ITERATING WITH CROWD

BY

DHARTHI NAGARAJ KASHYAP

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Adviser:

Associate Professor Brian Bailey
ABSTRACT

This thesis studies how novice designers iterate on their designs using the feedback received from a popular online forum - Reddit. The goal was to understand the potential of online forums to serve the need for external feedback during the design process. The thesis also strives to understand the feedback and what motivated iterations. This could potentially aid development of rubrics for receiving design feedback using online forums.

This thesis also looks at the difference in perceived quality of designs between iterations and if the perceived quality improves. Domain experts were asked to judge (blind to condition) which iteration more effectively satisfied the design goals originally stated by the designer. The goal is to understand if the evolution of a design better satisfies the design goals due to the feedback received online.

Sentiment analysis was performed to test if certain qualities of feedback potentially trigger iterations on a design. Additionally, the thesis studied how much of the feedback received on the designs through Reddit was used to make design changes. The other aspect is to understand if the degree of transformation between designs correlates to the number of perceived differences.
Dedicated to my mother - for being a pillar of strength and teaching me how to be one too.
ACKNOWLEDGMENTS

First and foremost, I would like to thank my thesis advisor, Prof. Brian Bailey for his support and unwavering patience. He gave me freedom to explore and design the research domain space and gave me guidance when I needed it. Research would not have been this enticing if not for him.

I would like to thank my parents for simply being who they are and to have created their lives around my brother and I as centers of their world. My mother’s encouragement has not only made me who I’m but has filled me with courage to face anything in life. I would like to thank my brother for giving me the confidence to take anything head on. I would also like to thank my fiancé Amith for picking me up whenever I felt like I want to quit.

I learned a lot under Prof Craig Zilles when I functioned as a TA under him. He taught me not to beat myself when things do not go my way and to believe in myself. As an international student, this eased me up when I was intimidated.

This would be incomplete without thanking my friends Megha Ailavajhala, Sachin Kashyap who were family away from home. Even though I’m not mentioning many people who supported me through my journey, thank you so much for being there for me. I will always cherish my life in Champaign as some of the best days of my life.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION..................................................................................1

CHAPTER 2: RELATED WORK...........................................................................5

CHAPTER 3: METHODOLOGY..........................................................................8

CHAPTER 4: DATA ANALYSIS AND RESULTS............................................18

CHAPTER 5: DISCUSSION..............................................................................29

REFERENCES..................................................................................................31
CHAPTER 1
INTRODUCTION

Design education is an iterative process which requires the author of the design to aggregate design goals and what the design should communicate, reflect on the design practices and create a design that best matches the intended goals. After receiving feedback on the design, the design process is repeated to produce new iteration of designs. The revised design may be significantly different from the original design or could have some tweaks based on the feedback received to better match design goals. This iterative process is inherent to design education.

Feedback is the other aspect that is stressed on during design education. Traditionally, design schools materialize the feedback process as a teacher-student interaction in context of an assignment or project, or as a peer interaction. It’s important to receive feedback to understand the goals and expectations set. Feedback is a staple of iterative design process.

In today’s world, education is evolving to online schools, collaborating on projects remotely and more importantly, learning is a continuous lifelong process. This imitation of the traditional world is transforming to an online process, however, there are unique challenges. Feedback is a slower process with time differences and delay in turnaround time.
After interviewing several designers of varied backgrounds, it was clear that online feedback is becoming intrinsic part of design process. Platforms like Dribble, Behance, Instagram, Pinterest, etc are becoming hotspots to receive design feedback. These platforms attract designers of different backgrounds and expertise levels and hence, the feedback received is a lot different than the feedback received in a classroom or from a peer who has similar expertise levels.

Textual feedback is a common way of receiving feedback on designs. Online platforms are organized to receive textual feedback and non textual feedback such as likes, pins and shares.

Reddit is an excellent platform to receive textual feedback apart from receiving non textual feedback like upvotes/downvotes. One of the advantages of Reddit is that it’s possible to have a dialogue between person receiving feedback and the person giving feedback. Figure 1.1 shows the snapshot of a dialogue taking place. The other nice feature is to be able to upvote/downvote feedback received. This makes Reddit extremely interactive.

After looking at several subreddits of Reddit like logo_critique, design_critiques etc, we realised that certain authors of design came back with new iteration of design after receiving feedback for the previous iteration of design. Therefore, there was evidence that the iterative process is being mimicked online. This is not evident on platforms like
Dribble or Behance because these platforms present one’s work like a portfolio. The essence of iterative process is not captured like the way it is in Reddit.

However, a small population was iterating on the designs. This motivated us to study and understand feedback on Reddit, what made people come back and if there was improvement in quality of design through the iterative process.

Figure 1.1: Example of a dialogue between a user who posted the design and a user who gave feedback.

Research questions:

1. Does the perceived quality of the designs increase with iterations?

This was one of the foremost questions we had about the iterative process and feedback. Previous work has always showed that the quality of the creative success improves over iterations [10,11,19,28].
2. What sort of feedback makes designers iterate vs not iterate?
If iterative process was carried out, what was it about the feedback that encouraged iteration? We were interested in finding what made one iterate on his/her design.

3. When designs are presented without labeling i.e without revealing which is the “before feedback” and “after feedback” designs, do domain experts pick the before or the after version? Is there a clear preference here?

What will a simple A/B test results be? This is another perspective on the 1st result. We wanted to see if the designs are not only improving, but also being desired to match their goals.

4. We study the differences between the “before feedback” design and “after feedback” design. How many of these differences come from the feedback received on Reddit?

We wanted to see how much of this feedback was put into use by authors of design who received this feedback. If there was a genuine improvement in quality of the design, did the improvements really come from Reddit?

These questions gives us a clue as to the use of online communities and how these communities are contributing to the design process. Understanding the benefits and limitations of these communities is important as the design professionals increasingly leverage digital and online tools in their process.
CHAPTER 2
RELATED WORK

Iteration is at the core of design process. For a long time now, design education has emphasized on the necessity of iteration. Feedback is the basis of iteration and it helps a designer understand the differences between the design goal and the interpretation of the design by the intended audience [1]. Knowing these differences aids one to iterate and produce quality designs that best matches the design goal. Design critique and feedback are crucial for creative success [10,11]. Research has shown that designers who receive feedback during the iterative process produce higher quality designs as opposed to those who do not [19,28]. Feedback helps novices understand the key principles in a design domain, compare design alternatives [12,26], and justify the assumptions behind their work and understand the interpretation of their work [1].

There are many avenues to get design critique and feedback. The traditional studio critique is community activity where someone presents preliminary work and then critics—often teachers and peers—provide feedback to improve the design [13,14]. Peer feedback can also foster effective communication and collaboration among designers [8,18]. Feedback process enhances designer’s ability to create effective designs and understand how to give feedback to other designers [20,25].

In today’s age of breaking barriers, design education can be pursued from sitting anywhere around the world. There has also been a rise of online communities. Early
work on social translucence highlighted its potential for supporting different types of communication and collaboration, letting people learn through observation and imitation, and enabling people to notice and follow social conventions [29]. According to Amin & Roberts [16], professional communities share characteristics that emphasize the importance of social interaction for learning. Online communities provide another outlet but the quantity and quality of feedback typically falls below users’ expectations [22, 23]. Professionals often participate in online special-interest groups in order to gain professional status and reputation, as well as to improve work-relevant skills [9]. To achieve these aims, people have traditionally turned to communities of practice (CoPs). These communities share some characteristics like focusing on the domain of interest, learn by interacting and engaging in joint activities and discussions. Helping one another and information sharing are some of the key cornerstones of learning. This co-operative environment leads to a collection of experiences, stories and ways of solving problems [21]. Designers use the examples they find online to get ideas and learn how to create specific designs [17, 29].

With the advancement in technology, it’s possible to get crowd feedback [29] and feedback from online communities like Pinterest, Dribble, Behance etc as well. Even though the results showed crowd feedback is not cosmetic and has depth to the feedback received, the same study went on to show that the crowd and the domain experts did not always on how well a design satisfied design principles and the stated goals of the designer [29].
There is enough evidence about the importance of iterative process and feedback, and with the rise of technology, online communities are becoming important to establish reputation and as a pillar of learning.
Designers use different platforms like Behance, Pinterest, Dribble, and Reddit to showcase their work and get feedback. We approached designers to understand the feedback they were expecting from each of these platforms. We then looked at each of these platforms to perceive the feedback and its usefulness in improving quality of design. The Reddit community had the most constructive criticisms with insights to the design and the creator’s goals and other attributes of design itself like colors, the visual design and the interpretation it lead to. After looking at several subreddits like “design_critique”, “web design”, “design” and “graphic design”, we decided to collect data from “design_critique” subreddit because the posts were from diverse people from different parts of the world, different age groups and encompassing both novices and experts.

The methodology section is divided into data - collection and domain – expertise feedback. Data analysis and the results are described in the next chapter. Data – collection section describes how data was collected over a time period of ten weeks, and was refined further for the next phase. A data point consists of the design, the creator’s goal for the design as well as the feedback received on the designs. Domain - expertise feedback section focuses on the procedure for determining the perceived quality of design in relation to the creator’s goals and perceived quality of the feedback received on these designs. An A/B test was also conducted to understand if there is a
preference for iterations, and the degree of transformation between different iterations of designs and what were some of the main differences between the designs.

3.1 Data collection:

We collected data after observing “design_critiques” for over ten weeks. We were looking for posts indicating the redditor is coming back with a new iteration of the design based on the feedback received previously. Figure 3.1 shows two examples which illustrate how Redditors indicated that they were here for an iteration. The example on top shows one such post where the user has explicitly stated the previous post and the feedback received on the previous iteration of design. The example at the bottom shows how just the title of the post indicated to an update of a previous design.

![Example of how reddit user posts wordings that indicated an iteration over designs they had previously posted on sub reddit.](image)

After looking at the time frame that reddit users took to iterate on their designs, we decided to not consider certain posts which indicated the 2nd/ “after feedback” iteration came after more than a month from the 1st / “before feedback” iteration. There were also cases of multiple posts of the same design that were created by the sub reddit users which was usually done due to a lack of feedback. These posts were ignored and not considered as a data point for this study.
The data set can be primarily divided into 3 different categories:

1. **Complete data**: Complete data points consist of designs that have a “before feedback” iteration which has been designed by the authors to best fit the objective/goal; these data points also consist a “after feedback” iteration design which were created after receiving feedback on Reddit for the “before feedback” iteration. Complete data points were mostly graphic in nature - posters, business cards, menus for restaurants and door hangers.

2. **Partial data**: Partial data was mostly website data. Even though partial data is pretty much like complete data, the difference between partial and complete data is that we do not have access to the “before feedback” iteration of the website design. However, we do have the feedback received on the design.

3. **Control data**: Control data was just randomized data points that just had 1 iteration of design.

<table>
<thead>
<tr>
<th></th>
<th>Complete data</th>
<th>Partial data</th>
<th>Control data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data</td>
<td>32</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>points</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1**: The number of data points in different categories of data

There are 32 data points in the complete data category, 21 data points in the partial data category and 25 data points in the control data category. 4 of the 32 data points in
the complete data category had 3 iterations instead of 2. We dropped this data for calculation purposes because these were an anomaly for the rest 28 points of data and yet small in number to stand as an individual category of data for analysis purpose.

3.2 Domain expertise feedback:

The next step was to determine if the perceived quality of the design improved over iterations. Design perceptions could change given the context. Hence, we believed that the quality of the designs should be determined against the creator's goals. We then wanted to look at the measure of the usefulness of the feedback received on these designs, keeping the design itself and creator's goals as reference to this measurement. We were also intrigued by what domain experts would prefer if different iterations of designs were presented to them without actually labeling them. The other simple measurements that accompanied along were the degree of transformation between the 1st and 2nd iterations as well as the main differences that are visible. We thought it would be interesting to correlate the degree of change and perceived quality of designs over iterations.

After narrowing down on what our measurements had to be, the study was broken down into 3 different phases. We bundled phase 1 and phase 2 to be one task. Phase 3 alone was categorized to be another task. This was done so because phase 1 and 2 had pointers to the stages of time at which design was produced. For example, consider figure 3.3; this is one piece of feedback that was given for a design on second
iteration. We wanted phase 3 to be answered without looking at the feedback or knowing what design came first. We hired freelancers on Upwork (previously known as oDesk) to do these tasks. Phase 1 and 2 were designed to take about 1.5 hours – 2 hours to complete. Phase 3 required about 40 minutes – 45 minutes to finish. These tasks were designed to be questionnaire forms designed on google forms. For each of the tasks, we hired four experts. Each of them had at least 2-3 years of work experience in design with some of the designers having more than 8 years of experience. All of them had definitive experience in graphic design apart from experience in finer visual elements like illustrations, logos etc. The hired domain experts were compensated for their time.

The 3 different phases are:

1. **Phase 1** involved rating the perceived quality of designs on a 5 point scale where 1 suggested the perceived quality of design did not match the goal it was created for at all and 5 indicated the perceived quality of the design matched the goal perfectly. The 1st and 2nd iteration designs of complete data were rated. The 2nd iteration alone of partial data was rated. The control data was rated on 5 point scale as well. Figure 3.4 shows the snapshot of how the task looked on google forms.

2. **Phase 2** involved rating the usefulness of feedback received on designs on a 5 point scale where 1 indicated that the feedback was not at all useful and 5 indicated that the feedback was very useful. In figure 3.2, we can see that the
original poster would gain more insight into a feedback by having a dialogue with the person who left a particular piece of feedback. However, we chose to only use the initial feedback that was left to measure the usefulness of the feedback. This was done for all 32 data points of complete data, 25 data points on control data and only on the 2nd iteration design and feedback of the partial data. Figure 3.4 shows the screenshot how phase 2 looked from an expert’s perspective.

3. **Phase 3** was the A/B test and hence, it was conducted on the 32 points of complete data. Figure 3.5 shows the screenshot of how phase 3 looked from an expert’s perspective. Phase 3 consists of 3 parts:

   a. Part 1: 1st and 2nd iteration of designs were presented without specifying the respective label of the iteration and experts were asked to pick what design they preferred to best match the goal the design was created for.

   b. Part 2: Part 2 measured the degree of transformation/difference between the designs on a 5 point scale where 1 meant the designs are not that different, and 5 meant the designs are completely different.

   c. Part 3: This part was a free text box to specify the major differences between the two designs.
Figure 3.2: Example of a dialogue between a user who posted the design and a user who gave feedback.

Good improvement, better form than the old one.

How will this logo look in grayscale or black and white? I’m concerned that the wood texture will get lost in those iterations.

Figure 3.3: Example of a feedback that points to existence of previous versions of design.
Iterating with the Crowd

* Required

Page 3 of 230

GOAL: I am part of a engineering team working for a large company for my studies and being the only designer was tasked with coming up with all the visuals. The team is happy with this but I feel as if it needs more.

![Screenshot of Phase 1 form](image)

**2. Rate this design**
Rate the perceived quality of the design based on the creator's design goal.

1 2 3 4 5

Poor ⬜ ⬜ ⬜ ⬜ ⬜ Excellent

« Back  Continue »

1% completed

Figure 3.4: Screenshot of Phase 1 form
Figure 3.5: Screenshot of Phase 2 form
After these tasks were done, we moved to data analysis and studying the results and their significance. This is presented in the next chapter.
This chapter describes all the tests we did on the data and all the results obtained from the tests.

When we took a look at the distribution of the perceived quality of the designs, the distribution is normal as shown in figure 4.1. We looked at some of the measures of central tendency and the results are listed in tables 4.1, 4.2 and 4.3.
When we take a look at table 4.1, we can see that the means for 2\textsuperscript{nd} iteration are higher than the means for 1\textsuperscript{st} iteration. It is also comparable to partial data as these are 2\textsuperscript{nd} iteration of designs for website data. The means for control data are much lower than the means for the overall data set or any of the other categories of data.

**Table 4.1 : Means for the perceived quality of design rated by the experts**

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st iteration</td>
<td>2.571428571</td>
<td>2.785714286</td>
<td>2.821428571</td>
<td>2.5</td>
</tr>
<tr>
<td>2nd iteration</td>
<td>3.178571429</td>
<td>3.178571429</td>
<td>3.071428571</td>
<td>3.464285714</td>
</tr>
<tr>
<td>Partial data</td>
<td>3.173913043</td>
<td>3.347826087</td>
<td>3.173913043</td>
<td>2.869565217</td>
</tr>
<tr>
<td>Control data</td>
<td>2.68</td>
<td>2.88</td>
<td>2.76</td>
<td>2.76</td>
</tr>
<tr>
<td>For all data</td>
<td>2.870689655</td>
<td>2.956896552</td>
<td>2.905172414</td>
<td>2.870689655</td>
</tr>
</tbody>
</table>

When we take a look at table 4.2, we can see that the standard deviations for the perceived quality of design rated by the experts.

**Table 4.2 : Standard deviation for the perceived quality of design rated by the experts**

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st iteration</td>
<td>0.9200874125</td>
<td>1.397276262</td>
<td>1.248808956</td>
<td>1.347150628</td>
</tr>
<tr>
<td>2nd iteration</td>
<td>0.8629652338</td>
<td>1.278122797</td>
<td>1.11980724</td>
<td>1.426784597</td>
</tr>
<tr>
<td>Partial data</td>
<td>0.9176629355</td>
<td>0.9334586382</td>
<td>0.7873265148</td>
<td>1.116070802</td>
</tr>
<tr>
<td>Control data</td>
<td>0.988264472</td>
<td>1.332916602</td>
<td>0.9695359715</td>
<td>1.3</td>
</tr>
<tr>
<td>For all data</td>
<td>0.9466662091</td>
<td>1.327909383</td>
<td>1.029803259</td>
<td>1.335340194</td>
</tr>
</tbody>
</table>
Table 4.1 shows the standard deviation for different categories of the data set. We can see that the standard deviations for the 2nd iteration of designs and partial data is lower than other categories which suggests higher number of data points closer to the mean. We can see that the standard deviations of 1st iteration of designs is higher which suggests a more spread out distribution of data points and away from the mean.

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st iteration</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2nd iteration</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Partial data</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Control data</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>For all data</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 4.3 : Modes for the perceived quality of design rated by the experts**

Similarly, there is an increase in the mode of 2nd iteration of designs as opposed to the 1st iteration of designs. We can even see the mode of the partial data has increased. Control data does better in terms of mode than 1st iteration of designs.

For the next part, we wanted to have a look at the feedback and study why people came back for feedback with further iterations of design. It seems logical to assume that authors of the designs came back with further iterations because they were happy with the feedback they received and it was valuable in producing better perceived quality of
designs as shown in tables 4.1, 4.2 and 4.3. The questions now arose as to what was special about the feedback given and how much of this feedback was actually used to produce revised iterations.

We approached the composition of the feedback from a sentiment analysis perspective. We aggregated the feedback given for the 1st iteration of designs of the complete data and the feedback given to the 1st iteration of designs for the partial data as well and performed sentiment analysis. The categories we used to label the feedback are positive, negative and neutral. We compared this set of feedback distribution against the control data’s feedback composition. The results are shown in Figures 4.2 and 4.3. We used IBM Watson’s AlchemyAPI to perform the sentiment analysis and the results are very intriguing and yet what we expected to see.

![Distribution of feedback for people who came back with new iteration designs](image)

*Figure 4.2: Distribution of feedback for people who came back to Reddit with new iteration of designs*
For designers who came back with new iteration of designs, there is 54.1% of critical feedback as opposed to 42.6% critical feedback for authors of these designs who did not come back. Further, critical feedback accounts for the majority of all feedback gotten to pave way to new iteration of designs whereas positive feedback accounts for majority of feedback in Figure 4.3 for people who did not come back. Retrospectively, we can also see that praise may not necessarily lead to improvement as opposed to offering considerable amount of critical feedback.

Neutral feedback gotten on both these sets of populations remains more or less the same amount. This result is extremely useful in understanding the composition of a good feedback.
The other aspect that we were interested in studying is how much of the feedback is actually being used to iterate and transform designs to attain higher perceived quality of designs to match the goals the designs were being created for. Phase 3 collected a free text response to what the major differences between the designs were. This information was collected for all 32 data points of complete data. Omitting the 4 data points that had 3 iterations, we studied the 28 data points.

A total of 99 unique key differences were found based on the key differences listed by the domain experts. 71 of these differences came from Reddit feedback. This means at least 71.71% of the changes done on the “after feedback” designs came from Reddit feedback. Of course, there is a good possibility that some of these feedback might be repeated during feedback from other mediums like peer reviews, mentor feedback etc. However, this does tell us that Reddit is a good place to get feedback and from the results obtained in 1, the feedback could potentially increase the perceived design quality.

Further, at least for 57.14 % of the sample space of designs, at least 80% of the design changes was reflected in Reddit feedback. This number increases to a good 75% of the designs if we consider 70% of the design changes matching Reddit design feedback.
Figure 4.4: Distribution of the key differences between the designs

Figure 4.4 shows the distribution of key differences. The blue bars of the graph indicate the differences that have a possible origin from Reddit feedback for the 1\textsuperscript{st} iteration of designs. The red bars indicate the number of key differences that did not originate from the Reddit feedback. It's evident from the graph that the blue bars dominate the graph. There are a lot of stand alone blue bars and there are a very few number of cases where the red bars equal the blue bar and one case where the red bar over takes the blue bar. This bar very clearly tells us that most of the changes that were made to the designs in the 2\textsuperscript{nd} iteration came from feedback received on Reddit on the 1\textsuperscript{st} iteration of designs. This means that the feedback received on Reddit is leading to design...
changes and in turn improving the quality of design as shown in the previous results in this study.

After studying the results so far, we wanted to see if the feedback and the design changes are leading to improved design iterations which match the goals of the designs better. We ran a A/B test to do this and this was phase 3.

<table>
<thead>
<tr>
<th>Design Iteration</th>
<th>Domain expert 1</th>
<th>Domain expert 2</th>
<th>Domain expert 3</th>
<th>Domain expert 4</th>
<th>Mean</th>
<th>Highest</th>
<th>Lowest</th>
<th>Total data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st design iteration</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>9.5</td>
<td>12</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>2nd design iteration</td>
<td>22</td>
<td>16</td>
<td>17</td>
<td>19</td>
<td>18.5</td>
<td>22</td>
<td>16</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 4.4: Summary of the preferred design iteration for the domain experts

From table 4.4, we can see that the domain experts mostly chose 2nd iteration as the design they preferred among the 2 designs presented to them. Expert 2 preferred the least number of 2nd iterations of all the 4 domain experts and the number was still 16 of 28 designs. This means that even Expert 2 preferred 2nd iteration for 57% of the cases in a sample size of 28 data points.

On an average, the 4 domain experts preferred the 2nd iteration 18.5 times out of a total
of 28 points. This makes it 66.07% preference for the 2nd iteration of the designs which is what we expected to see. We ran a chi square test to make sure the preference is 2nd iteration of design.

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st iteration of designs</td>
<td>38</td>
<td>56</td>
</tr>
<tr>
<td>2nd iteration of designs</td>
<td>74</td>
<td>56</td>
</tr>
<tr>
<td>Total number of observations</td>
<td>112</td>
<td>112</td>
</tr>
</tbody>
</table>

Table 4.5: Chi-squared test table

H₀ = There is no preference for designs (1st iteration designs preferred = 2nd iteration designs preferred)

Hₐ = There is a preference (1st iteration designs preferred ≠ 2nd iteration designs preferred)

Chi squared equals 11.571 with 1 degrees of freedom.

The chi squared value rejects the null because the value is statistically significant with p value less than 0.05. Therefore, there is a strong preference for the 2nd iteration of designs. Therefore, we can clearly see that the feedback from Reddit is helping with iterations of better quality designs and the experts prefer the iterations performed with the feedback inculcated in design changes.
Additionally, we wanted to see if there is a correlation between the difference in quality of designs between iterations and the total number of differences between them. However, the correlation coefficient r’s value is -0.066 which is not statistically significant. Therefore, the difference in quality does not vary the number of differences. The covariance between them is very low. This is demonstrated in figure 4.5.

Figure 4.5: Correlation between Number of differences between the designs and the difference in perceived quality of the designs
Figure 4.6 plots the degree of difference/ transformation rating by experts vs the number of differences. We expected to see a significant correlation between the two i.e., higher the number of differences between the iterations of designs, higher the degree of transformation/difference. The line plots above do not show that correlation. When Pearson’s r, correlation coefficient was calculated, it was not statistically significant.
CHAPTER 5
DISCUSSION

This study addresses a number of key points. The results show that the quality of the design improves over iterations which is what we had expected based on previous studies in the domain. An A/B test was run to understand what iteration of designs do the experts prefer when presented with the goal, and the domain experts mostly chose the 2nd iteration of designs which is consistent with the results we have obtained for improvement in perceived quality.

The other key aspect we looked at is the motivation the feedback provides for iteration. The sentiment analysis showed how negative valence in the feedback can prompt additional iteration.

Feedback is essential for iteration. It is necessary to close gap between the design goal and the interpretation of the design. Therefore, each piece of feedback is important to understand the gap and close it. The results obtained show that feedback is taken seriously for making design changes and producing the next iteration. A good percentage of design changes seemed to be motivated by the feedback received on the designs.
Future research work could potentially study how “cosmetic” or “deep” the intended feedback and design changes were and what was the correlation to the degree of transformation. When we looked at degree of transformation/ difference between designs and the total number of key differences between designs, the correlation was statistically insignificant. There is a need to find out why this correlation is statistically insignificant. This could also be addressed by studying how much depth the feedback had while recommending design changes. This study paves way for future work that includes developing rubrics for giving feedback in online communities. This could help people receiving feedback and the person giving it because the expectations are clear. Online communities are able to create an effective environment for receiving feedback, giving feedback and learning domain knowledge despite differences in time and geography. Online communities could pave way for a whole new way of receiving feedback that can be used for design changes and to meet the goals for which the design was created.
REFERENCES


