Button Bonanza: A Comparative Study of Swipe and Touch Characteristics in Mobile Run-and-Dodge Games

Jorian Berkhout
j.l.berkhout@uu.nl

Kwinten Jacobs
k.h.s.jacobs@students.uu.nl

Alissa Vermast
a.vermast@students.uu.nl

Ana Patricia Márquez Sánchez
a.p.marquezsanchez@students.uu.nl

Evelien van Workum
e.m.m.vanworkum@students.uu.nl

Naomi Vogelpoel
n.j.f.vogelpoel@students.uu.nl

Abstract
A comparative study was conducted to determine the difference in performance, intuitiveness and enjoyment of using either virtual buttons or swiping controls when playing a run-and-dodge smartphone game. The controls and characteristics studied were chosen based on previous research. An app was developed in which the participants self reported their experienced enjoyment and intuitiveness after playing the game. User performance was measured automatically during the game. The findings are in line with prior studies, and suggest swiping is better than tapping for all metrics. Further research is warranted to see the effect of the learning curve and phone position on these findings.

Author Keywords
Mobile game, game controls, swiping, virtual buttons, mobile interaction, user study, endless runner, run-and-dodge game

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *Haptic devices;* User studies; *Gestural input;*

Button Bonanza

For this study, the game 'Button Bonanza' was developed. A player uses either buttons or swiping gestures to guide a character along three parallel lanes. In these lanes three different types of obstacles are present: thimbles to jump over, hanging buttons to duck under, and large stacks of buttons that should be dodged entirely by moving left or right. How one performs these actions differs per level.

The game is currently available to download through the Google Play Store:

https://play.google.com/store/apps/details?id=com.DefaultCompany.ButtonBonanza

Introduction
When designing a mobile game, developers have to consider different decisions, such as the interaction method. According to design heuristics by Korhonen and Koivisto [15], it is important for usability that the controls are convenient and flexible, consistent, and similar to established controls. The aim of this study is to determine the better interaction style for mobile run-and-dodge games out of two interaction styles: virtual buttons and swipe gestures.

To determine the optimal controls for this specific type of game, three metrics were evaluated: intuitiveness, performance and enjoyment. These metrics were chosen because effectiveness and entertainment are found to be good metrics for evaluating a game [25] and intuitiveness of the controls affects player enjoyment as well [27].

By answering the following research question, mobile game developers can make more informed decisions when designing run-and-dodge games:

How do swipe controls compare to virtual buttons in terms of intuitiveness, performance and enjoyment in a mobile run-and-dodge game?

Related work
A review by Chaichitwanidchakol and Feungchan identified the existing types of mobile game interaction [4]. However, the number of studies comparing button and swipe interactions on mobile devices in the context of games is limited. A relevant empirical study has been done by Browne and

Figure 1: Screenshot of Button Bonanza with partly transparent buttons as input method

Anand [3], who compared button, swipe, and accelerometer interactions. No significant differences between buttons and swipe gestures were found in their 'scroll shooter' game. In the current study, a similar study design is proposed, investigating these controls in a different genre.

Studies by Negulescu et al. [21] and Kujala [16] have compared tapping, swiping and other interaction techniques while performing another task (i.e., walking, driving), and found that although performance did not differ significantly, participants spent less time looking at the screen for swiping and motion gestures. Additionally, Kujala stated that more pointing accuracy was needed for buttons. It also has been well established in literature that button size [23, 5] and placement [10, 29] can affect performance and usability depending on context. A study on website interfaces on smartphones by Dou and Sundar found that horizontal swiping positively affects desire to continue to use a website, and that swiping controls could induce feelings of enjoyment in users [7]: these effects may apply to mobile games as well.

Intuitiveness is mainly based on users having familiar and past experiences, and can therefore differ greatly per person [18]. Additionally, intuitiveness of interactions is often fast and unconscious, causing it to be difficult to be described or self-reported by participants [2]. Instead, studies often rely on general performance, familiarity and workload measures [1, 19]. Intuitiveness can thus be characterized by speed, high levels of efficiency and accuracy, and a lower level of conscious awareness of the cognitive processing taking place [18].

Hypotheses

In this study, two of the most common interaction styles are considered: swiping and virtual buttons [4]. In order to determine the best of these styles, player performance is measured, as better controls lead to better performance. However, performance alone does not suffice, because player enjoyment is vital for the success of a game as well. Even though increased performance leads to increased enjoyment [28], a study by Klimmt et al. found that more difficult controls do not necessarily reduce player enjoyment [14]. This can be attributed to the fact that less precise controls can increase the level of challenge, which can affect player enjoyment both ways [24, 22]. Consequently, player performance alone is insufficient to determine the best interaction style and enjoyment should be evaluated as well.

According to Chaichitwanidchakol et al. [4], there are two types of mobile game interactions: natural and non-natural. Natural interactions reflect their real-world counterpart, whereas non-natural interactions commonly involve (virtual) buttons. Although neither of the interaction styles evaluated in this study resemble their physical counterparts, swiping can be considered more intuitive as it mimics the movement performed by the player character. Intuitive controls are important for both usability and player enjoyment according to the heuristics defined by Desurvire et al. [6]. This is represented in the first set of hypotheses:

H1a: Swiping is more intuitive than interactions using virtual buttons.
H1b: Swiping will be perceived as more enjoyable than virtual buttons.

Tapping is generally a shorter action compared to swiping. Tapping virtual buttons can be registered at the moment a finger is detected by the touchscreen, while a swipe is typically only registered after the finger is lifted from the screen. In our developed game "Button Bonanza", the swipe is already registered halfway through the swipe motion. Swipe gestures can be faster for complex interactions such as typ-

Gameplay

A video that demonstrates the gameplay of Button Bonanza can be seen here:

https://youtu.be/wGtkV1fNyRs

Controls

Virtual buttons
The buttons used are partly transparent (see Figure 1) to ensure the positioning and size of the buttons are visible to the user, but do not obscure the view.
Swiping
The swiping gestures can be done anywhere on the screen, and follow conventional distances scaled to real world distance.

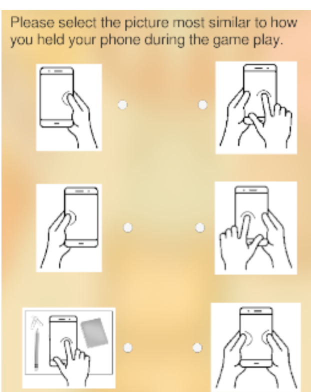

Figure 2: Question how people held phone during game play

ing (e.g. SwiftKey [20]) because gestures can be connected subsequently. In Button Bonanza, the interactions are short and separate, diminishing this advantage. As quick reactions are vital in dodging games, the second set of hypotheses state:

H2a: Tapping virtual buttons leads to increased performance.
H2b: Tapping virtual buttons can be more easily timed than swiping.

Methodology

A within-subject design was chosen for this study, since it accounts for individual differences between participants. As mentioned before, the experiment consisted of having participants playing the Button Bonanza game, with two different types of controls: swiping and tapping virtual buttons. The order in which the participants played with these controls was randomised.

Participants
The experiment was conducted with 25 participants who completed the entire study (13 male, 11 female, 1 other). The ages of the participants ranged from 19 to 60 (M = 26.44, SD = 9.20). One of the participants reported having a physical condition that interfered with their motor skills, but their results did not seem to be outliers. The participants reported spending an average of 2 hours per week playing mobile games. All participants were gathered by the researchers by means of convenience sampling, and all gave their informed consent prior to participating in the experiment. The participants were not compensated for their time, as participation was completely voluntarily.

Materials
The participants were instructed to download the Button Bonanza app from the Google Play Store [13], using their own mobile phone with Android. The Unity 3D engine [26] was used to develop the game. The data collected with the app was automatically sent out to and saved in a Firestore database [12] only accessible to the researchers.

All questionnaires were added inside the app for the participants' convenience. Performance of the game is measured automatically. Enjoyment was measured using a short, in-game version of the well-established GEQ (Game Experience Questionnaire), the iGEQ [11, 17]. As the Button Bonanza game did not include a story line, the component dealing with story (*Sensory and Imaginative Immersion*) were disregarded. The NASA R-TLX questionnaire [9] (which is used more often in mobile contexts [8]) measures perceived cognitive workload and was used in this study to get an indication of the perceived intuitiveness (see also Related work).

Experiment procedure
After downloading and opening the app (see Figure 1), the participant was guided through the experiment. Participants were allowed to hold their phone as they wished during the experiment, as long as it was in portrait mode (to minimize any confounding variables). They were asked to perform the experiment alone, without being disturbed.

Upon opening the application, the experiment was explained and participants were asked for their consent. Next up were questions about demographics (i.e., gender, age), physical conditions possibly interfering with fine motor skills and mobile gaming experience, which were all expected to be confounding variables [3, 1, 21]. The actual experiment started with a practice session of 20 seconds, where the participants were randomly assigned swiping or virtual button controls. Next, the participants played two subsequent levels of one minute each with this control set. When finished, the participant were asked to fill out the question-

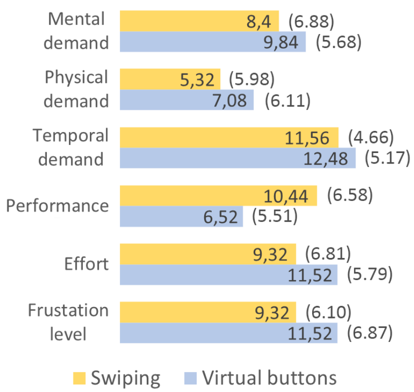

Figure 3: Descriptive statistics [$M(SD)$] of NASA R-TLX scores

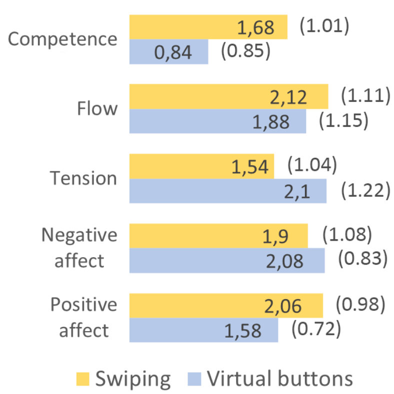

Figure 4: Descriptive statistics [$M(SD)$] of iGEQ questionnaire results

naires described in the Materials section. This process was repeated for the other control set. The experiment lasted about 15 minutes.

Data collection and storage
During the experiment, quantitative data was collected. As mentioned before, the participants started by answering questions about their demographics. After each level they played, the participants were asked i) how they held their phone during the game play (See Figure 2), and to fill in ii) in the NASA R-TLX questionnaire [9], and iii) the iGEQ questions [11]. To get a detailed picture of the performance, the number of correctly dodged obstacles ('achieved score') was measured, as well as the number of incorrect, missed and poorly timed swipes/taps. The collected data was anonymized by assigning every participant an ID; this way no participant can be recognized when looking at the data. The data was kept at the latest until the end of September 2021.

Data analysis
The data was transformed from a JSON to a CSV file and all incomplete entries (20 in total) were removed. For all dependent variables (i.e., intuitiveness, performance and enjoyment), two-tailed paired-samples t-tests or Wilcoxon-signed rank tests were used to compare the results between the independent variables (i.e., swiping versus tapping). The assumptions for these tests were checked by comparing means and variances, plotting histograms and conducting Shapiro-Wilk tests to check the normality. Everywhere an alpha of 0.05 was applied.

Findings

Intuitiveness
Regarding the intuitiveness (**H1a**), a paired-samples t-test was conducted to compare the R-TLX scores for the swiping and tapping conditions. Swiping (M=9.08, SD=4.64) was found to have a significantly lower subjective workload than tapping (M=11.31, SD=3.76); t(24)= - 2.82, p = .009. Therefore, hypothesis **H1a** was accepted.

For its individual components (see Figure 3), paired-samples t-tests were conducted as they were all normally distributed, too. Significant differences were only found for the *Performance* and *Frustration level* component. The perceived *Performance* was significantly higher for swiping than tapping; t(24) = 3.10, p = .005. Furthermore, significantly more *Frustration* was perceived for tapping; t(24) = -2.22, p = .036. Note that for each metric with its components, the test statistics and p-values are shown in the appendix.

Enjoyment
Regarding the enjoyment (**H1b**), a paired t-test was conducted for the total (i.e., summed) iGEQ, and shown to be significantly higher for swiping (M=13.26, SD=3.65) than tapping (M=10.62, SD=3.34); t(24) = 3.81 , p = .009. Therefore, **H1b** was accepted.

Concerning its individual components (see Figure 4), a t-test also reported a significant difference in the respective *Competence* scores; t(24) = 3.87, p = .001. Additionally, a significant difference in *Tension* was reported for the two conditions; t(24)= -2.75, p = .011. A Wilcoxon Signed-Rank test indicated that the *Positive affection* ranks of swiping were significantly higher than those of the buttons; z = 2.353, p = .017.

Performance
To investigate hypothesis **H2a**, various statistical tests were performed for each metric (see also Table 1). A paired-samples t-test was conducted on the achieved scores of each level as well as summed up. Swiping had a significantly higher summed score than tapping for each of them, except Level 1. For this level, the swiping score was higher

	Swiping	Virtual Buttons
Achieved score	24.32 (11.47)	19.20 (9.17)
Incorrect inputs	6.00 (9.23)	7.88 (7.45)
Poorly timed inputs	23.44 (8.38)	25.64 (7.60)
Missed inputs	2.24 (5.16)	3.28 (7.80)

Table 1: Descriptive statistics [$M(SD)$] of summed performance scores

but not significant; $t(24) = 1.90$, $p = .070$.

Wilcoxon signed-rank tests had to be used for the number of incorrect inputs because of non-normally distributed data. Tapping led to significantly more incorrect inputs than swiping for Level 1 as well as for the total number of incorrect inputs; $z(24) = 50.5$, $p = .040$ and $z(24) = 73.5$, $p = .048$ respectively. No other significant results were found.

For hypothesis **H2b**, t-tests were also conducted for the number of poorly timed inputs. For the number of missed swipes, Wilcoxon tests had to be conducted. The only significant result found was that for Level 2 tapping appeared to have a higher number of poorly timed inputs than swiping; $t(24) = -3.10$, $p = .005$. Hypotheses **H2a** and **H2b** were therefore both rejected.

Discussion

General discussion of the results

As stated in the Findings section, hypotheses **H1a** and **H1b** were accepted, and **H2a** and **H2b** were rejected. These findings are mostly in line with other studies. In contrast to Browne and Anand [3], this study presents significant differences between tapping and swiping for both enjoyment and performance. Similar to the study by Dou and Sundar [7], swiping was shown to positively affect enjoyment. Performance was also found to be higher for swiping compared to tapping, possibly due to the attention needed for tapping (cf., [21, 16]). The findings of this study contribute to knowledge about interaction methods on mobile devices, which may help game designers make more informed choices.

Limitations and future work

The current study has a couple of limitations. Firstly, this study is limited in what kind of interaction method it compares. This presents the possibility to compare other interaction styles such as tilt or area tapping in future work.

Secondly, this study does not take the learning curve into account when measuring the performance. The duration of this study did not allow for tests spanning out over multiple sessions in order to grasp the effect of the learning curve of the various interaction methods. Furthermore, the order in which the players encountered the conditions was not saved. Although the order was counterbalanced and should therefore not have an effect, future studies should verify the assignment of conditions. Lastly, participants held their phones in different ways during the experiment. How participants held their phone might be influenced by different factors. The participants were allowed to hold their phone in whatever way they liked, as they would be the most comfortable in this scenario and it would mimic the real world usage the best. The results of this research might have been different if all participants would have been forced to use the phone in the same manner.

Conclusion

To summarize, the results of the present study indicate that a) swiping is more intuitive than using virtual buttons; b) swiping is perceived as more enjoyable than virtual buttons; c) swiping may be more easily timed than swiping; and d) tapping virtual buttons does not lead to increased performance. These findings lead to the conclusion that swiping is a better control scheme for mobile run-and-dodge games than virtual buttons when looking at performance, and perceived intuitiveness and enjoyment.

Acknowledgements

We would like to express our gratitude to Dr. W.O. (Wolfgang) Hürst for guidance of this research. We would also like to thanks the peer-reviewers who provided helpful comments on a previous version of this paper.

REFERENCES

[1] Alethea Blackler and Jörn Hurtienne. 2007. Towards a unified view of intuitive interaction: Definitions, models and tools across the world. *MMI-Interaktiv* 13 (08 2007), 36–54.

[2] Alethea Blackler, Doug Mahar, and Vesna Popovic. 2010. Older Adults, Interface Experience and Cognitive Decline. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10) (OZCHI '10)*. Association for Computing Machinery, New York, NY, USA, 172–175. DOI: http://dx.doi.org/10.1145/1952222.1952257

[3] Kevin Browne and Christopher Anand. 2012. An empirical evaluation of user interfaces for a mobile video game. *Entertainment Computing* 3, 1 (2012), 1–10. DOI: http://dx.doi.org/10.1016/j.entcom.2011.06.001

[4] Pitsanu Chaichitwanidchakol and Witcha Feungchan. 2018. Exploring mobile game interactions. *International Journal of Electrical and Computer Engineering* 8, 5 (2018), 3954–3965. DOI:http://dx.doi.org/10.11591/ijece.v8i5.pp3954-3965

[5] Jessica Conradi, Olivia Busch, and Thomas Alexander. 2015. Optimal touch button size for the use of mobile devices while walking. *Procedia Manufacturing* 3 (2015), 387–394. DOI: http://dx.doi.org/10.1016/j.promfg.2015.07.182

[6] Heather Desurvire and Charlotte Wiberg. 2009. Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration. In *Online Communities and Social Computing*, A. Ant Ozok and Panayiotis Zaphiris (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 557–566. DOI: http://dx.doi.org/10.1007/978-3-642-02774-1_60

[7] Xue Dou and S.Shyam Sundar. 2016. Power of the Swipe: Why Mobile Websites Should Add Horizontal Swiping to Tapping, Clicking, and Scrolling Interaction Techniques. *International Journal of Human-Computer Interaction* 32, 4 (2016), 352–362. DOI: http://dx.doi.org/10.1080/10447318.2016.1147902

[8] Mattias Georgsson. 2020. NASA RTLX as a Novel Assessment Tool for Determining Cognitive Load and User Acceptance of Expert and User-based Usability Evaluation Methods. *European Journal for Biomedical Informatics* 16, 2 (2020), 14–21. DOI: http://dx.doi.org/10.24105/ejbi.2020.16.2.14

[9] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. Elsevier Science Publishers B.V., North-Holland, Netherlands, 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[10] Tymoteusz Horbiński, Paweł Cybulski, and Beata Medyńska-Gulij. 2020. Graphic design and button placement for mobile map applications. *The Cartographic Journal* 57, 3 (2020), 196–208. DOI: http://dx.doi.org/10.1080/00087041.2019.1631008

[11] Wijnand A. IJsselsteijn, Yvonne A.W. de Kort, and Karolien Poels. 2013. *The Game Experience Questionnaire*. Technische Universiteit Eindhoven, Eindhoven, North-Brabant, Netherlands.

[12] Google Inc. 2021a. Firebase. (2021). https://firebase.google.com/

[13] Google Inc. 2021b. Google Play Store. (2021). https://play.google.com/store

[14] Christoph Klimmt, Tilo Hartmann, and Andreas Frey. 2007. Effectance and control as determinants of video game enjoyment. *Cyberpsychology & behavior* 10, 6 (2007), 845–848. DOI: http://dx.doi.org/10.1089/cpb.2007.9942

[15] Hannu Korhonen and Elina M. I. Koivisto. 2006. Playability Heuristics for Mobile Games. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 9–16. DOI: http://dx.doi.org/10.1145/1152215.1152218

[16] Tuomo Kujala. 2013. Browsing the information highway while driving: Three in-vehicle touch screen scrolling methods and driver distraction. *Personal and Ubiquitous Computing* 17, 5 (2013), 815–823. DOI: http://dx.doi.org/10.1007/s00779-012-0517-2

[17] Effie L.-C. Law, Florian Brühlmann, and Elisa D. Mekler. 2018. Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 257–270. DOI:http://dx.doi.org/10.1145/3242671.3242683

[18] Simon Lawry, Vesna Popovic, Alethea Blackler, and Helen Thompson. 2019. Age, familiarity, and intuitive use: An empirical investigation. *Applied Ergonomics* 74 (2019), 74–84. DOI: http://dx.doi.org/10.1016/j.apergo.2018.08.016

[19] Mitchell McEwan, Alethea Blackler, Peta Wyeth, and Daniel Johnson. 2020. Intuitive Interaction with Motion Controls in a Tennis Video Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 321—333. DOI:http://dx.doi.org/10.1145/3410404.3414242

[20] Microsoft. 2021. Download Microsoft SwiftKey the smart keyboard and get more done. (2021). https://www.microsoft.com/en-us/swiftkey?activetab=pivot_1%3Aprimaryr2

[21] Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank. 2012. Tap, Swipe, or Move: Attentional Demands for Distracted Smartphone Input. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. Association for Computing Machinery, New York, NY, USA, 173–180. DOI:http://dx.doi.org/10.1145/2254556.2254589

[22] Serge Petralito, Florian Brühlmann, Glena Iten, Elisa D. Mekler, and Klaus Opwis. 2017. A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5087–5097. DOI: http://dx.doi.org/10.1145/3025453.3026047

[23] Kristian Adi Nugraha Restyandito. 2017. The Effectiveness of Button Size on Mobile Device Based on Hand Dimension. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2017)*, Vol. 2. IAENG, Hong Kong, China, 916–920.

[24] Mike Schmierbach, Mun-Young Chung, Mu Wu, and Keunyeong Kim. 2014. No One Likes to Lose. *Journal of Media Psychology* 26, 3 (2014), 105–110. DOI: http://dx.doi.org/10.1027/1864-1105/a000120

[25] Jhonny Paul Taborda, Jeferson Arango-López, Cesar A. Collazos, Francisco Luis Gutiérrez Vela, Fernando Moreira, Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, and Sandra Costanzo. 2019. Effectiveness and Fun Metrics in a Pervasive Game Experience: A Systematic Literature Review. In *New Knowledge in Information Systems and Technologies*. Springer International Publishing, Cham, 184–194. DOI: http://dx.doi.org/10.1007/978-3-030-16187-3_18

[26] Unity Technologies. 2021. Unity. (2021). https://unity.com/

[27] Andrew Thorpe, M. Ma, and A. Oikonomou. 2011. History and alternative game input methods. In *2011 16th International Conference on Computer Games (CGAMES)*. IEEE, Louisville, KY, USA, 76–93. DOI: http://dx.doi.org/10.1109/CGAMES.2011.6000321

[28] Sabine Trepte and Leonard Reinecke. 2011. The Pleasures of Success: Game-Related Efficacy Experiences as a Mediator Between Player Performance and Game Enjoyment. *Cyberpsychology, behavior and social networking* 14 (02 2011), 555–7. DOI:http://dx.doi.org/10.1089/cyber.2010.0358

[29] Jinghong Xiong and Satoshi Muraki. 2016. Effects of age, thumb length and screen size on thumb movement coverage on smartphone touchscreens. *International Journal of Industrial Ergonomics* 53 (2016), 140–148. DOI: http://dx.doi.org/10.1016/j.ergon.2015.11.004

Appendix: Test statistics and p-values for each metric and component

For each metric (user performance, intuitiveness/subjective workload, and experienced enjoyment), a two-sided, paired-samples t-test is used if the difference of swiping-tapping pairs follow a normal distribution; otherwise, a Wilcoxon signed-rank test is used. They can be recognized by their test statistic: the first have test statistics starting with "t(24)=", the latter "z=".

	Test statistic	**P-value**
Achieved score	$t(24) = 2.80$	$p = .010$
Incorrect inputs	$z = 48.5$	$p = .177$
Poorly timed inputs	$t(24) = 0.76$	$p = .453$
Missed inputs	$z = 27.0$	$p = .190$

Table 2: Advanced statistics of performance scores for the tutorial level

	Test statistic	**P-value**
Achieved score	$t(24) = 1.90$	$p = .070$
Incorrect inputs	$z = 50.5$	$p = .040$
Poorly timed inputs	$t(24) = 0.52$	$p = .608$
Missed inputs	$z = 3.0$	$p = .223$

Table 3: Advanced statistics of performance scores for Level 1

	Test statistic	**P-value**
Achieved score	$t(24) = 2.84$	$p = .009$
Incorrect inputs	$z = 46.5$	$p = .440$
Poorly timed inputs	$t(24) = -3.10$	$p = .005$
Missed inputs	$z = 9.0$	$p = .395$

Table 4: Advanced statistics of performance scores for Level 2

	Test statistic	**P-value**
Mental demand	$t(24) = -1.78$	$p = .088$
Physical demand	$t(24) = -1.67$	$p = .108$
Temporal demand	$t(24) = -0.79$	$p = .436$
Performance	$t(24) = 3.10$	$p = .005$
Effort	$t(24) = -1.95$	$p = .064$
Frustration level	$t(24) = -2.22$	$p = .036$
Total R-TLX score	$t(24) = -2.82$	$p = .009$

Table 5: Advanced statistics of R-TLX scores

	Test statistic	**P-value**
Competence	$t(24) = 3.86$	$p = .001$
Flow	$z = 26.0$	$p = .089$
Tension	$t(24) = -2.75$	$p = .011$
Challenge	$t(24) = -1.30$	$p = .205$
Negative affect	$t(24) = -1.84$	$p = .079$
Positive affect	$z = 22.5$	$p = .017$
Total iGEQ score	$t(24) = 3.81$	$p = .001$

Table 6: Advanced statistics of iGEQ scores