

Introduction

SmartWatches are watches with touch interfaces that are worn on the wrists of users. Users of these devices have access to the internet and applications. Since these devices serve as a watch alternative, it is common for these devices to have under 2 inch screen sizes. With such a screen size, more traditional keyboard implementations needed to be tested against other implementations for usability improvements. This paper aims to examine two SmartWatch Keyboard implementations; a standard keyboard implementation and a zoom approach where the keyboard enlarges into the users specified input area. In total there were 3 independent variables of interest; device type, size and participant age. These variables were evaluated for their impact on participant error rates and time of task completion and the results are presented here using descriptive statistics and repeated task ANOVA.

Methodology

To explore the relationship between the two text entry implementations, the participant age and the input size on accuracy and time a full factorial within-subject design was presented on a touch screen device. There were 4 participants and three independent variables and with two levels for keyboard size (0.3, 0.5), device type (Normal, Zoom) and three levels for age (Teenage age<20, Young Adult age<40, Late Adult age<60). Totaling 12 conditions. Error was measured as a percentage over the total amount of inputs that incorrectly matched the target phrase. Time of task completion was evaluated in seconds. The tests were counterbalanced by splitting up the procedure for which the participant completed the tasks. There were 3 blocks where participants completed 6 tasks per block. Those were then divided into 3 different text entries were presented and two different device types and two sized screens. The aim of such a design is to minimize inconsistent results, to get good data from all the tasks and to accurately explore how the age reflected in the results. Following and during the study, users gave qualitative feedback about their experience. The target phrases reflected a full range of letters of the alphabet and were inspired by Prof. Scott MacKenzie target phrase report.

The experimental software was programmed with a menu (list of task buttons) where the user was asked to complete each task. The menu allowed for an easier navigational interface through the study and gave appropriate instructions for when the participant should take breaks between the trial blocks. Once the participant began to type, error rate and time of completion of task was taken. Error rate was calculated as measurement of the incorrect input characters over the total target input. The files were saved as txt file where userId, age, type of study, input phrase, target phrase, error rate, time in seconds was received.

The aim of such a design was to give the participants enough exposure to the text entry procedure so that the results stay consistent, to control for the effects of nuisance variables, to explore the relationship between how physical size keyboard differences relate to implementation (thus why 0.3, and 0.5 were chosen respectively) and see how the participant age reflects their speed and error.

Hypotheses

The following are hypotheses for the experiment and there justifications:

Null Hypothesis: There is a difference between input devices with respect to error rate for text entry (Fitt's Law)

Alternative hypothesis: There are no differences between the input devices with respect to error rate for text entry.

Goal of experiment: To test whether there is a significant differences in error of text entry experiments.

Justification: The zoom board is novel implementation that allows for users for more area to select characters that increases the width and minimizes the distance of the target character/

Null Hypothesis: There is no difference between input devices with respect to time for text entry

Alternative hypothesis: There are no differences between the input devices with respect to time for text entry.

Goal of experiment: To test whether there is a significant differences in error of text entry experiments over task time.

Justification: The zoom board may require a learning curve to use given that its a novel approach to text entry that participants are not familiar with, however, given the small screen users may be more careful in input selection using the normal implementation

Null Hypothesis: Age has a negative linear effect on text entry error rate and time for task completion.

Alternative hypothesis: Age has a neutral or positive linear effect on text entry error rate and time for task completion.

Goal of experiment: To test how differences age impact the dependant variables.

Justification: The older one is, generally, the less exposed they are to using touch interfaces.

Null Hypothesis: There are no differences between the size of a SmartWatch keyboard within the general boundaries of SmartWatch sizes

Alternative hypothesis: There are negative differences between the size of a SmartWatch keyboard within the general boundaries of SmartWatch sizes

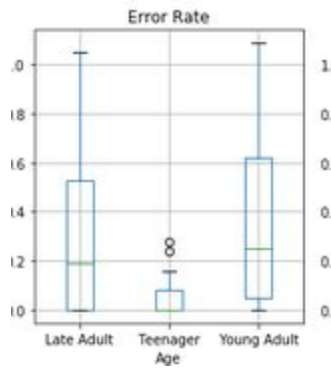
Goal of experiment: To test how the physical implementation size of a keyboard effects accuracy and time

Justification: Differences in SmartWatch sizes are not very drastic or as familiar, thus would not impact the accuracy and time significantly

Results and Analysis

The analysis presented several important insights. Generally for error rate, the mean error rate was roughly 25% with a 30% standard deviation, a minimum of 0 and a maximum of 108%. For the time taken to complete each task that mean was 42.17 seconds with a standard deviation of 24.8 seconds with the fastest time taking 10 seconds per task and the slowest taking 113 seconds.

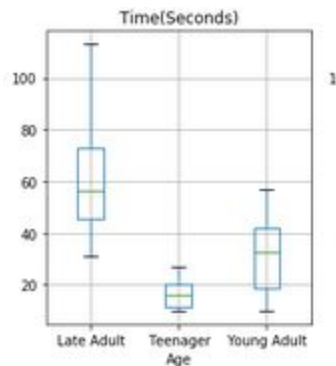
In creating a boxplot to compare the three independent variables and error, by observation, the teenager performed much better than the other two age groups.



There is no significant difference in the error mean of the teenage group against the other two groups($P=0.33$), however, the variances in input errors are much drastic.

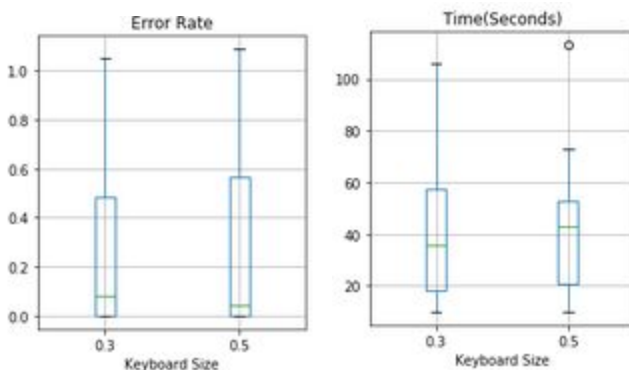
There is a significant difference between all groups with respect to time completion. Although the young adult made similar errors, their performance was a lot faster. The teenage group performed the fastest with a p-value of 0 when compared with the Late Adult group and a P-value of 0.00045 with the Young Adult.

There is also a significant difference between the Young Adult and Late Adult group ($P=0.005$).



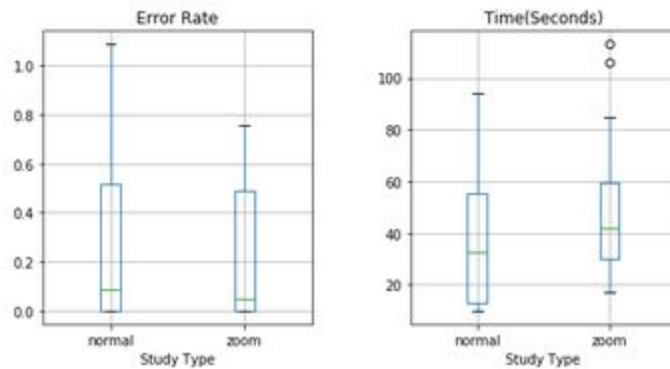
	N	Mean	SD	SE	95% Conf. Interval
Age					
Late Adult	36	0.303083	0.311587	0.051931	0.201298 0.404869
Teenager	18	0.054389	0.086407	0.020366	0.014471 0.094307
Young Adult	18	0.353722	0.364088	0.085816	0.185522 0.521922

There was no significant difference between the error rate and time with respect to the physical keyboard size as suspected.



	N	Mean	SD	SE	95% Conf. Interval
Keyboard Size					
0.3	48	0.249958	0.284844	0.041114	0.169375 0.330541
0.5	24	0.260792	0.357900	0.073056	0.117602 0.403982

Between the device types there were no significant differences between the error rates and time of task completion.



	N	Mean	SD	SE	95% Conf.	Interval
Study Type						
normal	36	0.271361	0.340873	0.056812	0.160009	0.382713
zoom	36	0.235778	0.276293	0.046049	0.145522	0.326033

Doing a full descriptive analysis of the relationship of all the independent variables and error rate and time respectively we receive:

Study Type	Keyboard Size	Age	N	Mean	SD	SE	95% Conf.	Interval
normal	0.3	Late Adult	12	0.353500	0.330006	0.095264	0.166782	0.540218
		Teenager	6	0.061000	0.097616	0.039851	0.002891	0.159109
		Young Adult	6	0.130000	0.241937	0.098770	-0.063590	0.323590
	0.5	Late Adult	6	0.231833	0.368516	0.150446	-0.063041	0.526708
		Teenager	3	0.058000	0.050229	0.029000	0.001160	0.114840
		Young Adult	3	0.898667	0.223290	0.128916	0.645990	1.151343
zoom	0.3	Late Adult	12	0.334833	0.288326	0.083232	0.171698	0.497969
		Teenager	6	0.046000	0.112677	0.046000	-0.044160	0.136160
		Young Adult	6	0.366000	0.314361	0.128337	0.114459	0.617541
	0.5	Late Adult	6	0.210000	0.310348	0.126699	-0.038330	0.458330
		Teenager	3	0.014333	0.024826	0.014333	-0.013760	0.042427
		Young Adult	3	0.231667	0.255171	0.147323	-0.057087	0.520420

Study Type	Keyboard Size	Age	N	Mean	SD	SE	95% Conf.	Interval
normal	0.3	Late Adult	12	59.250000	19.484843	5.624790	48.225412	70.274588
		Teenager	6	11.666667	1.966384	0.802773	10.093232	13.240102
		Young Adult	6	16.000000	4.857983	1.983263	12.112804	19.887196
	0.5	Late Adult	6	51.833333	11.703276	4.777842	42.468762	61.197904
		Teenager	3	11.333333	1.154701	0.666667	10.026667	12.640000
		Young Adult	3	36.666667	18.770544	10.837179	15.425796	57.907537
zoom	0.3	Late Adult	12	63.833333	20.984121	6.057594	51.960449	75.706217
		Teenager	6	22.833333	4.400758	1.796602	19.311994	26.354673
		Young Adult	6	37.166667	6.177918	2.522124	32.223303	42.110030
	0.5	Late Adult	6	62.833333	27.686940	11.303146	40.679167	84.987500
		Teenager	3	19.000000	1.732051	1.000000	17.040000	20.960000
		Young Adult	3	48.000000	5.291503	3.055050	42.012101	53.987899

To complete a two way repeated measure ANOVA test, I compared the keyboard sizes and study types with respect to time and error rate respectively. There was no significant variation amongst the groups between these groups.

Time (Seconds):

=====

ANOVA SUMMARY

=====

Source	SS	ddof1	ddof2	MS	F	p-unc	p-GG-corr	np2	eps
Study Type	425.391	1	3	425.391	6.652	0.082	0.082	0.689	1.000
Keyboard Size	6.891	1	3	6.891	0.059	0.823	0.823	0.019	1.000
Study Type * Keyboard Size	0.016	1	3	0.016	0.000	0.986	0.986	0.000	1.000

Error rate:

=====

ANOVA SUMMARY

=====

Source	SS	ddof1	ddof2	MS	F	p-unc	p-GG-corr	np2	eps
Study Type	0.022	1	3	0.022	2.416	0.218	0.218	0.446	1.000
Keyboard Size	0.000	1	3	0.000	0.009	0.929	0.929	0.003	1.000
Study Type * Keyboard Size	0.053	1	3	0.053	1.034	0.384	0.384	0.256	1.000

ANOVA uses the F-test to determine whether the variances between group means is larger than the variances of the observations within the groups. In the data we can conclude that populations come from populations with the same variance for both error rate and time taken for both the independent variables.

Discussion

Before discussing the results and their relation to the hypothesis presented, it is important to note that the study had only 4 total participants for the study. You typically would want 12 participants for HCI studies to get appropriate outcome results. Given that we also had three types of age groups, it could be argued that there should be 12 participants per age group as too get a further understanding of population outcomes. Other limitations of the study also included not testing for how the participants were seated when completing the study or doing a survey of the participants regarding their familiarity to touch devices, text entry systems and related technologies. These influence both the hypotheses and outcomes of the research and will be considered in the discussion.

There were some striking results based on the outcomes of the study. For one, age does not necessarily have a negative linear effect on text entry errors but it does have an effect on time. The teenage group performed significantly better, which is consistent with the null hypothesis, however, there are not significant differences between the late and young adult group, especially due to error rate, which goes against the idea of a linear negative relationship (increased error rate and time the older the age). I believe I have attained the results I did given the limited amount of study participants and exposure to a survey. The one participant in the young adult category may have desired to finish as quickly as possible, but not as accurately.

The physical SmartWatch size was also interesting. It was consistent with the null hypothesis that there are no significant differences, but if we break it down to the young adult group, the smaller screen size had far better results in terms of mean time than the larger counterpart (P value = 0.0022). Thus, the hypothesis was too general to account for group differences between means. This was similar with respect to the device type. This was only true for teenagers, and not the other groups. It is difficult to explain why this is the case, possibly due to the device the user was using (different smartphone) or because the user is familiar with using a smaller text entry keyboard.

The other hypothesis were consistent with the results that were found, with smaller variances amongst each individual group. Doing ANOVA with the within group variances in the trial showed that variances between the two independent variables with respect to time and error rate. Given that there were multiple variables being tested, for some studies such as the age group difference with respect to time, a between study may have been more fruitful given that we could test them simultaneously. This would be preferable if there were more test subjects.

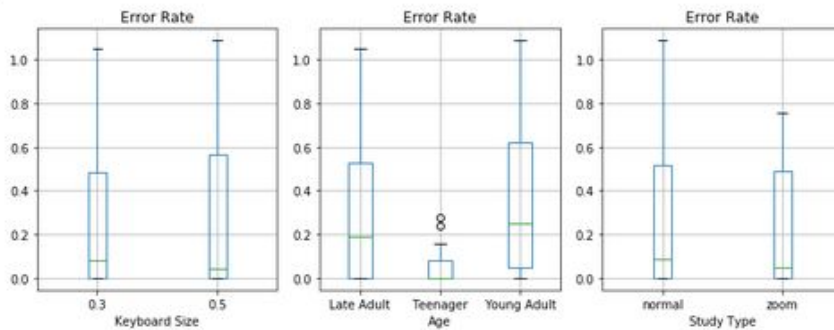
Conclusion

As a result, the testing shed some interesting insights with relation to the test devices, the keyboard size and age of participants with respect to error rate and time. Based on the results, testing between group for age related and device type difference may be promising as well as a different approach to the counter balancing technique used, where there are less tasks but larger difference between the tasks themselves in terms of design, and the study is counter balanced so that the participant does a different type of task then the other every third time to account for the age.

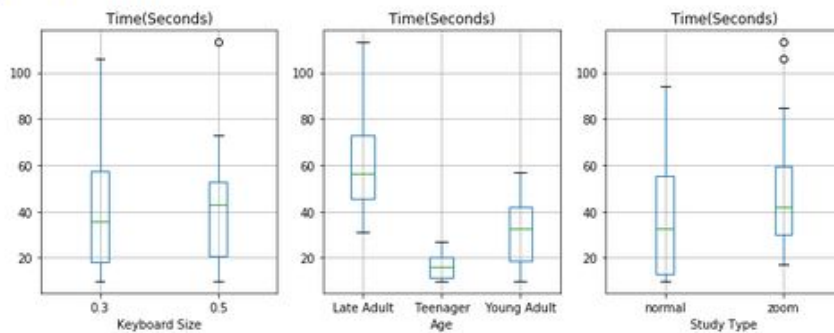
Appendix

Descriptive Statistics

```
2]: import matplotlib.pyplot as plt
df["Keyboard Size"] = df["Keyboard Size"].astype(float)
df["Time(Seconds)"] = df["Time(Seconds)"].astype(float)
df["Error Rate"] = df["Error Rate"].astype(float)
fig, (ax1,ax2,ax3) = plt.subplots(1,3, figsize=(10,4)) # 1 row, 3 columns
df.boxplot('Error Rate', by='Keyboard Size', ax=ax1)
df.boxplot('Error Rate', by='Age', ax=ax2)
df.boxplot('Error Rate', by='Study Type', ax=ax3)
plt.suptitle("")
plt.tight_layout()
```



```
3]: fig, (ax1,ax2,ax3) = plt.subplots(1,3, figsize=(10,4)) # 1 row, 3 columns
df.boxplot('Time(Seconds)', by='Keyboard Size', ax=ax1)
df.boxplot('Time(Seconds)', by='Age', ax=ax2)
df.boxplot('Time(Seconds)', by='Study Type', ax=ax3)
plt.suptitle("")
plt.tight_layout()
```



```
6]: #Descriptive Statistics
df['Error Rate'].describe()
```

```
6]: count    72.000000
   mean      0.253569
   std       0.308595
   min       0.000000
   25%       0.000000
   50%       0.053000
   75%       0.517000
   max       1.087000
   Name: Error Rate, dtype: float64
```

```
: df['Time(Seconds)'].describe()
```

```
: count      72.000000
   mean       42.166667
   std        24.817077
   min        10.000000
   25%        19.500000
   50%        41.500000
   75%        57.000000
   max       113.000000
   Name: Time(Seconds), dtype: float64
```

Out[70]:

	N	Mean	SD	SE	95% Conf. Interval
Age					
Late Adult	36	0.303083	0.311587	0.051931	0.201298 0.404869
Teenager	18	0.054389	0.086407	0.020366	0.014471 0.094307
Young Adult	18	0.353722	0.364088	0.088816	0.185522 0.521922

```
In [71]: xp.summary_cont(df.groupby(['Keyboard Size']))['Error Rate']
```

Out[71]:

	N	Mean	SD	SE	95% Conf. Interval
Keyboard Size					
0.3	48	0.249958	0.284844	0.041114	0.169375 0.330541
0.5	24	0.260792	0.357900	0.073056	0.117602 0.403982

```
In [72]: xp.summary_cont(df.groupby(['Study Type']))['Error Rate']
```

Out[72]:

	N	Mean	SD	SE	95% Conf. Interval
Study Type					
normal	36	0.271361	0.340873	0.056812	0.160009 0.382713
zoom	36	0.235778	0.276293	0.046049	0.145522 0.326033

```
In [75]: xp.summary_cont(df.groupby(['Study Type', 'Keyboard Size', 'Age']))['Error Rate']
```


Out[75]:

Study Type	Keyboard Size	Age	N	Mean	SD	SE	95% Conf. Interval
normal	0.3	Late Adult	12	0.353500	0.330006	0.095264	0.166782 0.540218
		Teenager	6	0.081000	0.097616	0.039651	0.002891 0.159109
		Young Adult	6	0.130000	0.241937	0.098770	-0.063590 0.323590
	0.5	Late Adult	6	0.231833	0.368516	0.150446	-0.063041 0.526708
		Teenager	3	0.058000	0.050229	0.029000	0.001160 0.114840
		Young Adult	3	0.898667	0.223290	0.128916	0.645990 1.151343
zoom	0.3	Late Adult	12	0.334833	0.288326	0.083232	0.171698 0.497969
		Teenager	6	0.046000	0.112677	0.046000	-0.044160 0.136160
		Young Adult	6	0.366000	0.314361	0.128337	0.114459 0.617541
	0.5	Late Adult	6	0.210000	0.310348	0.126699	-0.038330 0.458330
		Teenager	3	0.014333	0.024826	0.014333	-0.013760 0.042427
		Young Adult	3	0.231667	0.255171	0.147323	-0.057087 0.520420

In [77]: `rp.summary_cont(df.groupby(['Age']))['Time(Seconds)']`

Out[77]:

	N	Mean	SD	SE	95% Conf. Interval
Age					
Late Adult	36	60.138889	20.119859	3.353310	53.566402 66.711376
Teenager	18	16.555556	5.962848	1.405457	13.800860 19.310251
Young Adult	18	31.833333	14.549105	3.429257	25.111990 38.554677

In [78]: `rp.summary_cont(df.groupby(['Study Type']))['Time(Seconds)']`

Out[78]:

	N	Mean	SD	SE	95% Conf. Interval
Study Type					
normal	36	37.000000	24.750180	4.125030	28.914941 45.085059
zoom	36	47.333333	24.122307	4.020385	39.453380 55.213287

In [79]: `rp.summary_cont(df.groupby(['Keyboard Size']))['Time(Seconds)']`

	N	Mean	SD	SE	95% Conf.	Interval
Keyboard Size						
0.3	48	41.729167	25.538240	3.686127	34.504357	48.953976
0.5	24	43.041667	23.817201	4.861666	33.512802	52.570532

```
rp.summary_cont(df.groupby(['Study Type', 'Keyboard Size', 'Age']))['Time(Seconds)']
```

			N	Mean	SD	SE	95% Conf.	Interval
Study Type	Keyboard Size	Age						
normal	0.3	Late Adult	12	59.250000	19.484843	5.624790	48.225412	70.274588
		Teenager	6	11.666667	1.966384	0.802773	10.093232	13.240102
		Young Adult	6	16.000000	4.857983	1.983263	12.112804	19.887196
	0.5	Late Adult	6	51.833333	11.703276	4.777842	42.468762	61.197904
		Teenager	3	11.333333	1.154701	0.666667	10.026667	12.640000
		Young Adult	3	36.666667	18.770544	10.837179	15.425796	57.907537
zoom	0.3	Late Adult	12	63.833333	20.984121	6.057594	51.960449	75.706217
		Teenager	6	22.833333	4.400758	1.796602	19.311994	26.354673
		Young Adult	6	37.166667	6.177918	2.522124	32.223303	42.110030
	0.5	Late Adult	6	62.833333	27.686940	11.303146	40.679167	84.987500
		Teenager	3	19.000000	1.732051	1.000000	17.040000	20.960000
		Young Adult	3	48.000000	5.291503	3.055050	42.012101	53.987899

TWO - WAY REPEATED MEASURE ANOVA

<
>

Source	SS	ddf1	ddf2	MS	F	p-unc	p-GG-corr	np2	eps
Study Type	425.391	1	3	425.391	6.652	0.082	0.082	0.689	1.000
Keyboard Size	6.891	1	3	6.891	0.059	0.823	0.823	0.019	1.000
Study Type * Keyboard Size	0.016	1	3	0.016	0.000	0.986	0.986	0.000	1.000

Source	SS	ddof1	ddof2	MS	F	p-unc	p-GG-corr	np2	eps
Study Type	0.022	1	3	0.022	2.416	0.218	0.218	0.446	1.000
Keyboard Size	0.000	1	3	0.000	0.009	0.929	0.929	0.003	1.000
Study Type + Keyboard Size	0.053	1	3	0.053	1.034	0.384	0.384	0.256	1.000

Logged data:

[illegible]

Consent Forms:

STUDY PROTOCOL TEMPLATE

Project Title: SmartWatch Text Entry Research Study

Investigators: Matthew Byra matthew.byra@mail.utoronto.ca

Background and purpose of Research: The purpose of our study is to understand how users of all ages help us derive requirements for the design of novel interactive computational media that are intended to be useful to smartwatch users. A brief description of our design concept is: to compare and analyze smartwatch interactions with device size, device type and age.

Participant selection and eligibility: Participants will be chosen from teenagers, young adults and middle aged individuals. They will be identified via survey and selected according to age requirements being met. In general, they will be characterized by age.

Procedure: We will brief participants about the purpose of the study, explain the attached consent form to them, and ensure that they consent to participate and sign the consent form. We will then engage the participants in within subject user testing. We will also with their permission make observations as follows: error rate, time taken to complete task, age and screen size.

Voluntary Participation & Early Withdrawal: The participation in this study is entirely voluntary, and participants are free to cease participation at any time, for any reason, without the need to give any explanation. At their request, we will delete any of their data and it will not be used in our analysis or any subsequent reports or presentations.

Relationships: Our relationship to the participants may be described as follows: friends and family.

Risk and benefit: There are no anticipated risks associated with participation in this study, beyond those associated with everyday use of computer (e.g. participants may feel that they have wasted their time). The only benefit will be to contribute to the education of the investigators.

Compensation: Participants will receive no compensation.

Information sought: The information to be sought is described in the attached application which has the implemented research study.

Privacy and confidentiality: Information will be kept confidential by the investigators. Names or other identifying or identified information will not be kept with the data. The only other use will be to include excerpts or copies in the assignment submitted, but names and other identifying or identified information will not be submitted.

CONSENT FORM TEMPLATE

Consent Form: SmartWatch Text Entry Research Study

I hereby consent to participate in a study conducted by Matthew Byra for an assignment in University of Toronto Computer Science 428, Human-Computer Interaction.

I agree to participate in this study the purpose of which is to analyze the effects of smartwatch text entry interactions, device size and participant age.

I understand that

- the procedures to be used are within subject full factorial application study.
- I will receive no compensation for my participation.
- I am free to withdraw before or any time during the study without the need to give any explanation.
- all materials and results will be kept confidential, and, in particular, that my name and any identifying or identified information will not be associated with the data.

Naomi Robinson
Participant's Printed Name

Naomi Robinson
Participant's Signature

09-11-2019
Date

MATTHEW BYRA
Experimenter Name

Mh
Experimenter's Signature

CONSENT FORM TEMPLATE

Consent Form: SmartWatch Text Entry Research Study

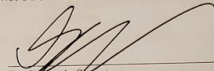
I hereby consent to participate in a study conducted by Matthew Byra for an assignment in University of Toronto Computer Science 428, Human-Computer Interaction.

I agree to participate in this study the purpose of which is to analyze the effects of smartwatch text entry interactions, device size and participant age.

I understand that

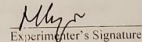
- the procedures to be used are within subject: full factorial application study.
- I will receive no compensation for my participation.
- I am free to withdraw before or any time during the study without the need to give any explanation.
- all materials and results will be kept confidential, and, in particular, that my name and any identifying or identified information will not be associated with the data.

Patrick Byra
Participant's Printed Name


Participant's Signature

09/11/19
Date

MATTHEW BYRA
Experimenter Name


Experimenter's Signature

CONSENT FORM TEMPLATE

Consent Form: SmartWatch Text Entry Research Study

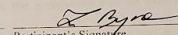
I hereby consent to participate in a study conducted by Matthew Byra for an assignment in University of Toronto Computer Science 428, Human-Computer Interaction.

I agree to participate in this study the purpose of which is to analyze the effects of smartwatch text entry interactions, device size and participant age.

I understand that

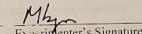
- the procedures to be used are within subject: full factorial application study.
- I will receive no compensation for my participation.
- I am free to withdraw before or any time during the study without the need to give any explanation.
- all materials and results will be kept confidential, and, in particular, that my name and any identifying or identified information will not be associated with the data.

ZOEIA BYRA
Participant's Printed Name


Participant's Signature

09-11-2019
Date

MATTHEW BYRA
Experimenter Name


Experimenter's Signature