Evaluating User Interface Design Based On *Don Norman's 7 Principles*: An Investigation Into The Legitimacy Of Principled User Interface Design

Adrian Fearman, Joseph Yong, Jeff Wang, Talayeh Amiri Tokaldany, Alice Chai
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Ontario, Canada M3J 1P3
afearman@my.yorku.ca, josephy5@my.yorku.ca, Jeffslw@my.yorku.ca, tala9776@my.yorku.ca, chaia@my.yorku.ca

ABSTRACT
The design of modern user interfaces follows strict, principled guidelines. Although most of these principles are open to being interpreted by the user interface's designer, the repeated implementation across platforms has set a specific standard regarding how user interfaces are expected to look and respond. The principles have become embedded because users have come to expect a certain response to interaction, not necessarily because it is inherently better or naturally intuitive. The familiarity of these designs thus impacts a user's approach to interaction, as well as their satisfaction. Questions arise about what happens when these design principles are not implemented and how that impacts users. In this paper, we will determine whether an unprincipled user interface design can outperform a principled user interface design. Our findings show that user interface design that does not follow standard design principles performs much worse than principled design with lower user satisfaction and performance.

Keywords
User interfaces, accessibility, usability, design principles, UI, modern-day, conventional, unconventional, principled, unprincipled

INTRODUCTION
Before the inception of the modern, user-focused internet, as it is known today, decisions to streamline interface design were not prioritized. There was no standard practice regarding how the user interface should be designated or what it should look like [4]. Consequently, the early Graphical User Interfaces were full of crazy new ideas, such as Floating Flashing navigation and moving text [4]. However, with time, the importance of user interface design became more apparent, as it is evident in the early interface design decisions used in Apple products. Apple's success would initiate the revolution to modernize the design and structure of user interfaces. Formal consolidation of these design practices, as seen in *Don Norman's 7 Principles* [6], would change how UI designers, companies, and developers design and structure UI for their products, thus creating the present-day standard. A user interface's usability and aesthetics can play a role in the user's confidence in performing their task [2], which could potentially reduce the amount of time and frustration it would take the user to get the results they wanted. It is also of interest to e-commerce brands, as UI design has measurable impacts on online consumer behavior [1]. Because of this, there is a convincing argument that a well-designed or principled user interface would perform better–or at least be more efficient–than those that are not. But then, a question strikes: does the end-user equally benefit from this type of design? Or is it the designers and those few corporations that would benefit from a design like this?

It introduces a factor that users do not stop to think about; it is a pure coincidence that there have not been any new design philosophies that could rival the modern-day design standard, brought upon by the likes of figures such as Apple and *Don Norman's 7 Principles* [6], and the amount of control these design principles, standards, and guidelines have over how a user interface should look is worrisome. A quick browse of Apple's App Store is an example of how widespread this standardization has become. What happened to creativity or innovation? Has the general public become too complacent with what they have? When the Command-Line Interface (CLI) was the standard of user interface design, designers still criticized the model and sought to make user interfaces look and be better, and as a result, today's Graphical User Interfaces (GUIs) were created [8]. However, the same approach is not being taken toward modern-day user interface designs, and this may be the reason why designers have been stuck with the same playbook for modern user interface design for years. After all, if it works for the user, individuals, and companies like Steve Jobs or Apple, why bother to reinvent the wheel?

This is why it has become a common misconception that designers who utilize current design principles can expect better performance due to the design being more accessible and more catered toward the user. Since there are no other alternatives within the sphere of UI design, no one could object to that claim.

Therefore, this research challenges and investigates the legitimacy of one of the prominent design principles within user interface designs, *Don Norman's 7 Principles* [6]; it questions whether a user interface design that does not follow *Don Norman's 7 Principles* [6] can perform as well as the user interface designs that do follow them in terms of user accessibility, usability, and performance. However, to answer that question, this proposition had to be put to the test. To replicate such an environment, sample groups were utilized to go through a designated UI type that they were instructed to use. They were put to the test as they were placed in different UI environments. With a thorough analysis of the data collected, it can be determined whether this paper's hypotheses are supported or not.

Related Works

Throughout the years, various studies have been done on the many aspects of user interface designs. As a result, there are many sources on the topic, some of which discuss concepts similar to what this research paper aims to further explore.

A paper from 2021's CHI conference, Accessibility of Command Line Interfaces, researched how accessible Command Line Interfaces (CLIs) are to users [7]. This paper studies how accessible and usable CLIs are in the present times, where Graphic User Interfaces (GUIs) are the standard user interface type. When compared to the user interfaces of any modern-day app or software–which reflect much of the values within modern-day design principles–most CLIs (such as Unix's Vi editor) do not follow such design principles. Therefore, it can be deemed that the user interfaces within CLIs are badly designed. In addition, the authors looked into their participant's performance, tracking how much time it took each participant to finish their task while using the provided CLIs. This research aims to evaluate and track each participant's reactions to certain tasks with an outdated user interface. This paper, however, will also include an evaluation of conventional user interfaces, going further than just evaluating unconventional user interfaces.

In another related work, a journal paper called *An Interaction Model for Visualizations Beyond The Desktop*, Jansen, and Dragicevic utilize different experimental input and display UI types to understand the interactive relationship between computers and humans and evaluate its effectiveness [3]. The research concept and objects of study are similar to what we are studying, as we are both investigating unconventional means of conveying a user's input and feedback.

Another related work would be a conference paper called, *A comparison of accelerometer and touch-based input for mobile gaming*. Medryk and MacKenzie compared two mobile control methods to see which is more effective [5]. The idea and nature behind this study are similar to what we are researching. But, unlike the paper, we are not comparing different control methods. Instead, we are evaluating and investigating different types of user interface designs.

In another research paper*, The influence of user interface design on consumer perceptions: A cross-cultural comparison*, it was found that there were differences in the way people of various cultural backgrounds may perceive and react to key website elements such as color and price [1]. This research paper's conclusion implies that there is a possibility that there are no universal truths when it comes to user interface design and development, which is similar to what we argued in the introduction. This research also employs a similar experimental design, having participants interact with our user interface systems and gather both physical (derived directly from the interactive system) and emotional data (in the form of an interview).

METHOD

Participants

The research experiment included 12 voluntary participants recruited from populations within and outside of York University. We acquired the aforementioned sample size through our connections with friends, family, and/or colleagues. In an unbiased approach, we recruited participants from a variety of different age groups, which ranged from 19 to 53. All participants in our research self-identified themselves as male or female. Seven participants were male; five were female. The participants recruited had a range of different cultural backgrounds. Most of our participants listed their occupations as students. But we could not gather information about their program of study as they did not want to disclose it because of privacy reasons. Furthermore, all participants used computers and smartphones regularly, with a minimum of 3-4 hours spent on their devices daily. Participants were given no incentives or compensation for this study. Some participants had prior experience with UIs, either from developing or designing them; four participants had prior experience, while eight participants had not. We took this fact into account to prevent skews in our data.

We extensively recorded their personal information, such as their age, occupation, gender, computer usage, etc. The way we used to identify our participants is by using their name initials and providing participant numbers, which helps us to organize their data in our spreadsheet.

Apparatus

We utilized several different devices and software to study and test our participants.

For hardware devices, we used home desktops, laptop computers, and smartphones.

The desktop or laptop computer device has at least one monitor display, a display resolution of at least 1920 x 1080 pixels, a keyboard and mouse input device, and is running *Windows* or *macOS*. Participants used their own devices while participating in our study due to the remote nature of the procedure we set for our participants. Because of this, each of the specifications on the participant's computers was different from the others. We assumed some participants may also have other alternative input devices, such as a touch screen or gaming controller, that they could have used with their computer devices. To standardize this, we asked the participants to use only their keyboard and mouse as their input devices. Figure 1 shows an example of the type of computer a participant may have used.

The smartphone was the other device that our participants used. The smartphone device has at least one screen display, a display resolution of at least 1080 x 1920 pixels, a touchscreen input, and is running *iOS* or *Android*. Participants used their smartphones while participating in our study. However, unlike computer devices, we did not have to be concerned about smartphones having unique specifications or features since most smartphone features and hardware are standardized. If specific smartphones had unique features, we asked our participants to not use them in our trials. Because of this, we did not have to implement any standardization measures for the participants who used their smartphones. Figure 1 shows an example of the type of smartphone a participant may have used.

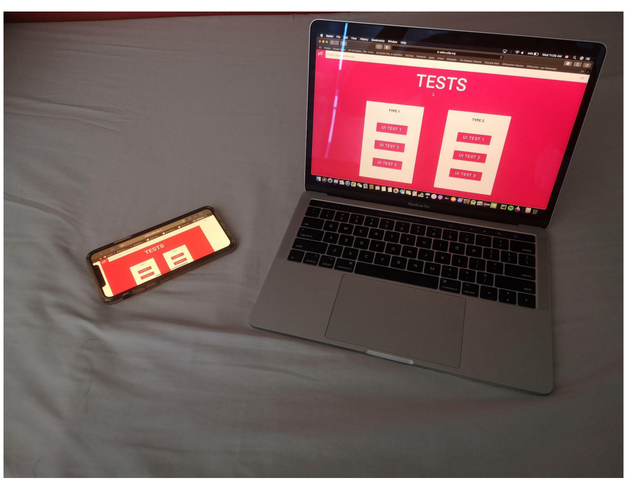

Figure 1. *iPhone 12* (released Fall 2020) and *Macbook Pro* (released 2019).

For software, participants were given an instruction PDF, containing the link to access our prototyping software. It holds our UI tests and data-recording functionalities.

The prototyping software was developed in *Processing* and *Uizard*. See Figure. 2. UI of the prototype.

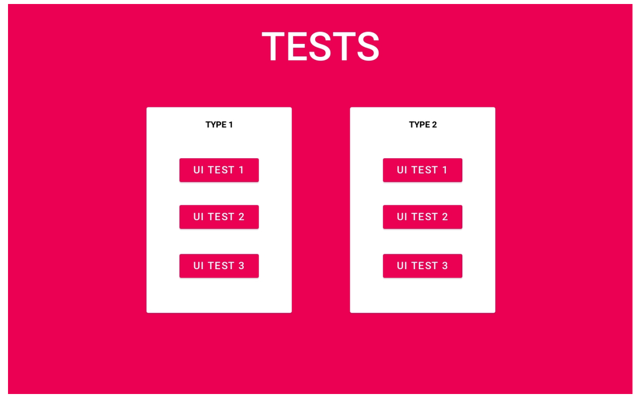

Figure 2. UI testing prototype user interface.

There were two types of user interface (UI) tests: one that followed the design principles of user interface design and one that didn't. Type 1 is our UI test that follows the user interface principle, *Don Norman's 7 Principles* [6]. This UI type represented the UI design and layout of present times, reflecting many of the UI used in today's apps, like Uber, Spotify, and so on. Type 2 is our UI test that doesn't follow the user interface principle, *Don Norman's 7 Principles* [6]. This UI type represented the UI design and layout of the bad, experimental, or outdated designs found on the internet. We could see this type of UI in apps like the original internet explorer. These types of tests were used to test and study our participants and made up the bulk of the research. Each of the trials had one task for the participant to do. It included purchasing an item on an e-commerce-like website, navigating to bookmark a blog post on a blog-like website, and sending a direct message on a communications website. In Figure. 3, you can see the comparison between UI Test 1 Trial (Type 1, Type 2).

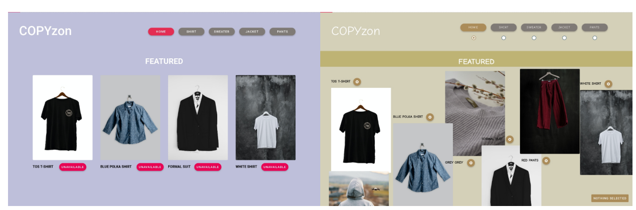

Figure 3. UI Test 1 Trial E-commerce (Type 1, Type 2).

There was also data-collecting software to collect data from the participants. There are two types of data-collecting software: one built into the test and one separate from the test. The built-in software is a data collection software that was built into the prototype, giving the UI tests. It tracked how long the user spent and how many mouse clicks were performed on each specific part of the trial. This data was automatically put into a .txt file upon task completion, which was sent back to the researchers for analysis (see Figure 4a). When the task was completed, the application prompted the user to download a .txt file with their results. These files were sent to the researcher through the questionnaire (see Figure 4b).

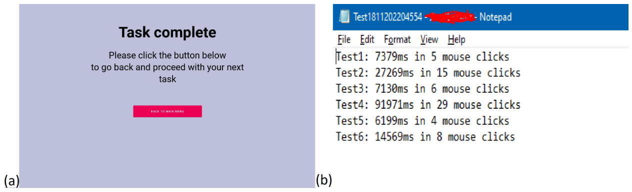

Figure 4. a) Task completion. b) Results activity.

The separate software was no different from the built-in ones, but instead of being built into the UI tests, we achieved it through a separate software. For example, to record the time spent on each trial, we used the stopwatch feature in the clock app on *iPhone* to do so.

However, in these cases, the researchers were further involved with the participants and their trials for some of the separate data-collecting software (for example, monitoring the time spent on each trial on the separate stopwatch app). We had taken this into account to not have skews in the data.

Procedure

The procedure for the participants in our research started with the formation of groups. We split the 12 participants into two groups of four and eight, respectively. Group 1 started with Type 1 first, then Type 2. Group 2 did the opposite of group 1; they started with Type 2 first, then Type 1. The participants sat in their room, acting like the enclosed booths in a typical lab environment.

Once the groups have been formed, the participants will prepare to meet with the researchers on an online call. The researcher first introduced themselves, briefly introduced the objectives of the experiment, and asked if the participant had any questions before starting. Then, the researchers provided the participant with a PDF that contained a set of instructions and links to our prototype software and questionnaire. If the participant is having any setup issues, the researcher will help to resolve them. Otherwise, the researcher is on standby to monitor their behavior.

When the participant is ready, based on which group they were assigned to, the participant starts with UI tests under their given UI type. After each trial, the participant is allowed to read the PDF again before continuing. The participant was not allowed to have any practice and had to finish their trials in one go. However, the participant is allowed to restart if problems relating to the software arise (e.g. testing software crashed during trials). When the participant finished all of their trials, they started their second (and last) attempt with the other UI type that they haven't done yet. The trials were similar to the previous test they had, but the design of the user interface will be different. After the six UI test trials were completed, participants were instructed to send a .txt file back to the researchers and were given a questionnaire to obtain further quantitative and qualitative data. The first item on the questionnaire asked the participant to enter their initials. This is a way of consent given to us for their participation in this experiment and keeps a level of anonymity for the participant. This anonymity will allow the participants to provide more honest feedback. Other items included participant personal information, such as occupation, age, computer usage, participant levels of frustration, ease of use, UI preference, and further elaboration of such items in the form of open-ended comments. Participants could have requested a one-on-one online interview as an alternative to the online questionnaire. The questions in the interview will be the same as the ones on the questionnaire. Due to the nature of a virtual face-to-face interview, it provided us with more information and insights that the questionnaire couldn't uncover. When finished, the participant can finally leave.

Design

For this research, we employed a 2 x 2 within-subject design. We tailored the design of our research as a 2 x 2 Latin square to examine the impact of the two independent variables with their two different levels on the two dependent variables.

The independent variables for our research were the different UI types that the participants faced and the different input devices that the participants used. Each of these variables would have two different levels, as stated below.

For the two levels in our different UI type variables, one of the levels was UI Type 1, while the other was UI Type 2.

For the two levels in our different input device variables, one of the levels was the mouse and keyboard input of a desktop/laptop computer, while the other was the touch screen input of a mobile device.

The dependent variable for our research was the time it takes for a participant to complete each trial and their level of frustration while doing their task within each of the UI types. While these variables are quantitative, a questionnaire was used to obtain additional quantitative data, such as frequency of device usage and level of frustration, and additional qualitative data, such as user preference and ease of use.

Within our research study, we have calculated that the total number of trials happening in our study is 72 trials (12 participants, two UI types, and three trials per UI type).

RESULTS AND DISCUSSION

After we collected all the results and data from our participants, we transferred all of that into a spreadsheet which we were able to use for graphs, tables, and analysis.

The task completion time results collected from our participants can be seen in Table 1. P1 to P4 belong to Group 1, while P5 to P12 belong to Group 2. We used the data from Table 1 to calculate the average time it took the participants to complete each of the trials within their respective groups. With this information, we were able to compile it into two visualized graphs. Figure 1 shows the average time participants took to complete each of the trials within the first group, while Figure 2 shows the average time participants took to complete each of the trials within the second group. We also recorded the participant's level of frustration within the given questionnaire, which can be seen in Table 5.

Users would often label their experiences with inaccessible and unusable user interface designs as frustrating, difficult to use, and inefficient. Thus, we can use this to determine how accessible and usable our user interface designs are.

Our data shows that participants are able to complete their tasks much faster with user interfaces in Type 1 than in Type 2, regardless of what group they were put into. Both participants in Group 1 and 2 have been able to achieve similar speeds and times. Although there are some big differences among some of the times from the participants, we were able to rule it out through these factors.

Recall how we instructed some of the participants to use their smartphones instead of their computers. Some participants who had a huge time difference amongst their peers were users who were using mobile devices. Within the questionnaire, we were able to stem this from the mobile version of the testing software. The tests in the testing software were not properly optimized for mobile users. The testing software didn't have a responsive UI which threw many participants off and the click recognition was sometimes spotty.

> P3: "id like to speak with the person who designed how it would work on mobile. it sucks. it doesn't work. spent a minute trying to hit a tiny button. desktop portion works fine..."

Taking this into account, this difference wouldn't be because of the participants but rather due to some minor unavoidable problems that the participants had to face. Hence, if we were able to provide them with a better mobile experience, then the participants should have better or similar times to those on desktop/laptop computers.

Another point that was also pointed out was the fact that participants forgot their task midst their trial. This prompted them to look back to read the instructions again, which would eat up more time in the process. Because of that, their time was longer as they had to spend time remembering their task. We did implement some precautions within the testing software, but it does not fully prevent situations like this. Hence, any time differences caused by this are inevitable, as we cannot fix one's tendency to forget. However, thanks to our precautions, the number and impact of cases like this are very minimal. Thus, we were able to proceed with analyzing the data without scrutiny.

Overall, Group 1 averagely takes 14.21 secs to finish one Type 1 UI Test and 22.64 secs to finish one Type 2 UI Test. Group 2 averagely takes 10.27 secs to finish one Type 1 UI Test and 57.05 secs to finish one Type 2 UI Test. Each of our participants (regardless of what group they have been assigned) averagely takes 11.58 secs to finish one Type 1 UI Test and 45.58 secs to finish one Type 2 UI Test, as shown in figure 3.

Table 1

Task Completion Time (s) Results from the participants

Participant	Type 1 - Trial 1	Type 1 - Trial 2	Type 1 - Trial 3	Type 2 - Trial 1	Type 2 - Trial 2	Type 2 - Trial 3
P1	6.72	52.67	14.22	42.32	22.84	14.71
P2	9.98	16.68	12.07	18.99	30.39	12.21
P3	4.22	7.75	6.34	16.15	41.32	10.46
P4	8.4	22.67	8.8	26	23	13.2
P5	37.31	12.52	5.95	22.59	16.42	56.46
P6	7.38	7.13	6.2	27.27	91.97	14.57
P7	6.64	9.13	6.73	14.45	47.28	11.15
P8	5.09	10.48	4.86	21.84	28.21	14.62
P9	4.49	10.69	4.63	12.7	28.09	38.7
P10	9.19	9.93	10.22	69.54	231.67	24.68
P11	7.4	12.87	6.76	60	37	28
P12	12.5	21.8	16.6	26.75	400.47	44.8

Table 1. Task Completion Time (s) Results from the participants.

Table 5

Level of Frustration Results from the participant's questionnaire (1 to 5 scale, higher means very frustrated)

Participant	Type 1 - Trial 1	Type 1 - Trial 2	Type 1 - Trial 3	Type 2 - Trial 1	Type 2 - Trial 2	Type 2 - Trial 3
P1	1	1	1	5	5	5
P2	2	2	2	4	5	5
P3	1	1	1	5	5	4
P4	1	2	1	2	2	1
P5	1	1	1	4	3	2
P6	1	1	1	4	5	2
P7	1	2	2	5	5	5
P8	1	2	1	4	5	3
P9	3	1	1	3	1	1
P10	1	1	1	1	1	1
P11	1	1	1	5	5	5
P12	1	1	1	2	1	3

Table 5. Level of Frustration Results from the participant's questionnaire.

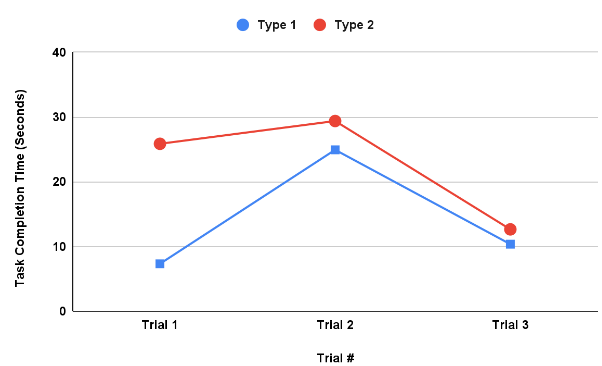

Figure 1. Group 1's Average Task Completion Time of Each Trial.

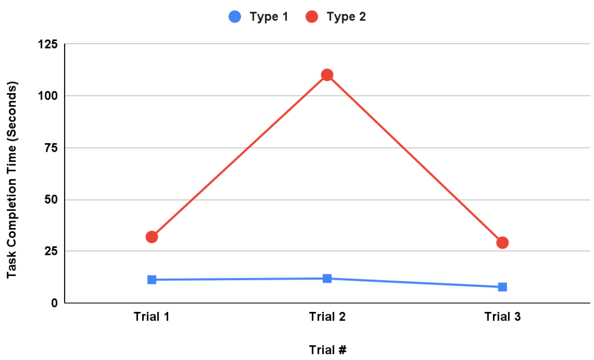

Figure 2. Group 2's Average Task Completion Time of Each Trial.

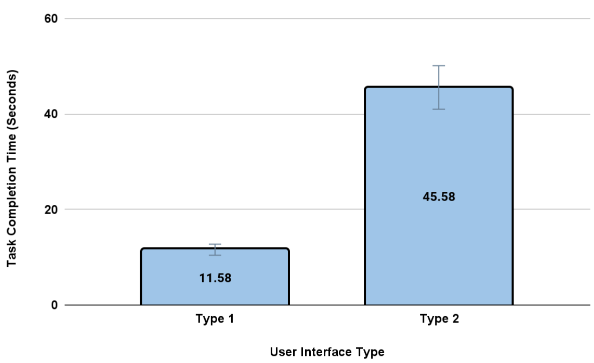

Figure 3. Average time for a participant (regardless of what group) to finish a Type 1 or Type 2 UI test.

To check whether the differences in our data are significant, we crunched our data into an ANOVA software to find out. The ANOVA table that we obtained can be seen in Table 2. Through our ANOVA analysis, we determined that the effect of user interface type on task completion time was statistically significant ($F(1, 11) = 7.800$, $p < .05$). This means that the different user interface types significantly influence the participant's task completion time. We also determined that the effect of user interface test trials on task completion time was statistically significant ($F(2, 22) = 3.874$, $p < .05$. This means that each of the user interface test trials also significantly influences the participant's task completion time. The difference within the interaction effect between the user interface type and user interface test trials was not statistically significant ($F(2, 22) = 2.190$, $p > .05$). This means that the different user interface types and their trial do not significantly influence each other to make a tremendous difference within our data. We also documented the resulting main effect in Table 3 and the effect sizes in Table 4.

Table 2

ANOVA Table for Task Completion Time (s)

Effect	df	SS	MS	F	p
Participant	11	32380.558	2943.687		
UI type	1	20801.201	20801.201	7.8	0.0175
UI type_x_Par	11	29336.891	2666.99		
UI test trials	2	16240.12	8120.06	3.874	0.0362
UI test trials_x_Par	22	46109.996	2095.909		
UI type_x_UI test trials	2	9893.907	4946.954	2.19	0.1357
UI type_x_UI test trials_x_P	22	49695.518	2258.887		

Table 2. ANOVA Table for Task Completion Time (Secs).

Table 3

Main Effect Means

Type of Mean	Participant Number, UI type Number, or Trial Number	Calculated value
Grand mean		28.58111111
Participant means	P0	25.58
	P1	16.72
	P2	14.37333
	P3	17.01167
	P4	25.20833
	P5	25.75333
	P6	15.89667
	P7	14.18333
	P8	16.55
	P9	59.205
	P10	25.33833
	P11	87.15333
UI type means	Type 1	11.58389
	Type 2	45.57833
UI test trials mean	Trial 1	19.91333
	Trial 2	49.7075
	Trial 3	16.1225

Table 3. Main Effect Means.

Table 4

Effect Sizes | Partial-eta-squared (np2)

Effect	Calculated value
np2 f1(UI type)	0.414878
np2 f2(UI test trials)	0.260467
np2 f1xf2(UI typexUI test trials)	0.166035

Table 4. Effect Sizes, Partial-eta-squared (np2).

In the end, we have evidence to prove that non-principled designs do not outperform principled designs. Within Figure 1, Figure 2, and Table 1, we can see how most participants perform faster in Type 1 (which is our modern-day designs). In table 5, we can see how Type 1 was favored by the participants for the least frustrating. Since the majority of participants reported that Type 1 is easy to use and least frustrating within the questionnaire, we can determine that principled designs also outperform non-principled designs in terms of accessibility and usability. However, despite the results, some things need to be pointed out about our data.

While we found that the testing software worked perfectly on the devices that it was developed on, this was not the case when participants were using it. To continue with the research, we provided an older backup version of the testing software for mobile users only. Although the backup version was made using a different tool, it was almost identical to the latest version of the testing software (which was made in *Processing*). However, since we had to resort to the backup version, we could not utilize the data-collecting software built into the latest version of the testing software and had to resort to manually timing the participant to record the time it took them to finish their trial. Because of this, the time we recorded from mobile users was not fully accurate, as there would be a delay in starting and stopping the stopwatch due to a person's reaction time. Because of this, the time could be off by a few milliseconds. We tried our best to reduce that delay, but we cannot fully rely on this to guarantee near-accurate timings. Thankfully, the times we received were not too heavily impacted by this, allowing us to continue with our research.

CONCLUSION

In conclusion, we determined that principled user interface designs perform better than unprincipled designs. Thus, the legitimacy of design principles like *Don Norman's 7 Principles* [6] will remain strong, as it helps to make our user interfaces more accessible, usable, and better performing. However, regardless of that, we still want to stress the importance of experimentation. The incorporation of rigid design principles into user interfaces has caused a standardization in appearance and function, but it has also caused stagnation. For further development in the user interface/user experience field, designers should strive to try to make something new and unique. What we have determined in this study is that we shouldn't exactly dismiss the principles, guidelines, or standards created by modern-day designs, but we should build upon these structures, integrating innovative ideas into the field. Designers and user interface developers should be able to decide how they want to make their user interfaces. Whether that is to utilize unconventional or conventional means is up to them. At the end of the day, we hope that we can potentially expand upon this with better testing software, more time, and a more polished hypothesis. Perhaps in the future, we can explore deeper into an understanding of what truly makes a good user interface design good.

ACKNOWLEDGMENT

We would like to thank our participants for helping out in our study.

REFERENCES

1. Cheng, Wu, C.-S., & Leiner, B. (2019). The influence of user interface design on consumer perceptions: A cross-cultural comparison. *Computers in Human Behavior*, 101, 394–401. https://doi.org/10.1016/j.chb.2018.08.015

2. Guo, Wang, X.-S., Shao, H., Wang, X.-R., & Liu, W.-L. (2020). How User's First Impression Forms on Mobile user Interface?: An ERPs Study. *International Journal of Human-Computer Interaction*, *36*(9), 870–880. https://doi.org/10.1080/10447318.2019.1699745

3. Jansen, Y., & Dragicevic, P. (2013). An Interaction Model for Visualizations Beyond The Desktop. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2396–2405.https://doi.org/10.1109/TVCG.2013.134

4. Mayer, R. (2020, December 8). *The era of UI unification*. Medium. https://uxdesign.cc/the-era-of-ui-unification-8c00a90bb2ea

5. Medryk, S. and MacKenzie, I. S., A comparison of accelerometer and touch-based input for mobile gaming, *Proceedings of the International Conference on Multimedia and Human-Computer Interaction - MHCI 2013*, (Ottawa, Ontario: ASET, Inc, 2013), 117.1-117.8.

6. Norman. (2013). *The Design of Everyday Things: Revised and Expanded Edition* (Rev. and expanded ed.). Basic Books.

7. Sampath, H., Merrick, A., & Macvean, A. (2021). Accessibility of Command Line Interfaces. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3411764.3445544

8. Usabilla. (2017, May 19). *A Short History of Computer User Interface Design*. Medium. https://medium.theuxblog.com/a-short-history-of-computer-user-interface-design-29a916e5c2f5