Stats 230 - Final Project

Andrew Cambridge Kirsten Chapman Madison Wozniak

Sterling Bradshaw

***Abstract*—Abstract?**

***Index Terms*—Operating System, Android, iPhone, Human-Computer Interaction**

I. Introduction

The debate between which smartphone operating system is "best" can be quite contentious: with some phone users staunchly defending iOS and others fully supporting Android. As such, we seek out to quantify which operating system is truly optimal. For this study, we will be examining the ease of use that users have with both operating systems; looking specifically at the time it takes for them to complete phone-based tasks. Using this time data, we seek to determine which operating system is truly more 'user friendly' by examining the differences in speed between the two operating systems taking into account the participants' native operating system (i.e., the phone they currently use). Consequently, our research question is:

(RQ1a): Is Android or iOS the more user friendly operating system based on speed?

(RQ1b): How does a users' native operating system effect this?

As we begin our study, we expect that the effect of operating system on the time it takes to complete a task will change based on whether or not it is the user's native operating system.

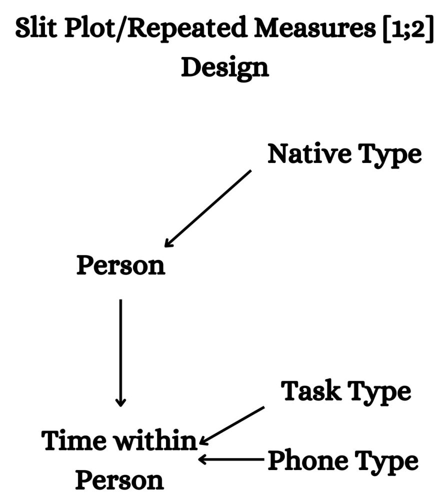

Fig. 1. The experimental design of our split plot/repeated measures study

II. Design and Data Collection

A. Experimental Design

The experimental design we used was a split-plot/repeated measures design with one between-unit factor and two within-unit factors (SP/RM[1;2]). Our large experimental unit was a person, and the between-units factor assigned to person was native phone type of the subject (iPhone, Android, or Other). The experimental unit within the person block is a time within a person, which was assigned a task type and a phone type. Our tasks had two levels, task 1 and task 2. Phone type also included two levels, iPhone and Android. Figure 1 shows the diagram of our experimental design.

B. Statistical Model

The model we used for the SP/RM[1;2] experiment was $Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \eta_l + (\alpha\gamma)_{ik} + (\alpha\eta)_{il} + (\gamma\eta)_{kl} + (\alpha\gamma\eta)_{ikl} + \epsilon_{ijkl}$ with Y_{ijkl} representing each observation, μ representing the grand mean, α_i representing the native phone type effect with two levels (iPhone, Android), $\beta_{j(i)}$ representing each person, γ_k representing the phone type effect with two levels (iPhone, Android), η_l representing the task type effect with two levels (screenshot task, settings task), $(\alpha\gamma)_{ik}$ representing the native phone type-phone type interaction effect, $(\alpha\eta)_{il}$ representing the native phone type-task type interaction effect, $(\gamma\eta)_{kl}$ representing the phone type-task type interaction, $(\alpha\gamma\eta)_{ikl}$ representing the 3-way interaction effect of all three factors, and ϵ_{ijkl} representing the residual. The hypotheses we want to test are the following:

1. $H_0 : \alpha_1 = \alpha_2$ vs. $H_A : \alpha_1 \neq \alpha_2$
2. $H_0 : \gamma_1 = \gamma_2$ vs. $H_A : \gamma_1 \neq \gamma_2$
3. $H_0 : \eta_1 = \eta_2$ vs. $H_A : \eta_1 \neq \eta_2$

Error		df	SS	MS	F-Stat	P-Val
Subject ID	**Native Brand**	1	37463	37463	3.347	0.0767
	Residuals	32	358199	11194		
Within	**Phone Type**	1	11711	11711	2.262	0.13591
	Task Type	1	173816	173816	33.567	8.72e-08 ***
	Native Brand:Phone Type	1	87459	87459	16.890	8.35e-05 ***
	Native Brand:Task Type	1	55408	55408	10.700	0.00149 **
	Phone Type:Task Type	1	51792	51792	10.002	0.00209 **
	Native Brand:Phone Type: Task Type	1	62093	62093	11.991	0.00080 ***
	Residuals	96	497109	5178		

TABLE I

$p < 0.05$: *; $p < 0.01$: **; $p < 0.001$: ***

4. $H_0 : (\alpha\gamma)_{11} = \cdots = (\alpha\gamma)_{22} = 0$ vs. H_A : At least one of the $(\alpha\gamma)_{ik}$ is different from the others
5. $H_0 : (\alpha\eta)_{11} = \cdots = (\alpha\eta)_{22} = 0$ vs. H_A : At least one of the $(\alpha\eta)_{il}$ is different from the others
6. $H_0 : (\gamma\eta)_{11} = \cdots = (\gamma\eta)_{22} = 0$ vs. H_A : At least one of the $(\gamma\eta)_{kl}$ is different from the others
7. $H_0 : (\alpha\gamma\eta)_{111} = \cdots = (\alpha\gamma\eta)_{222} = 0$ vs. H_A : At least one of the $(\alpha\gamma\eta)_{ikl}$ is different from the others

C. Power Analysis

Based on the 2 groups of between block factors (native_type with levels of Android and iOS), our 34 participants enable us to detect a difference of 30 seconds between software types, given expected standard deviation of 33 (anything beyond 100 seconds from the mean would be considered too weird) and a significance level of .05, with approximately 95% power. A significant issue with this experiment is that the actual standard error of the data was about 99.44. While our expected standard deviation indicates 95% power, power calculated with the standard error of the data is only 23%, not at all large enough to make any claims with certainty. The remainder of this report discusses significance of factor effects based on p-values, but be aware that the conclusions are grossly under powered.

D. Data Gathering Methodology

Data was gathered using the procedure outlined in our script (see Appendix). Each participant was given instructions about how the experiment would go. They were told not to use the search function, when to start, and what to say when they were done. Before each observation, the participant was allowed to review the task they were to perform. The person was then given a phone and a timer started until they completed the task (see the tasks listed in the script found in the Appendix).

Each of the tasks was done on each of the two phones. Both the order in which the phones were given and the order of the tasks within each phone were randomly assigned. The following code was used to create the random order for each participant, followed by how each number was interpreted. The randomization of how the phones and tasks were intermixed allowed us to randomize potentially confounding effects caused by the order of the phones or the order of the tasks, and thus manage and deal with variation caused by things such as learning curves and similarities or differences between tasks and phones. If a participant could not complete the task, their time was recorded as the maximum time of 10 minutes (600 seconds).

III. Data Analysis

Refer to table I for our Anova table.

A. ANOVA Assumptions

The normality assumption was checked both within subjects and between subjects. Graphs of each of their histograms indicate both error distributions resemble normal distributions. To check that the errors have common standard deviations, the standard deviations of both errors were calculated, with the resulting values: between subjects had a standard deviation of 103.485 and within-subject was 70.0545. Because the ratio of standard deviations is less than 2:1, they have similar standard deviations (code for these graphs and calculations can be found in the Appendix).

B. Factor Effects

With a .05 significance level, we see that all our within block factors other than the phone type were statistically significant. The Native Brand, a between block factor, was not significant. Thus, we fail to reject the null hypotheses that effect sizes are different from each other and zero for both the Native Brand main effect and Phone Type main effect. We reject the null hypotheses for all other main effects and interaction effects in the model.

We will examine exactly how much of the variability in the response variable was due to each factor individually using each factor's η^2.
$\eta^2_{task} \approx 0.19$ tells us that the task type factor accounts for about 19% of the variability in the response, $\eta^2_{native*phone} \approx 0.093$ tells us that the native type by

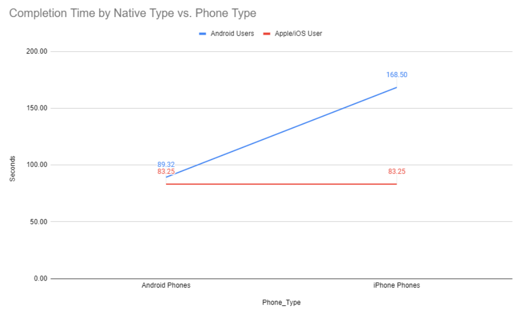

Fig. 2. The interaction between Native_Brand and Phone_Type

phone type interaction factor accounts for about 9% of the variability in the response, $\eta^2_{native*task} \approx 0.059$ tells us that the native type by task type factor accounts for about 6% of the variability in the response, $\eta^2_{phone*task} \approx 0.055$ tells us that the phone type by task type accounts for about 6% of the variability in the response, and $\eta^2_{native*phone*task} \approx 0.066$. tells us that the 3-way interaction accounts for about 7% of the variability in the response.

In response to our first research question, we conclude that with a p-value of 0.136 the phone type does not affect the time it takes for a subject to complete a task on the phone. Our other question was if the subjects native phone type would affect the response time for the different levels of phone type (or in other words, if there is a significant interaction between the native phone type and phone type factors). With a p-value of 0.000084, this interaction effect is definitely significant. From the $\eta^2_{native*phone}$ we see this interaction accounts for about 9% of the variability in the response variable. Figure 2 shows our interaction graph.

IV. Conclusion

Three key findings emerged from our data and analysis. Firstly, phone type alone was not a significant factor in the time it took a participant to complete the tasks. This indicates that with all else held constant, switching the phone that a participant was using for a certain task would have no significant effect on the time it took them to complete said task. Secondly, we found that the task types themselves were highly significant. Meaning that a change in task had a significant effect on the time it took to complete a task. Finally, we found that the interaction of phone type and native type was significant. This possibly indicates that a BYU student who uses an Android phone may find it harder to move to an iPhone than an iPhone user would find it to move to an Android.

These conclusions are supported by the p-values created in the anova table

Our study sample was not representative of the entire population, it primarily consisted of BYU students. As we used convenience sampling, we cannot make any solid inferences about the broader population. We can though state some cause-and-effect since we randomly assigned the order in that participants would complete the task.

In order to mitigate some of the limitations of our work, future studies should gather a larger, demographically representative sample. Increasing the sample size will also increase the power, and in turn, will allow us to detect smaller differences in the data (we can currently only detect a 30-second difference). This experiment could have been improved by tighter quality control of the treatment process and environment. Due to a larger standard deviation than expected, this experiment was also significantly under powered and would be able to yield reliable results given the increased sample sized mentioned above.

APPENDIX

a) Anova Code:

```
library(tidyverse)
# the below code sets the factors in the data
phones$Native_Brand ← as_factor(phones$Native_Brand)
phones$Phone_Type ← as_factor(phones$Phone_Type)
phones$Task_Type ← as_factor(phones$Task_Type)
phones$Subject_ID ← as_factor(phones$Subject_ID)

#performs the anova analysis on our statistical model and provides the anova table summary
phones.aov ← aov(Completion_Time ∼ Native_Brand + Error(Subject_ID) + Phone_Type + Task_Type + Native_Brand:Phone_Type + Native_Brand:Task_Type + Phone_Type:Task_Type + Native_Brand:Phone_Type:Task_Type, data=phones)
summary(phones.aov)
```

b) Power Test Code:

```
library(tidyverse)
#determines what level of power we will have with the given information
power.anova.test(groups=2, between.var=var(c(0,30)), within.var=(100/3)^2, sig.level=.05, n=34)
phones ← read_csv("Phones.csv")
sd(phones$Completion_Time)
```

c) Order Randomization Code:

```
sample(1:4,4)
Where:
1 = iPhone, Task 1
2 = iPhone, Task 2
3 = Android, Task 1
4 = Android, Task 2
```

d) Histograms:

```
# graphing histograms of both between and within residuals to check normality assumption
hist(phones.aov$Within$residuals)
hist(phones.aov$Subject_ID$residuals)
```

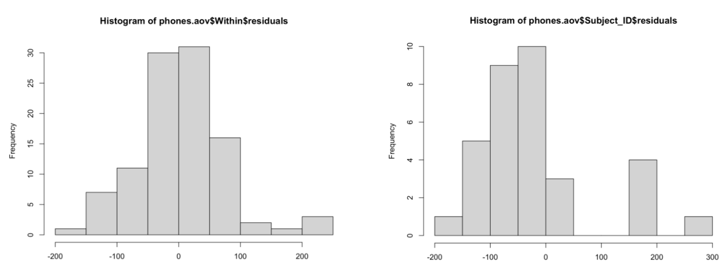

Fig. 3.

e) S Assumptions Code:

```
# compute standard deviations of within and between residuals for similar standard deviation assumptions
sd(phones.aov$Subject_ID$residuals)
```

sd(phones.aov$Within$residuals)

f) Protocol:

Introductory Script:

Determine the native type of the participant: On a day-to-day basis, what type of phone do you use? iPhone, Android, other. You will be given two working and unlocked phones and asked to complete two tasks on each of them. I will randomly give you one of the phones and then give you instructions on how to complete a randomly determined task. We will do this four times, such that you will complete each task on each phone. You may not use the search function to look for apps or settings. Please hold the phone screen so that I can see it and say "done" when you have completed the task. Here are the instructions for the first task. Please read them and then we will start in 30 seconds. Give the person the instruction sheet for a random task and start a 30 second timer. At the end of 30 seconds: Start the task when I say "go." Place the unlocked phone in front of the person and get your timer ready. "Go." Start the timer when you say go. Stop the timer when the person has said "done" and you can see that the task is complete. Follow along the entire time, without interrupting. If the person simply cannot complete the task or gives up, mark the time as "INF."

Task 1

Take a screenshot of the home screen. Save the screenshot to wherever the phone stores photos. Locate the image in the application used by the phone to store photos. Edit the image by applying a preset filter (does not have to be a specific filter, just any). Change the caption or title of the image to "you_are_awesome123!"

Task 2

Using the phone's settings application to change the time to 24-hour/military time format. Also in settings, turn the device into night/dark mode. Lastly, find the area of settings to manage battery health and turn the device onto power saving mode.

g) Data Sheet:

TABLE II
GENERATED BY SPREAD-LATEX

Observation_Number	Subject_ID	Proctor	Native_Brand	Phone_Type	Type_Match	Task_Type	Completion_Time
1	1	Andrew	Android	iPhone	false	Task 1 - screenshot	238
2	1	Andrew	Android	Android	true	Task 2 - settings	71
3	1	Andrew	Android	Android	true	Task 1 - screenshot	80
4	1	Andrew	Android	iPhone	false	Task 2 - settings	108
5	2	Andrew	Apple/iOS	Android	false	Task 1 - screenshot	300
6	2	Andrew	Apple/iOS	iPhone	true	Task 2 - settings	246
7	2	Andrew	Apple/iOS	iPhone	true	Task 1 - screenshot	162
8	2	Andrew	Apple/iOS	Android	false	Task 2 - settings	98
9	3	Andrew	Apple/iOS	iPhone	true	Task 1 - screenshot	93
10	3	Andrew	Apple/iOS	Android	false	Task 1 - screenshot	110
11	3	Andrew	Apple/iOS	Android	false	Task 2 - settings	93
12	3	Andrew	Apple/iOS	iPhone	true	Task 2 - settings	27
13	4	Andrew	Apple/iOS	iPhone	true	Task 2 - settings	41
14	4	Andrew	Apple/iOS	Android	false	Task 1 - screenshot	173
15	4	Andrew	Apple/iOS	Android	false	Task 2 - settings	114
16	4	Andrew	Apple/iOS	iPhone	true	Task 1 - screenshot	120
17	5	Kirsten	Apple/iOS	Android	false	Task 2 - settings	40
18	5	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	72
19	5	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	180
20	5	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	135
21	6	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	72
22	6	Kirsten	Apple/iOS	Android	false	Task 2 - settings	85
23	6	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	51
24	6	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	120
25	7	Kirsten	Apple/iOS	Android	false	Task 2 - settings	81
26	7	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	114
27	7	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	72
28	7	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	48
29	8	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	124
30	8	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	78
31	8	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	39
32	8	Kirsten	Apple/iOS	Android	false	Task 2 - settings	182
33	9	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	96
34	9	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	80
35	9	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	88
36	9	Kirsten	Apple/iOS	Android	false	Task 2 - settings	106
37	10	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	94
38	10	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	64
39	10	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	124
40	10	Kirsten	Apple/iOS	Android	false	Task 2 - settings	137
41	11	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	53
42	11	Sterling	Apple/iOS	Android	false	Task 2 - settings	62
43	11	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	170
44	11	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	200
45	12	Sterling	Android	iPhone	false	Task 2 - settings	60
46	12	Sterling	Android	Android	true	Task 2 - settings	72
47	12	Sterling	Android	Android	true	Task 1 - screenshot	179
48	12	Sterling	Android	iPhone	false	Task 1 - screenshot	600
49	13	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	36
50	13	Sterling	Apple/iOS	Android	false	Task 2 - settings	91
51	13	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	61
52	13	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	82
53	14	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	53
54	14	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	172
55	14	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	98
56	14	Sterling	Apple/iOS	Android	false	Task 2 - settings	135
57	15	Sterling	Android	iPhone	false	Task 2 - settings	55
58	15	Sterling	Android	Android	true	Task 2 - settings	73
59	15	Sterling	Android	Android	true	Task 1 - screenshot	74
60	15	Sterling	Android	iPhone	false	Task 1 - screenshot	120
61	16	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	62
62	16	Sterling	Apple/iOS	Android	false	Task 2 - settings	42
63	16	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	18
64	16	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	137
65	17	Sterling	Android	Android	true	Task 2 - settings	80
66	17	Sterling	Android	Android	true	Task 1 - screenshot	93
67	17	Sterling	Android	iPhone	false	Task 1 - screenshot	135
68	17	Sterling	Android	iPhone	false	Task 2 - settings	42
69	18	Sterling	Android	Android	true	Task 1 - screenshot	184
70	18	Sterling	Android	Android	true	Task 2 - settings	70
71	18	Sterling	Android	iPhone	false	Task 1 - screenshot	600
72	18	Sterling	Android	iPhone	false	Task 2 - settings	79
73	19	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	44
74	19	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	184
75	19	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	69
76	19	Sterling	Apple/iOS	Android	false	Task 2 - settings	60
77	20	Sterling	Apple/iOS	Android	false	Task 1 - screenshot	78
78	20	Sterling	Apple/iOS	Android	false	Task 2 - settings	52
79	20	Sterling	Apple/iOS	iPhone	true	Task 2 - settings	36
80	20	Sterling	Apple/iOS	iPhone	true	Task 1 - screenshot	184

TABLE III
GENERATED BY SPREAD-LATEX

Observation_Number	Subject_ID	Proctor	Native_Brand	Phone_Type	Type_Match	Task_Type	Completion_Time
81	21	Madison	Android	Android	true	Task 2 - settings	79
82	21	Madison	Android	iPhone	false	Task 2 - settings	47
83	21	Madison	Android	iPhone	false	Task 1 - screenshot	124
84	21	Madison	Android	Android	true	Task 1 - screenshot	59
85	22	Madison	Apple/iOS	iPhone	true	Task 2 - settings	47
86	22	Madison	Apple/iOS	Android	false	Task 1 - screenshot	61
87	22	Madison	Apple/iOS	Android	false	Task 2 - settings	58
88	22	Madison	Apple/iOS	iPhone	true	Task 1 - screenshot	53
89	23	Madison	Apple/iOS	Android	false	Task 1 - screenshot	91
90	23	Madison	Apple/iOS	Android	false	Task 2 - settings	70
91	23	Madison	Apple/iOS	iPhone	true	Task 1 - screenshot	49
92	23	Madison	Apple/iOS	iPhone	true	Task 2 - settings	31
93	24	Madison	Android	iPhone	false	Task 1 - screenshot	78
94	24	Madison	Android	Android	true	Task 1 - screenshot	55
95	24	Madison	Android	iPhone	false	Task 2 - settings	47
96	24	Madison	Android	Android	true	Task 2 - settings	39
97	25	Madison	Android	Android	true	Task 2 - settings	205
98	25	Madison	Android	iPhone	false	Task 2 - settings	133
99	25	Madison	Android	Android	true	Task 1 - screenshot	158
100	25	Madison	Android	iPhone	false	Task 1 - screenshot	600
101	26	Madison	Apple/iOS	iPhone	true	Task 1 - screenshot	118
102	26	Madison	Apple/iOS	Android	false	Task 2 - settings	138
103	26	Madison	Apple/iOS	Android	false	Task 1 - screenshot	101
104	26	Madison	Apple/iOS	iPhone	true	Task 2 - settings	59
105	27	Madison	Android	iPhone	false	Task 1 - screenshot	150
106	27	Madison	Android	iPhone	false	Task 2 - settings	72
107	27	Madison	Android	Android	true	Task 1 - screenshot	77
108	27	Madison	Android	Android	true	Task 2 - settings	92
109	28	Madison	Android	Android	true	Task 1 - screenshot	104
110	28	Madison	Android	iPhone	false	Task 1 - screenshot	231
111	28	Madison	Android	Android	true	Task 2 - settings	71
112	28	Madison	Android	iPhone	false	Task 2 - settings	37
113	29	Madison	Android	Android	true	Task 2 - settings	20
114	29	Madison	Android	iPhone	false	Task 1 - screenshot	600
115	29	Madison	Android	iPhone	false	Task 2 - settings	78
116	29	Madison	Android	Android	true	Task 1 - screenshot	139
117	30	Andrew	Apple/iOS	iPhone	true	Task 2 - settings	69
118	30	Andrew	Apple/iOS	Android	false	Task 1 - screenshot	110
119	30	Andrew	Apple/iOS	iPhone	true	Task 1 - screenshot	71
120	30	Andrew	Apple/iOS	Android	false	Task 2 - settings	79
121	31	Andrew	Android	iPhone	false	Task 2 - settings	59
122	31	Andrew	Android	iPhone	false	Task 1 - screenshot	178
123	31	Andrew	Android	Android	true	Task 1 - screenshot	86
124	31	Andrew	Android	Android	true	Task 2 - settings	73
125	32	Andrew	Android	Android	true	Task 1 - screenshot	84
126	32	Andrew	Android	iPhone	false	Task 1 - screenshot	77
127	32	Andrew	Android	Android	true	Task 2 - settings	45
128	32	Andrew	Android	iPhone	false	Task 2 - settings	25
129	33	Kirsten	Apple/iOS	iPhone	true	Task 2 - settings	72
130	33	Kirsten	Apple/iOS	iPhone	true	Task 1 - screenshot	103
131	33	Kirsten	Apple/iOS	Android	false	Task 1 - screenshot	122
132	33	Kirsten	Apple/iOS	Android	false	Task 2 - settings	75
133	34	Kirsten	Android	Android	true	Task 2 - settings	53
134	34	Kirsten	Android	iPhone	false	Task 1 - screenshot	97
135	34	Kirsten	Android	iPhone	false	Task 2 - settings	48
136	34	Kirsten	Android	Android	true	Task 1 - screenshot	86