how to interact with technologies on their own. Our study strongly supported the notion that most SSTs do not meet this expectation of discoverability. Users cannot rely on the intuitive problem-solving approaches employed to learn other technologies. Frequently used systems were found in our study to be more usable, likely due to familiarity effects. However, many SSTs rarely enjoy a high frequency of use. Principles like guidance are critical for supporting the usability of devices with which users interact only infrequently.

Our study highlighted that SSTs which presented their instructions all at once tended to overwhelm users, who were unsure where to start. By contrast, those SSTs which present little or no information tended to confuse their users, who were instead unsure how to start. One approach to reconciling this tension in the design of future SSTs is by progressively disclosing guidance. Progressive Disclosure, a well-known interface design principle (Nielsen, 2006), calls for key information to be revealed incrementally and only as required. In the context of SSTs, this would allow users to cognitively process only what they are required to, thereby reducing the chance they become overwhelmed. This provides the basis for our first design consideration:

1. SST interactions are infrequent and unfamiliar and thereby happen in conditions of limited cognitive bandwidth. This cognitive budget needs to be balanced carefully, including through the use of techniques like progressive disclosure, guidance, and feedback.

One established interface pattern for implementing progressive disclosure is the wizard: a step-by-step linear sequencing of required decisions. This results in reducing the user's cognitive load, as users are not required to make several decisions at once.

Guidance need not be explicit to be effective. For example, sequencing multiple decisions to reduce cognitive load can be achieved without an overt wizard-like interface. Written guidance may not always be appropriate due to issues of accessibility, but in many cases, it can be substituted through the use of visual affordances of sequence. This approach to the design of SSTs removes the reliance on the user to read instructions to understand what to do next. Information can be arranged to afford specific actions. For example, ensuring the first step to the interaction is at the top of the device, and the last step at the bottom. Our findings highlight the ongoing challenges of SSTs and accessibility, especially as they relate to the task of user guidance. Further research into supporting users with diverse cognitive and physical abilities is needed in the area of SSTs. This is given greater urgency by the increasing employment of SSTs for fundamental social and civic tasks.

Equally important to guidance prior to an interaction, is the ongoing guidance in the form of feedback during the interaction. Feedback, a well-known usability principle (Nielsen & Molich, 1990), has been employed on many of the SSTs we evaluated, however, the execution was often insufficient. In several cases, this was caused by hardware that was not fit for purpose, such as the illegible-in-daylight screen on the parking meter. In other cases, the choice of feedback modality failed to take into account the context, such as audio feedback in a loud shopping center.

The consequence of poor feedback leads to a hostile relationship between the user and SST and by extension the organization that deployed it. When feedback is unintelligible, users feel unsupported and may conclude that the organization is uninterested in providing a quality experience. Many of these issues could have been easily avoided if context had played a greater role in the design process. We believe that this mismatch arises from applying broad design principles intended for predominantly handheld or desktop interaction to a context such as SSTs is ineffectual. Existing HCI evaluation methods fail to take into account the unique environmental, social, and cognitive operating context of SSTs. This points to an opportunity for SST creators, and forms our second design consideration:

2. The gold standard of any usability evaluation is to prototype in the field, but this is often costly or otherwise impractical. Despite this expense, context is critical in evaluating SST user experiences. Designers must consider carefully how to balance this cost with the need for rapid and meaningful design insights.

5.2. Rebuilding trust

All interactions are fundamentally relationships, and all healthy relationships are built on trust. A relationship cannot flourish in the absence of trust, and if a user's trust in the system is broken negative experiences will follow. Trust must be built over time, which presents a challenge for systems that are used infrequently.

Support can help built user trust, but many SST interactions occur in contexts where there is limited or no human support available. A greater reliance is placed therefore on the SST to inspire the requisite confidence and trust to deliver a good user experience. User relationships with interactive technologies are often tentative: it is easy for users to lose trust and hard for systems to win it back. This is particularly true when the interaction is infrequent, unfamiliar, and unsupported, as is the case with SSTs. When trust is broken and confidence is lost, a successful transaction becomes increasingly less likely.

Unfortunately, it is all too easy for many users to mistrust SSTs. Prior negative experiences with poorly designed SSTs can leave lasting impressions, meaning many users are primed towards distrust. When money is involved, this effect is further amplified. The fragile relationship between the user and SST creates an imperative: designers cannot afford a single misstep that might break that trust.

Our study suggests that usability issues are a primary driver of this problem. Many of the issues experienced by our participants likely could have been avoided by employing well-known HCI design principles. For example the Vending Machine we evaluated asked users to make payment before product selection. This caused confusion, as users could not understand how the machine could charge them the correct price upfront. This confusion led to a distrust of the system since users were unclear how much they would be charged. This was entirely predictable as it violates the well-known usability principle of maintaining consistency (Molich and Nielson 1990).

For designers to build on fragile trust, they must ensure that SSTs are presented as familiar, safe, welcoming, and simple. While known HCI usability principles are a start, they do not go far enough in taking into account the unique context of an SST. For example, the principle of visibility (Norman, 1995) states that user confidence is greater when an element is more prominent. For some SSTs (including the train ticket machine in our study), an increase in visibility may make the interaction more public, which can create performance anxiety in users. Further research is required to adapt usability principles – which were primarily designed in the era of the personal computer – to the context of SSTs.

Psychological safety, defined as the knowledge that failures will not be punished (Newman et al., 2017), is known to be an important component of trust-based relationships. A user who fears they may be unable to recover from an error will become anxious and lose trust in the system. Therefore, providing assurance of recoverability is a key ingredient to maintaining a trust-based relationship. Delivering this assurance is not limited to the device. In some contexts, human assistance can augment the SST, such as the grocery store checkout. However, this assistance is not always welcome. Our study showed that it can frustrate or embarrass some users while simultaneously providing assurance of recoverability to others. Care must be taken to balance the assurance of support with maintaining user agency.

The very public nature of many SST interactions exacerbates these delicate trust issues, particularly when others were waiting to use the device. This waiting crowd increases the pressure to perform. People do not enjoy being embarrassed, especially in public, and certainly not by a machine. The resulting anxiety, coupled with the pressure to complete the interaction quickly, amplifies any usability issues present. If, in this environment of fear and haste, a user does make a mistake, a rapid deterioration in trust tends to follow. This underpins our third design consideration:

3. The nature of SST interactions means that it is particularly challenging to build and maintain trust with the user: relationships with SSTs are fleeting and fragile. Cognisant of this fragility, every macro-, and micro-interaction contained within an SST must place users' psychological safety at its center. Principles such as recoverability and visibility

are critical to maintaining this, but they need to be employed with nuance given the environmental, social, and physical peculiarities involved.

6. Conclusion

Despite the important role that SSTs play in our everyday interactions with businesses and government services, there is limited research into the usability of these devices. While several studies that focus on a single incarnation of SSTs have been conducted, our study is the first significant broad usability study on this class of technology. This study included a selection of SSTs that represent the wide variety of SST implementations available. This allowed our team to better understand the recurring usability issues that occur across SSTs as a category, rather than our insights being limited to a particular context, use case or manufacturer.

The outcome of our investigations was consistent and clear: users encountered severe usability issues when using SSTs, which resulted in a negative user experience. The importance of these devices in society highlights the need for these problems to be addressed. Combining the SUS scores with our think-aloud interviews revealed several insights as to the causes of these usability issues. An absence of appropriate feedback, guidance, and coherence, as well as the overall unfamiliarity of SST interfaces, were key contributors. The need to balance the cognitive load and the need to build upon the fragile trust users have with SSTs arose as critical components for improving these experiences.

Several limitations of our study should be acknowledged. Firstly, while every attempt was made to make our sample of SSTs as broad as possible, we only studied seven devices out of thousands in common use. Secondly, we studied singular interactions with devices, not repeated or longitudinal ones, which may have softened the significant unfamiliarity issues we observed but could also have surfaced additional usability challenges not encountered in fleeting interactions. Thirdly, and perhaps most importantly, our participants skewed young and university-educated, which would have significantly affected the profile of prior experience with different interface modalities. We were, however, able to demonstrate that there was no age-dependent effect on the SUS component of our results. Our finding that SST usability is overwhelmingly poor should be understood with these limitations in mind.

Our results suggest potential avenues for a further, more detailed assessment of SST usability as a starting point for developing sector-wide standards and recommendations for change. It would be illuminating to directly assess cognitive load during SST interactions, either through survey instruments like the TLX (Hart & Staveland, 1988) or through physiological measurements (Ayres et al., 2021). This could help to triangulate some of the findings about information overload and validate our suggestions around progressive disclosure. Additionally, the ISO standard for usability (ISO 9241-11) suggests measuring effectiveness (ie, task completion) and efficiency (ie, time spent) in addition to satisfaction, which we measured via SUS. Our hypothesis would be that task completion rates might explain the heteroskedasticity observed in our SUS scores, where lower-scoring technologies had higher variance because a portion of users experienced catastrophic usability failures. Time efficiency would be harder to compare across SST categories, but it might provide a useful metric when comparing multiple interfaces for performing the same task.

SST manufacturers may claim that there are legitimate commercial and contextual reasons as to why the usability issues described in this paper cannot be resolved. Combined with the belief that SSTs are living on borrowed time, this has led to a demonstrable disengagement with innovation by the industry. This disengagement is often justified by the perpetual belief that SSTs will imminently be replaced by personal technologies such as smartphones, a belief which has repeatedly failed to come to fruition. Instead, automation in the service and retail sectors continues to drive demand for new kinds of SST Wirtz et al. (2021).

Our study suggests that designers must stop viewing SSTs as devices not worthy of our attention and instead see them as the unique interaction modality they are. The notion that SSTs are an unwanted legacy technology – undeserving of design investment – is a tragedy. SSTs possess an untapped potential to deliver meaningful and rich experiences between organizations and users.

Note

1. It should be noted that Kortum and Bangor (2013) also assessed ATMs, and gave them a much higher SUS score of 82.3, but that in their study crowdsourced participants were asked to recall their interactions with broad technology categories, while in our study participants performed specific tasks with a specific technology. Recalling past interactions may have biased memories towards familiar and simple ATM tasks (like cash withdrawal) while our tasks (depositing a cheque) were likely both less familiar and more complex.

Acknowledgment

This research was partially funded through support from the Henry Halloran Trust.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abdelaziz, S. G., Hegazy, A. A., & Elabbassy, A. (2010). Study of airport self-service technology within experimental research of check-in techniques case study and concept. *International Journal of Computer Science Issues*, *7*(3), 30.

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Castro, D., Atkinson, R. D., & Ezell, S. J. (2010). Embracing the self-service economy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1590982

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. In: J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (Vol. 3, pp. 222–248). Sage.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE Publications. https://books.google.com.au/books?id=hZ6kBQAAQBAJ

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Diaper, D., & Stanton, N. (2003). *The handbook of task analysis for human-computer interaction*. CRC Press.

Ekşioğlu, M. (2016). User experience design of a prototype kiosk: A case for the Istanbul public transportation system. *International Journal of Human–Computer Interaction*, *32*(10), 802–813. https://doi.org/10.1080/10447318.2016.1199179

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Freeman, M. (1992). Clarence saunders: The piggly wiggly man. *Tennessee Historical Quarterly*, *51*(3), 161.

Freeman, M. (2019). *Clarence saunders & the founding of piggly wiggly: The rise & fall of a memphis maverick*. Arcadia Publishing.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Gómez-Carmona, O., Casado-Mansilla, D., & López-de Ipiña, D. (2018). Multifunctional interactive furniture for smart cities. *Multidisciplinary Digital Publishing Institute Proceedings*, *2*(19), 1212. https://doi.org/10.3390/proceedings2191212

Grimes, M. A. (1947). The legality of parking meter ordinances and permissible use of parking meter funds. *California Law Review*, *35*(2), 235. https://doi.org/10.2307/3477078

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Hartson, H. R., Andre, T. S., & Williges, R. C. (2001). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, *13*(4), 373–410. https://doi.org/10.1207/S15327590IJHC1304_03

Henderson, H., Leong, T. W. (2017). Lessons learned: A study on user difficulties with parking meters. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 533–537).

Hoffman, M. S. (2000). Ergonomic and usability issues with accessibility in the development of systmes for the public sector. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 5, p. 221). https://doi.org/10.1177/154193120004402958

Jenner, B., Flick, U., von Kardoff, E., & Steinke, I. (2004). *A companion to qualitative research*. Sage.

Khairat, S., Coleman, C., Newlin, T., Rand, V., Ottmar, P., Bice, T., & Carson, S. S. (2019). A mixed-methods evaluation framework for electronic health records usability studies. *Journal of Biomedical Informatics*, *94*, 103175. https://doi.org/10.1016/j.jbi.2019.103175

Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the system usability scale. *International Journal of Human-Computer Interaction*, *29*(2), 67–76. https://doi.org/10.1080/10447318.2012.681221

Kristoffersen, S., & Bratteberg, I. (2010). Design ideas for it in public spaces. *Personal and Ubiquitous Computing*, *14*(3), 271–286. https://doi.org/10.1007/s00779-009-0255-2

Lemke, J. L. (2012). Analyzing verbal data: Principles, methods, and problems. In *Second international handbook of science education* (pp. 1471–1484). Springer.

Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. *Journal of Usability Studies*, *13*(3), 158–167.

Liew, M. S., Zhang, J., See, J., & Ong, Y. L. (2019). Usability challenges for health and wellness mobile apps: Mixed-methods study among mhealth experts and consumers. *JMIR mHealth and uHealth*, *7*(1), e12160. https://doi.org/10.2196/12160

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33*(3), 338–348. https://doi.org/10.1145/77481.77486

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review*, *27*(3), 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001

Nielsen, J. (2006). *Progressive disclosure*. https://www.nngroup.com/articles/progressive-disclosure/

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. In Proceedings of the Sigchi Conference on Human Factors in Computing Systems (pp. 249–256). ACM.https://doi.org/10.1145/97243.97281

Norman, D. A. (1995). The psychopathology of everyday things. In: *Readings in human–computer interaction* (pp. 5–21). Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-051574-8.50006-6

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5 Pt 2), 1189–1208.

Peres, S. C., Pham, T., Phillips, R. (2013). Validation of the system usability scale (sus) sus in the wild. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 192–196. https://doi.org/10.1177/1541931213571043

Polacco, A., & Backes, K. (2018). The amazon go concept: Implications, applications, and sustainability. *Journal of Business and Management*, *24*(1), 79–92. https://doi.org/10.10.6347/JBM.201803_24(1).0004

Segrave, K. (2015). *Vending machines: An American social history*. McFarland.

Shneiderman, B., Plaisant, C., Cohen, M. S., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson.

Sung, H. J., & Jeon, H. M. (2020). Untact: Customer's acceptance intention toward robot barista in coffee shop. *Sustainability*, *12*(20), 8598. https://doi.org/10.3390/su12208598

Turner, J. J., & Szymkowiak, A. (2019). An analysis into early customer experiences of self-service checkouts: Lessons for improved usability. *Engineering Management in Production and Services*, *11*(1), 36–50. https://doi.org/10.2478/emj-2019-0003

Wirtz, J., Kunz, W., & Paluch, S. (2021). The service revolution, intelligent automation and service robots. *European Business Review*, *29*(5), 909.

About the authors

Hamish Henderson is a Designer and a Consultant Human-Factors Engineer. He is an Adjunct Senior Lecturer of Practice at the University of Sydney, in the School of Architecture, Design and Planning. As a Urban Interfaces Lab, he explores how we can design more usable and useful Self-Service Technologies.

Kazjon Grace is a Senior Lecturer in Design at the University of Sydney, and the Director of the Designing with AI Lab. His background is in HCI, AI, and the cognitive science of creativity and design. His research explores human-AI interaction, with a particular focus on computational tools for design.

Natalia Gulbransen-Diaz is a PhD candidate in the School of Architecture, Design and Planning at the University of Sydney. Her work explores design practice in non-profit organisations, with a focus on design integration and organisational change. Through her research, Natalia emphasises the value of design in realising diverse social outcomes.

Brittany Klaassens is a PhD candidate in the School of Architecture, Design and Planning at the University of Sydney. Her work explores productivity culture within higher education, with a focus on redesigning interactive productivity tools. Through her research, Brittany emphasises the significance that design brings to student-led educational experiences.

Leong Tuck Wah is a Human-Computer Interaction and Participatory Design researcher and scholar, who has worked in multidisciplinary teams across Europe and Australia. Tuck's interests include human values, human experiences of digital and innovating methods to support more effective ways to co-design with underserved and marginalized communities.

Martin Tomitsch is Head of the Transdisciplinary School at the University of Technology Sydney, a founding member of the Media Architecture Institute and the Urban Interfaces Lab, and holds an honorary position at the University of Sydney. His books include "Making Cities Smarter" and "Design Think Make Break Repeat".