Controlled Study of Click Keyboards:

exploring the alphabet and phone pad

D101-Team 3

Anna Koniarska
31-1947 Pendrell St.
V6G 1T5
Vancouver BC

Brennan Harcus
14868 97th Ave
V3R 9V3
Surrey, BC

Ethan Johanson
9988 149 St
V3R 7W7
Surrey, BC

Abstract

A controlled experiment was conducted, comparing two different types of virtual keyboard, an alphabetic and a phone pad type. This experiment was done to assess which of the two would be preferable, if at all, in certain contexts. We hypothesized that the alphabetic keyboard would outperform the phone pad keyboard in speed and input accuracy. In our study, we used a program called kb.tcl which had different keyboard layouts that could easily be switched. The program also had the ability to record participants speed and input which was favourable for record keeping and analysis. Data files collected after each participant were stored and collected in excel. Questionnaires throughout the study were also distributed and filled out by participants. Results showed that there was no significant difference between the two keyboard layouts in speed, however, the phonepad keyboard had more errors which is a highly significant difference. This suggests that the alphabetic keyboard may be preferable as it produces less errors or a redesign of the phone pad keyboard. Also, we noticed some potential bias and concerns that can be found in the rest of the paper.

1. Introduction

We investigated two click keyboard interfaces, a phone pad keyboard and an alphabetic keyboard and tried to determine which of the two is better for use in a system. Most users are familiar with a standard qwerty keyboard, unlike the qwerty keyboard which is found on the majority of devices such as computers, the alphabetic and phone pad keyboard is more often found on portable devices such as phones which may be familiar to some users, in recent years there have been shifts to qwerty keyboards. Thus, it is important that we determine if the phone pad or alphabetic keyboards differ in anyway in regards to efficiency and accuracy when choosing between the two.

In section two of this document you can find a description of the experiment, which includes, the hypothesis' and null-hypothesis' used, participant descriptions, materials, potential bias and the method of the experiment. Section three is the results of the study, while Section four is our discussion which has an analysis of our outcomes and critical reflection. Section five is our conclusion and lists the outcomes of the study.

2. Description of the experiment

2.1 Introduction and hypotheses

As a group, we hypothesized that a virtual alphabetical keyboard is more efficient in terms of user accuracy and speed of typing, as opposed to the use of a virtual phone pad style keyboard.

Hypothesis:
-There is a significant usage speed difference between an alphabetical style and a phone pad style virtual keyboard.
-There is a significant error rate difference between an alphabetical style and a phone pad style virtual keyboard.

Null-hypothesis:
-There is no significant difference in usage speed between an alphabetical style and a phone pad style virtual keyboard.
-There is no significant difference in error rate between an alphabetical style and a phone pad style virtual keyboard.

Our experiment consisted of testing three participants, each of them fairly familiar with similar

keyboards. Our experiment consisted of having participants write a series of sentences with each typing layout, with the backspace or delete button removed in order to record errors. Observers would then compile data on the time taken via recording software, as well as note any errors made. We provided our participants with a pre-questionnaire , a post-questionnaire, as well as a brief detailing of the experiment. This experiment was repeated by other groups and then data was compiled.

2.2 Participants

Our demographic and sample contained technology savvy students, in or near their final year of study at Simon Fraser University. Participants were divided into two groups, one of which started with the phonepad keyboard followed by the alphabet keyboard, while the other set started with the alphabetic keyboard first. Our data was provided to us by our classmates in IAT 432 and Teaching Assistant.

2.3 Materials

Each test required a computer running the Windows operating system, although the utilized software was available for Mac OSX, we chose to run the test with the Windows version as it was more stable. We also required a fully functioning mouse peripheral for the computer.

Installed on each computer was the Tcl/TK environment,the programming language necessary to run the software and self installing script. Also, the program kb.tcl, available at http://carmster.com/432/ControlledStudy/kb.tcl, was installed, which contained the sample keyboards to be tested.

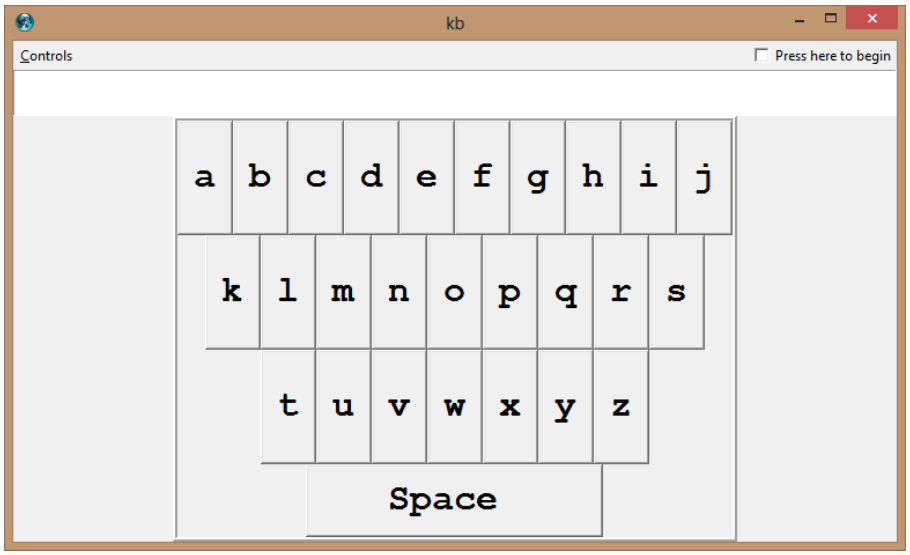

Fig 1. *Alphabetic Style Keyboard.*

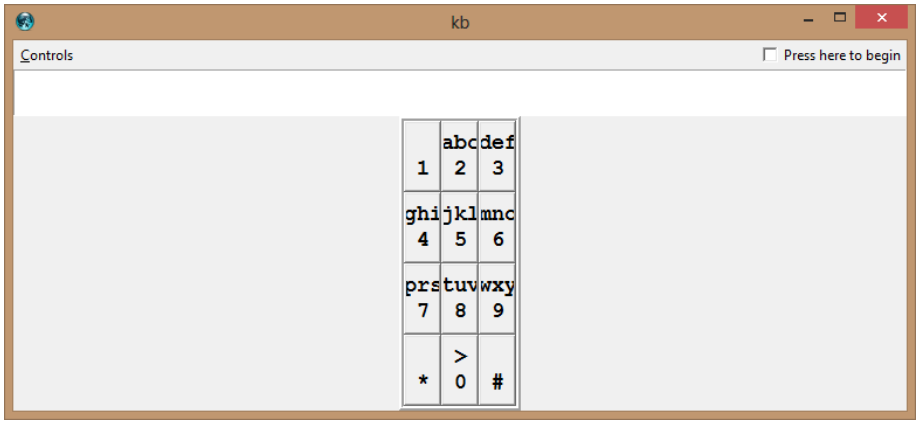

Fig 2. *PhonePad Style Keyboard.*

Forms were prepared for consent as well as for a pre-questionnaire, a mid-experiment questionnaire after each keyboard, and a post-questionnaire. Pencils or pens were also required to fill out forms.

2.4 Methods

Empirical methods were used to analyze data, while we also used qualitative methods, such as questionnaires to provide us with a different perspective. We asked participants for their consent and instructions were read to them, assuring that if they felt uncomfortable that they could stop at any time. Before starting the tasks we gave participants a pre-questionnaire, in order to determine participants experience levels with digital or physical keyboards. Also, participants were split into two groups, one group started with the phone pad keyboard first, while the remaining participants started with the alphabetic keyboard.

Before beginning the tasks, we selected a keyboard in the program, each participant alternating with the starting type. We demonstrated participants how to start and stop the recording process of the kb.tcl program. We requested that participants select the checkbox once they started and completed each sentence in order to record the amount of time it took to write. The text created by the keyboards output was only in lowercase and did not allow participants to back space, in order to record any errors which could suggest issues with the keyboards that would not be noticeable otherwise. For every participant, a script file with their data was saved after each conducted experiment.

Participants were instructed to do three tasks for each keyboard, using their computer mouse and selected keyboard interfaces. Each of the tasks required participants to write a 46 character long sentence with the digital keyboard and mouse. Instructions for each keyboard were provided as well as each sentence was given one at a time, in order to not overwhelm or rush participants. Once a participant completed the three tasks or sentences, they were asked in a mid-experiment questionnaire, what they thought about the keyboard they were using. After a

participant completed one keyboard and answered the question, they were asked to repeat the same tasks again with the other keyboard and were asked the same mid-experiment question.

The three sentences were:

Sentence 1 (46 characters)
- `come back next week and we will play some more`

Sentence 2 (46 characters)
- `after school my mom took us out to go shopping`

Sentence 3 (46 letters)
- `on the way down the stairs i saw an orange cat`

Once tasks were completed, a post-questionnaire was given, and participants were asked which of the two keyboards they preferred. This concluded the experiment for participants. Participants were given numbers to preserve their anonymity and their files from the test were renamed to have their participant ID. File data from the keyboard program was transferred into an excel spreadsheet for statistical analysis using standard deviation, the t-test and frequency tables. We noted the time in which participants completed tasks, writing speed and number of errors in each sentence for analysis and took into consideration participants comments about the keyboards.

2.5 Problems

One significant problem posed to us, as we cycled between our roles as participant or observer, is that we effectively watched each other do the test ahead of time; with the exception of course being the participant who went first. This allowed the remaining two participants to learn from the mistakes of this first, and quite possibly could have lead to the data being contaminated.

A recommendation for future testing would be to remove the keyboard away from the participants. Participants are used to using the keyboard as a medium for typing, but the experiment calls for exclusive typing via the mouse. This could potentially cause unnecessary errors and skewed data.

3. Results

Keypad Usage Speed (characters per minute)

Keypad Type	Phone Pad	Alpha Pad
Mean	52.404872	58.694359
Median	41.3	52
Standard Deviation	26.133203	20.17705

T-Test value for Speed of the two keyboards: 0.2381197

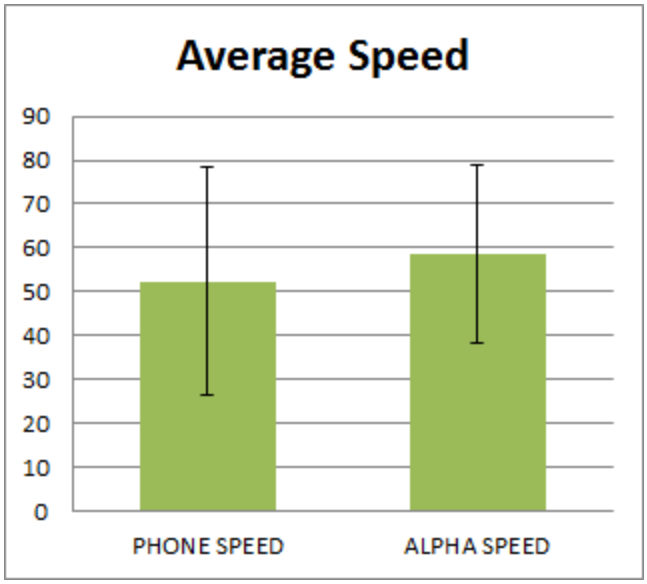

Fig 3. *Average characters per minute for the virtual keyboards.*

Keypad error rate

Keypad Type	Phone Pad	Alpha Pad
Mean	3.6969697	0.7269697
Median	2.5	0.5
Standard Deviation	4.7784264	1.1372486

T-Test value for Errors of the two keyboards: 0.0013675

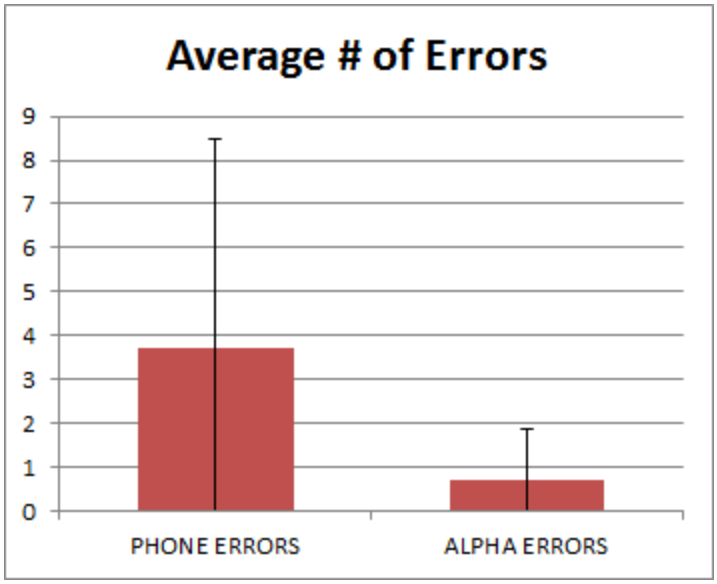

Fig 4. *Average errors per sentence (46 characters) for the virtual keyboards.*

Frequency of Results

Phonepad and Alphabetic Keyboard Speed

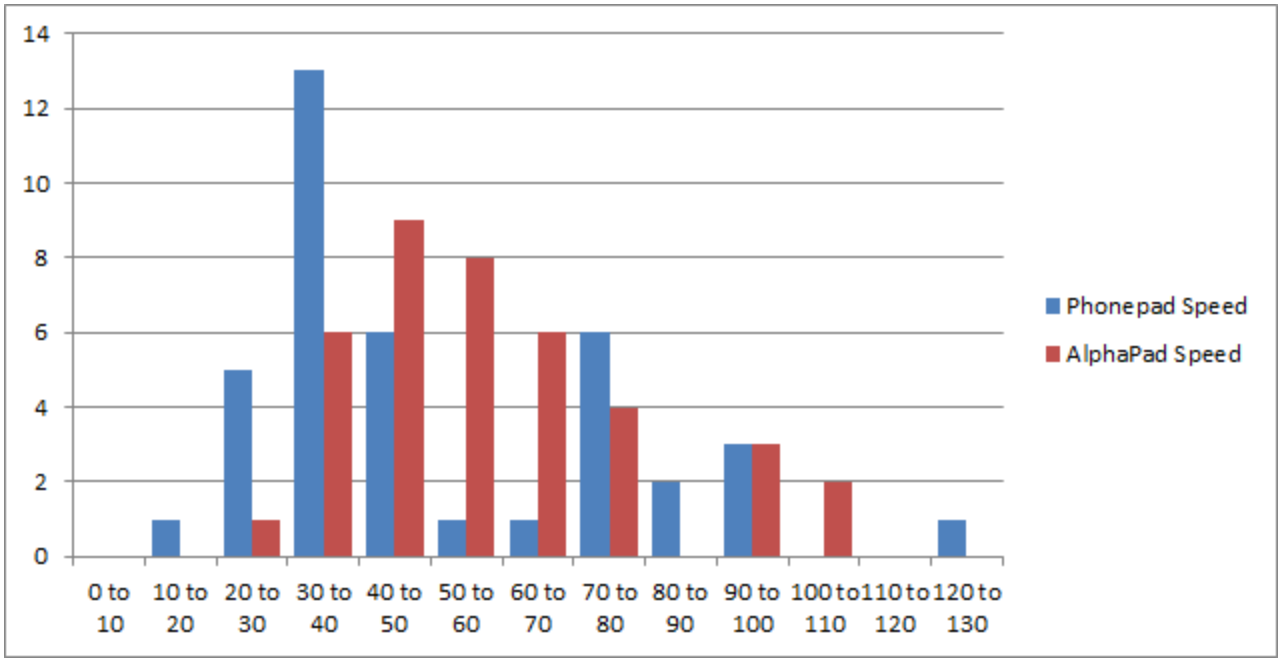

Fig 5. *Frequency graph for keyboard speed results (characters per minute).*

Phonepad and Alphabetic Keyboard Error Frequency

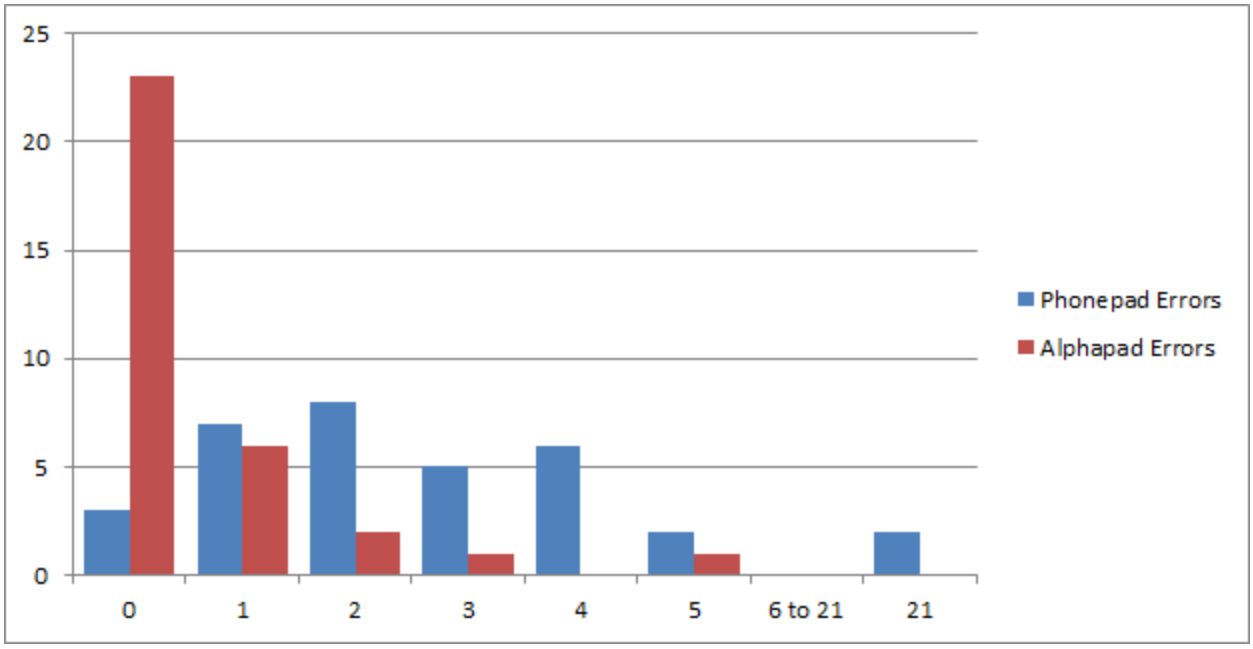

Fig 6. *Frequency graph for keyboard error count (rounded down).*

Data Anomalies

- Data from users on teams D102-T4 and D101-T2 consisted of the exact same data for the three person teams including the outlying 21.5 error rate, which was repeated across a user in each group.

4. Discussion

4.1 Interpretation of results

These results indicate that when it comes to the amount of time taken by participants to complete the tasks, there is no statistically significant difference between the two virtual keyboard layouts. Contrary to initial prediction, the time spent to complete each sentence did not vary enough to be considered a significant difference.

The errors however, were significantly different between the two. The alphabetic virtual keyboard averaged far fewer errors than the virtual phonepad. This could possibly indicate that the layout of the alphabetical keyboard afforded the user higher error avoidance, whereas the phone pad interface did not. This was suggested to be confirmed after running a t-test on the datasets, showing a statistically highly significant difference between the error rate on the alphabetical pad (lower error rate) and the phonepad style (higher error rate).

When looking at these results as a cohesive whole, the study suggests that while our hypothesis indicated that there would be a difference in speed between the keyboards, this has proven to be

false (as the t-test confirms the null hypothesis that there is no statistically significant difference). However, our hypothesis that the error rate would vary across layouts is likely true, shown by the highly significant difference in recorded results. This suggests that the alphabetic layout is the better performing of the two types, being ahead in one category and tied in the other.

4.2 Impact for practitioners

Our study found that the alphabetical virtual keyboard performed significantly better at reducing errors in comparison to the phonepad virtual keyboard, without a corresponding impact on average usage speed, and would thus suggest utilizing the alphabetical variant if a virtual keyboard is required. While it is purely speculation that the inclusion of a functional delete/backspace key would affect the speeds of use of the virtual keyboards, we feel that it is likely that future designs would find any potential speed reduction worth the ability to lower error rates to near zero.

While for our sample group, on average the tested virtual keyboard layouts performed quite similarly in speed, looking at individual results shows that many participants performed faster on one keyboard over the other. This suggests that implementations should consider incorporating both variants and allowing a user to select a preference that best fits their personal usage patterns. Also, these results suggest that there may need to be a revision in the design of the phone pad keyboard and further tested.

4.3 Critical reflection

The experiment itself had some unintentional biases that, in retrospect, should have been identified and dealt with. In particular, we probably should have removed the keyboard from the testing area. Even though the test is restricted to mouse use only, we had a few instances in which participants attempted to use the keyboard first before the mouse, possibly increasing the amount of time it took to complete the task.

Another issue was the selection of the participants, which were the students doing the report itself. Doing the experiment with the observers themselves creates a whole host of issues. A few of these being any biases that may arise as a result of knowing specifically what we are being tested for, and how our behaviour will affect our results, as well as being prepared for what was expected of us and the ability to learn from other participants' mistakes. This could cause a possible unintentional reduction in error.

4.4 Research agenda

Contrary to our initial hypothesis, the speed that the participants completed each of their tasks were, on average, very similar. However, the amount of errors that occurred between the two mediums were significantly different. Although we hypothesized that the alphabetical keyboard would result in fewer errors, we did not predict the similarity in time taken to complete each sentence.

If this study was to go on further, it would be interesting to see how a virtual keyboard that has the ability to allow participants to correct their mistakes compared to a keyboard that does not, such as that in our experiment. It may also be worth investigating whether alternative layouts such as QWERTY or DVORAK show any marked improvement (or difficulties), or whether the same patterns are shown on a touchscreen vs a mouse based system..

5. Conclusions

At the beginning of this experiment, we predicted in our hypothesis that the alphabetical virtual keyboard would out perform the virtual phone pad in terms of allowing participants to complete their tasks sooner, and by having a lower error frequency. To explore this hypothesis, we created an experiment during which participants would complete a set of three sentences using each virtual typing device whilst being recorded by a separate program. The experiment was prefaced by a pre-test questionnaire and followed by a post-test questionnaire, in order to collect demographic and experience information. During the test, we had each participant instructed by two observers, to monitor and instruct participants.

Our results indicated that the time taken for participants to complete their respective sentences are on average very similar from each other. However, the vast difference in errors that occurred was expected, with the alphabetical virtual keyboard reporting the fewest errors. This new information contradicted our null hypothesis in regards to the time taken to complete the tasks, however this was not the case in regards to error occurrence. We concluded from these findings that the layout of the alphabetical keyboard afforded the user higher error avoidance, whereas the phone pad interface did not, as evidenced by the large error disparity. T-tests on the datasets added further confirmation to this conclusion.

In reflection we evaluated our testing procedures and made recommendations to researchers seeking to replicate our experiments some improvements, such as expanding the participant pool and minimizing the testing environment. Finally, we discussed where this research could possibly go in the future, and what we would be most interested in pursuing should we desire. We would like to also thank Carman Neustaedter and Azadeh Forghani for their support in collecting data and teaching us the methods for conducting a controlled experiment.

References

Greenberg , S. (n.d.). *How to structure reports on controlled experiments in design evaluation*. Retrieved from http://carmster.com/432/ControlledStudy/reports.html

Appendix: Archival report

A. Raw Data

Please contact Azadeh Forghani for original data.

B. Statistical runs from computer

Include printouts of your statistical analyses.

C. Experimental consent forms

Please contact Azadeh Forghani for original consent forms.