

Usability and Behaviour Analysis of Prisoners using an Interactive Technology to Manage Daily Living

Kyle Boyd
Ulster University
Belfast, BT15 1ED
ka.boyd@ulster.ac.uk

Raymond Bond
Ulster University
Jordanstown, BT37 0QB
rb.bond@ulster.ac.uk

Stephen Gallagher
Ulster University
Coleraine, BT37 0QB
sm.gallagher@ulster.ac.uk

George Moore
Ulster University
Jordanstown, BT37 0QB
g.moore@ulster.ac.uk

Emer O'Kane
Core Systems, Belfast
emer@coresystems.biz

A computer system (Direct2Inmate) has been developed to assist prisoners to manage typical daily living tasks such as ordering meals, registering for educational programmes, making health appointments, personal entertainment and much more. The system is available worldwide via kiosks and tablet PCs. We investigate if this digital technology meets the needs of prisoners who have low computer and reading literacies. In considering a prisoner's persona, researchers have identified that emotions in prisons are volatile and can be heightened due to usability issues with interactive technologies, which can be disruptive and result in unwanted behaviours. With this in mind, we evaluated the system's user interface using usability testing and we recorded usability metrics in addition to the facial and verbal behaviours of prisoners whilst they interacted with the system.

Offender Technology, User Interfaces, User Experience, Usability Testing, Human-Computer Interaction

1. INTRODUCTION

Currently in the UK there are 85,641 prisoners in custody [1]. Studies suggest that between 22% and 47% of those prisoners do not have any formal qualifications [2], [3]. Furthermore, between 20-30% of prisoners have learning difficulties that affect their ability to cope within a criminal justice system. Indeed, it is estimated by the Prison Reform Trust that 60% of prisoners have a reading ability equivalent or less than that of a five year old child and around 40% of prisoners need specialist support for dyslexia [4]. Other sources state that 46% of people entering the prison system have literacy skills no higher than those broadly expected of an 11 year old child [5]. This prisoner persona has made it more challenging to develop an interactive technology that can be used to help prisoners manage typical daily living tasks such as sending social messages, making health appointments, participating in educational programmes, making electronic shop purchases, ordering meals, browsing and being entertained (Figures 1 and 2). Moreover, given emotions in prison are a key concern, sub-optimal usability of user interfaces can raise levels of frustration for the prisoner which can escalate to inappropriate behaviours and disruption [7,8,16,17]. Whilst the Direct2Inmate software was

designed for those with low literacy [9][10], conducting a usability experiment with prisoners who have never used the technology is necessary to measure its usability and to understand the problems prisoners have with interactive systems. We hypothesize that prisoners will require additional user guidance and that they will be more verbally and facially expressive when encountering usability issues in comparison to non-offenders who have standard literacy levels (a control group). This paper presents the work in progress of an experiment to test these hypotheses which to date involves undertaking a usability test of the Direct2Inmate system with prisoners and ex-prisoners.

2. DIRECT2INMATE TECHNOLOGY

The interactive prisoner technology called Direct2Inmate is a secure platform for prisoners to securely access information and services for themselves. It provides tools for prisoners to rehabilitate and successfully re-enter society through self-motivation.

The platform supports applications to provide prisoners with services such as electronic messaging, submitting requests/forms and shop ordering.

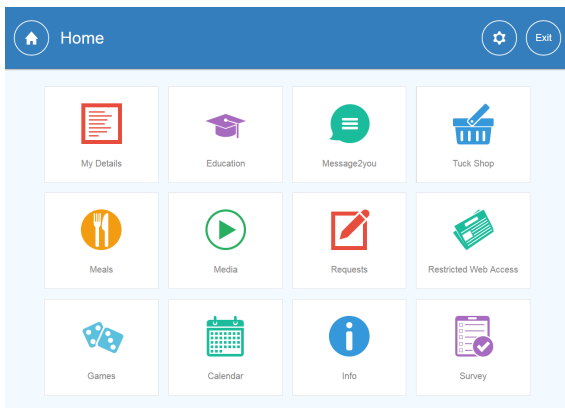


Figure 1: Main Dashboard of the Direct2Inmate Software

3. USABILITY METHODOLOGY

There are a number of methods that can be used to conduct a usability test. The most common approach is the concurrent 'think aloud' protocol (which was used in this study) where a participant verbalises their cognitive processes whilst they attempt a series of interactive tasks. This method helps demonstrate and highlight the usability issues being encountered as they interact with the system. The advantage of the think aloud protocol is that it offers a rapid approach to conducting and obtaining first hand insight into the thought processes associated with different tasks [8] [11]. The following tasks in this order were given to each user (verbally and in writing):

- (I) You want to buy some items from the shop. Please purchase 1 kit kat, 1 ice tea, 1 hand wash, 2 oranges
- (II) You have a headache. Please report this and ask to see a dentist.
- (III) You want to contact a friend. Send them a message.
- (IV) The food menu for your meals is available for week beginning 23rd May 2016 – make your selections
- (V) You want to sign up for a new education course – select one and enrol. (optional task)

The screen, audio and the user's facial expressions were recorded during their interactions with Direct2Inmate using screen-casting software. Eye tracking was also used to determine the visual hierarchy of the home screen. The Single Ease Question (SEQ) was asked before each task (how difficult do you expect the task to be?) and after each task (how difficult was the task?). SEQ is a 7-point rating scale to assess how difficult users find a task. Asking this before and after each task indicates if the system met the user's expectation. After each usability test is completed, each participant completed a post test usability questionnaire using

the Systematic Usability Scale (SUS) [13]. SUS is a tool for measuring usability and consists of a 10 item questionnaire with a five-scale Likert style response ranging from strongly agree to strongly disagree. After completion, a universal SUS score and percentile rank is given by benchmarking against a known distribution. Usability metrics were then computed.

2.1 Computed Usability Metrics

- (I) Time spent to accomplish each task (task completion times)
- (II) Frequency and severity of problems and usability errors participants encountered.
- (III) Successfully accomplished tasks (task completion rate)
- (IV) Un-successful task attempts (task failure rate)

For quantitative analysis, we used averages (mean and median), standard deviation and inter-quartile range. Hypothesis testing such as a t-test was used to test the differences between the pre-task and post-task SEQ scores which highlight whether any of the tasks did or did not meet the user's expectation.

The user videos were also qualitatively evaluated by a qualified behavioural analyst (SG) to assess behaviours from the facial and verbal data (14). The study was approved by Ulster University ethics filter committee.

4. RESULTS

The following are preliminary results from the usability test. A total of 15 participants (14 Males, 1 Female, mean age = 23.4±8.70) were recruited from a prison and a prisoner rehabilitation group in Northern Ireland. Educational levels of the subjects were low (10 had achieved levels less than high school and only 5 finished high school). Further subject profiling can be seen in Figure 3.

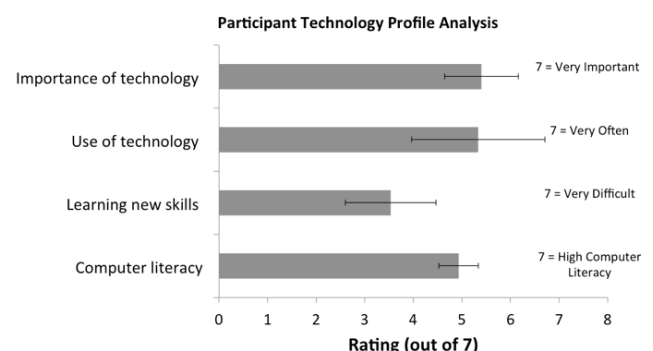


Figure 3: The participant technology profile analysis of those who took part in the study. Showing how users felt

about technology, its use, how difficult it was to learn and their own computer literacy from 0 to 7.

SUS scores can be seen in Figure 4. Mean SUS score was 80.8 (SD=14.93). This SUS score achieved a high percentile rank making it more user friendly than almost 90% of all other interfaces in the SUS distribution (achieving a B+ when grading on a curve).

Systematic Usability Scale Scores

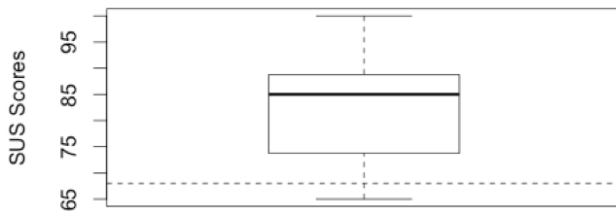


Figure 4: The systematic usability scale scores against the benchmark represented as a dashed line (68).

Task completion rates are: T1=93%, T2=87, T3=93%, T4=100%, t5=100%. task completion times for each task is shown in figure 5. as expected, shopping (t1) and meal selection (T2) tasks took the longest. Table 1 shows task completion times benchmarked against expert task completion times (the expert is the lead designer of Direct2Inmate). Task 2 took on average 4.79 times longer than the expert and task 1 took 4.18 times longer than an expert indicating that these tasks had some usability issues. Figure 6 also shows the SEQ ratings before and after each task. Interestingly, task completion rate, the difference between the mean subject task completion time and expert time along with the SEQ ratings all agree that task two has usability issues. This is interesting given this is seemingly a simple task (i.e. to contact a dentist using the system). This provides a form of cross validation using each of these usability metrics. Nevertheless, the pre- and post-task SEQ was not statistically significant ($p=0.36$) but had the greatest difference ($\Delta=0.86$).

Task Completion Times

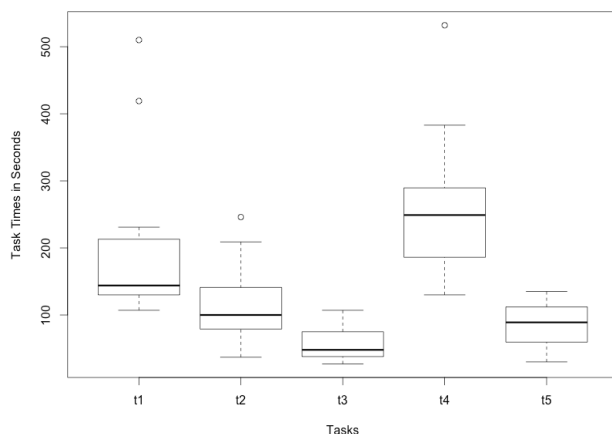


Figure 5: Task completion time box plots.

Table 1: Task completion times with benchmarking (task 5 is removed as only a small number of subjects attempted this optional task).

Task	Mean \pm SD	Expert Time	Delta	Factor
1	200.43 \pm 119.42	48	152.43	4.18
2	114.85 \pm 62.40	24	90.85	4.79
3	57.29 \pm 24.88	23	34.29	2.49
4	260.07 \pm 100.02	115	145.07	2.26

Single Ease Question Analysis

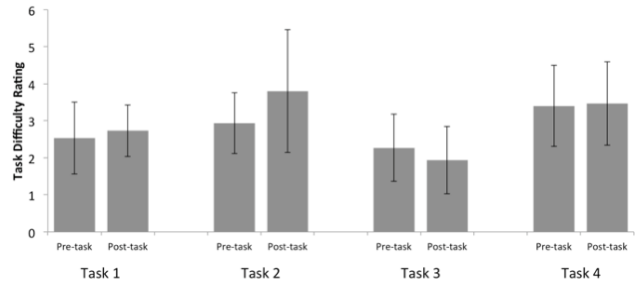


Figure 6: Single ease question ratings before and after each task (error bars=95% confidence intervals).

4.1 Observational Analysis and Usability Issues

One of the major issues was that prisoners were often left feeling uncertain at the end of a task as they expected user feedback even in cases where a designer might think that it is obvious that the user has completed a task. For example, in this experiment it was common for the user to express uncertainty after selecting all of their meals in task 4. Whilst it may seem obvious that a person has finished the task when they select their last meal for the last day, the prisoner still required reassurance that the task was completed and submitted. Such a usability issue could be solved using a placebo 'save/submit' button along with user feedback to mitigate uncertainty. Feedback should show users their location i.e. 'where I am, current status – what is happening, future status – what will happen next and outcomes and results – what just happened'. This study would support the need for 'micro' user feedback for prisoner-computer interaction. Another key observation was the number of typos committed when using the search engine to purchase items in the electronic tuck shop (task 1). This resulted in unwanted search results and frustration leading to time consumed tuck shop browsing to select desired products [15]. This was a common occurrence amongst prisoners and supports the need for more intelligent 'typo-friendly' search engines for low literate prisoners.

4.2 Behavioural Analysis

Behavioural patterns can act as indicators of frustration. For example, repeatedly pushing a button (which does not result in an expected outcome) is known as an extinction burst (18). We have all repeatedly pushed buttons on our television

remote control when the battery has depleted before we finally put the control down. Unfortunately, the extinction burst is often followed by extinction induced aggression, i.e., throwing the remote control across the room. In the case of the prisoner using the computer interface, the extinction induced aggression could be directed at the computer or other elements of the environment. Such patterns of responding could be mapped to increases in verbal aggression. In the present study, the usability of the interface was high enough to ensure that frustration did not reach such levels. Latency to respond, and increasing inter-response times (IRT) could also be indicative of the user not feeling competent in using the programme. Therefore extinction bursts, longer IRT, and increased latency could provide momentary data to suggest that prompts are required to avoid frustration. It was notable that when participants made an error that they could easily address, they did so with no signs of frustration. However, when multiple errors were made with short IRTs, frustration levels grew and were evidenced by verbal behaviours such as swearing, and exasperated noises. Multiple verbal behaviours with short IRTs were indicative of mounting frustration. For example, verbal behaviours increased when participants had difficulty locating shop items using the search engine. Subjects also displayed verbal behaviours when having difficulties caused by navigation issues. For example, participants became frustrated if they could not find a specific course title by manual search and therefore had to browse all courses instead, to find a similar course title. Generally, the participants performed well given that they had no instruction. Most issues involved the search engine, and navigational issues. Typical behavioural response patterns were evident and prove to be a useful metric in analysing competent use of an interface. Given the educational levels of the participants, computer literacy was quite high. This was evident across the various tasks. For example, online messaging and online shopping were fairly straightforward whereas tasks such as ordering meals proved more problematic. As a work in progress, we analysed the behaviours recorded in a subset of the videos (n=3). As a novel usability metric, we recorded the number of negative 'verbal outbursts' per task per subject. The number of verbal outbursts were: task 1=3, task 2=4, task 3=0 task 4=4. This would not normally be considered a metric given users are not normally verbally expressive but in this case they were. The number of verbal outbursts had concordance with the SEQ analysis (Figure 5) where task 3 had the best SEQ score and had no verbal outbursts yet task 1 had the worst SEQ score and had the most verbal outbursts. Also the number of use errors in this subset has concordance with SEQ and verbal outbursts where task 3 had the least errors (e=2) and task 2 had the most (e=8).

5. DISCUSSION AND FUTURE WORK

It could be seen that prisoners were very expressive both verbally and facially when interacting with the software, especially in cases of agitation and frustration. Given this attribute of the prisoner persona, there may be an opportunity to make use of facial expression based measures of affect during the design and day-to-day operation of the system. While the Think Aloud protocol is a widely respected and helpful design tool, it is also well established that the subjective comments elicited can be somewhat removed from their underlying causes, either consciously or subconsciously. Using facial expression analysis to help to determine the underlying emotions associated with tasks and reactions could prove highly valuable in forming a more reliable usability assessment protocol. Particularly given that the prisoners might not wish, or be able, to accurately articulate their feelings to others. There is a further opportunity in this regard, that goes beyond usability testing. There's a possibility to embed facial expression based affective computing algorithms within the system's normal operation, to help to detect when a prisoner is frustrated with a task and, for example, allow for a prison officer to intervene and provide assistance in real time. On-going and future work includes the recruitment of a control group (non-prisoners) to compare their usability metrics and emotional behaviours with the prisoner group. This will build evidence to show if prisoners are statistically more expressive, either due to their environment or persona.

6. CONCLUSION

This work evaluated the usability of a state of the art web based prisoner technology that assists inmates with typical daily living tasks. Whilst the usability of the system achieved a high SUS score, results show the need for more fine grain user feedback and typo-friendly search engines given the lack of literacy in the prisoner population. We found concordance between a series of usability metrics including task completion rate, Δ SEQs, verbal outbursts, the difference between the mean task completion time and expert times. An interesting observation was that prisoners were expressive during user interactions which provides an opportunity to implement affective computing (automatic facial expression analysis) to detect moments of frustration to adapt the user experience by providing real-time assistance to avoid unwanted behaviours. We acknowledge that this study is a work in progress and we are carrying out ongoing studies to better understand prisoner-computer interaction.

7. REFERENCES

1. "Prison population figures: 2015 - Publications - GOV.UK." [Online]. Available: <https://www.gov.uk/government/statistics/prison-population-figures-2015>. [Accessed: 29-Apr-2016].
2. P. R. Trust, "Prison: the Facts," 2013. [Online]. Available: <http://www.prisonreformtrust.org.uk/Portals/0/Documents/Prisonthefacts.pdf>. [Accessed: 29-Apr-2016].
3. UK Gov, "No Title," *Prior Qualifications of Adult OLASS learners*, 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/489417/bis-16-31-prior-qualifications-adult-olass-learners-2015.pdf. [Accessed: 05-May-2016].
4. "No One Knows - offenders with learning difficulties and learning disabilities." [Online]. Available: <http://www.prisonreformtrust.org.uk/uploads/documents/noknl.pdf>. [Accessed: 29-Apr-2016].
5. C. Taylor, "Prisoner education in the UK: a review of the evidence by Prisoners' Education Trust," *Prison Serv. Journal*, no. 223, pp. 44–51, 2016.
6. R. Metzner, "Through the Gateway: How Computers Can Transform Rehabilitation," 2013.
7. V. Knight, "Modus Vivendi : The cell , emotions , social relations and television," pp. 1–19, 2015.
8. A. Bruun, E. L.-C. Law, M. Heintz, and L. H. A. Alkly, "Understanding the Relationship Between Frustration and the Severity of Usability Problems: What Can Psychophysiological Data (Not) Tell Us?," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 3975–3987.
9. B. M. Chaudry, K. H. Connelly, K. A. Siek, and J. L. Welch, "Mobile interface design for low-literacy populations," *2nd ACM SIGHIT Int. Heal. informatics Symp.*, pp. 91–100, 2012.
10. N. Ahmad, U. Shoaib, and P. Prinetto, "Usability of Online Assistance From Semiliterate Users' Perspective," *Int. J. Hum. Comput. Interact.*, vol. 31, no. 1, pp. 55–64, 2015.
11. S. Krug, *Don't Make Me Think! A Common Sense Approach to Web Usability*, vol. Second Edi. Berkley: Newriders, 2009.
12. J. Sauro and J. R. Lewis, *Chapter 2 - Quantifying User Research*, vol. 11. 2012.
13. J. Brooke, "SUS - A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
14. Cooper J.O, Heron T.E, Heward W.L. *Applied behavior analysis (2nd ed.)* Upper Saddle River, NJ: Pearson; 2007.
15. Sun, Qin, and Nancy Spears. "Frustration Theory: toward an understanding of keyword search effectiveness and consumer responses." *Journal of Customer Behaviour* 10.1 (2011): 35-48.
16. Knight, Victoria. "Personal Control: Television, Emotion and Prison Life." *Remote Control*. Palgrave Macmillan UK, 2016. 121-152.
17. Crewe, B. and Bennett, J. eds., 2012. *The prisoner*. Routledge.
18. Lerman, D.C. and Iwata, B.A., 1995. Prevalence of the extinction burst and its attenuation during treatment. *Journal of Applied Behavior Analysis*, 28(1), pp.93-94.