Group 5: Adaeze Ezekwerre, Annie Doctrow, Lesley Pace, Marissa Villano
INFO 615 - Spring 22
Dr. Jina Huh-Yoo
5 June 2022

Experiment Plan: Insight Timer

Abstract

Meditation and mindfulness apps are increasingly being used in mental health discussions. We will investigate the impact these apps play in alleviating the pandemic's ongoing emotional effects on young adults. Specifically, we will focus on incorporating gamification into these apps. We hypothesize that this global change in the app will increase young adults' usage of the meditation and mindfulness app Insight Timer. Furthermore, if the daily frequency of user sessions increases over time, we expect our secondary indicator, daily mood, to demonstrate general improvement over the same time. The key metric, daily session frequency, will collect quantitative data. Because the application can track this critical measure without any outside interference, data collection will be unmoderated using user logs. The secondary metric will be used to collect quantitative and qualitative data via self-reporting and mood check-ins. If we add gamification to Insight Timer, the increased engagement will encourage user engagement and utilization of the app, which will improve the daily frequency of user sessions. Incorporating this will encourage users to achieve goals within the application and make tasks less overwhelming and more entertaining.

Introduction

As a result of the COVID-19 pandemic, the cost to protect physical health resulted in declined mental health. Consequences such as isolation, job loss, school closures, lack of childcare, restricted access to mental health care, and greater exposure likelihood put young adults at a higher risk of facing mental health issues during the pandemic (Panchal, et al 2021). Already at a higher risk of poor mental health even prior to the pandemic, Panchal et al write that fifty-six percent of young adults reported symptoms of anxiety and or depressive disorder during the pandemic (2021). To help combat these symptoms, virtual meditation and mindfulness resources are gaining popularity as free or low-cost solutions. A longitudinal observational study by Stecher et al that examined app-based meditation habits and the associated mental health benefits found that routine use of the app allowed for users to form reflexive meditation habits and that the users associated their routine use with greater perceived mental health benefits (2021). Furthermore, with the key being routine use of the app, a systematic review of 1,017 intervention studies targeting engagement of adult web and or mobile users suggests that “gamification can increase engagement in online programs and enhance related outcomes, such as learning and possibly health behavior” (Looyestyn 2017). Thus, as consequences of the pandemic linger on, this experiment aims to focus on increasing young adult utilization of the meditation and mindfulness app, Insight Timer, through introducing gamification.

Methods

Hypothesis

If we apply gamification to Insight Timer, then the added engagement will drive user utilization of the app, which will positively impact the user session daily frequency. Additionally, if the user session daily frequency increases over time, we expect to see our secondary metric, daily mood, to reflect general improvement over the same period of time.

Data Collection Plan

What Data to Collect?

To collect quantitative data that tracks use and consistency of the app, Insight Timer, our key metric will be session daily frequency. This existing metric communicates how many times a day a user participates in one of the app's mindfulness or meditation sessions.

Though not central to the success of this experiment, as a secondary outcome metric, we will track user mood through the app's daily mood tracker. The daily mood tracking feature on the app prompts the user to check-in by selecting an emoticon correlating to either "great," "good," "okay," "bad," or "awful." Once the overall mood has been selected, another page displays different suggested categories as to why the user feels the way they do and allows for them to select any of the suggested additional feelings to complete their mood check-in.

Why are we Collecting the Data?

It is our primary focus to observe whether the application of gamification in this mindfulness app will lead towards increased utilization on a consistent basis. Consistency can be measured through various metrics, however we felt this will be best exemplified through our key metric, session daily frequency. This metric was selected as it reveals current user behavior and methods of engagement within the app for individuals as well as overall trends for the user base. By utilizing this as our key metric, we can analyze for trends in the user's utilization in terms of average daily use and number of sessions consumed daily between the two test cells. Since we hypothesize that adding gamification to the app will increase daily user utilization rates, session daily frequency represents a direct correlation to the expected user behavior. Ideally, in other words, an increase in the metric will indicate that the implementation of the new feature is having the desired impact on users.

We also have a listing of secondary metrics focused on the emotional state of the user by inclusion of "affect" data. This is of deep intrigue to us considering the rise in mental health issues over the past few years driven by pandemic fatigue, virtual battle fatigue, and racial battle fatigue. This can be exemplified by the rise in distress disclosures (Zhang 2021) and people expressing negative sentiment in therapy (Dwyer 2022). If our primary outcome is successful, wherein users increase their daily utilization, it will be interesting to also look for this secondary outcome to analyze if their mood improved overtime, as well. While this is not central to the success of this experiment, it would reflect a positive user experience as a side effect result of the proposed change.

Already at a higher risk of poor mental health even prior to the pandemic, our more targeted focus on young adults is purposeful due to the belief that their routines of consistency in education and socialization were interrupted to their detriment as a result of the COVID-19 pandemic and we would like more insight into methods to improve their well-being. If our hypothesis proves successful in that adding gamification will increase utilization, then our hope would be that the susceptible age group gains utility in the materials consumed which will work to enhance their mental wellbeing over time. We also hope to see this as a secondary outcome through mood improvement.

When is the Data Collected?

We plan to collect longitudinal data to analyze how user behavior adjusts over time. Longitudinal data suits this study best since the primary metric focuses on increasing utilization rates of users in terms of consecutive days, thus the change we expect to see will happen over time, not in a snapshot or single day. For this reason, we will export daily user data on a weekly basis over the course of six months (June 2022 through December 2022). We expect six months to provide sufficient time for users not only to form a habit, but also to break one. Moreover, we can track for any impact the different seasons may impose on user's daily utilization, such as Seasonal Affective Disorder (SAD) heightening during the winter months causing users to seek more mindfulness or meditation compared to during other months. Each week will

be defined as beginning on Monday at 12:00AM local time and ending on Sunday at 11:59PM local time for a clean cut off. Given the nature of the COVID-19 pandemic and young adults' predisposition to mental health struggles, the data will be collected contextually to evaluate if the expected impact of gamification on our primary metric can withstand the multitudes of external factors that influence what actions the users being studied prioritize and how they spend their time.

How is the Data Collected?

We will be collecting quantitative data from the primary metrics. Because the program can track the key metric without any outside intervention, collecting this data will be unmoderated with the help of user logs. The secondary metric will be gathered by self-reporting using the mood check-ins. The secondary metric's data will be quantitative as well as qualitative. Quantitatively, we can take the overall mood selection and assign it a scaled numerical value to track over time. Qualitatively, we can utilize the explanatory selections and additional feelings to support any trends we may see in the user data.

The test cells will be randomly selected to ensure they are representative of the portion of the user base we hope to learn about. In terms of our user segmentation analysis, we segment by stable characteristics and will filter by user age to identify the young adults, which we define as individuals between the ages of 18 to 29, and location, which will be the United States. We believe that this age group utilizes technology similarly and share common values and expectations. Due to the exploratory and global nature of this experiment, we will not segment by behavior and will include young adult users of all experience with Insight Timer, such as new users, existing users, and current routine users.

How Much Data to Collect?

We will be collecting data from a large sample size, specifically 10% of users. As adding the gamification aspect to the app is not necessary for users to continue using the app, there is no urgency to implement this feature, though we hope the gamification will increase user interest and excitement, draw more users to the app, and retain current users. At the same time, the content consumed should contribute to healthy coping habits for mental wellbeing, but because there are many factors involved with a user's mental health and vary widely person to person, a large sample size is needed for us to see if there are significant trends in our secondary outcome metrics, rather than coincidences. For this reason, testing this feature with 10% of users over this time period will allow us to gain insight into the popularity of this gamification and feel comfortable saying gamification helped user's mental health.

Data Analysis Plan

Choice of Analysis Method

We hypothesize that by adding gamification to the app, the user's daily frequency will increase. By using an independent t-test to compare the test cells, we will be able to see the user's daily frequency with the gamification side by side with the control group without the gamification. We will utilize the standard p-value of .05 to ensure that differences between the two test cells are by random chance.

Analysis Method Rationale

Analyzing our data using an independent t-test will be appropriate for this experiment because our primary metric is quantitative, numerical, and continuous. This metric consists of collecting the daily frequency in which users participate in sessions on the app. With this being our key metric, using a t-test will allow us to analyze two samples, users utilization in terms of their daily average use, and the number of sessions they have consumed daily, to evaluate our hypothesis to see if adding gamification to the app had any influence on the users.

Discussion

Interpretation of Results

If we see a statistically significant increase in the session daily frequency in cell B compared to that of cell A, then our experiment yielded an expected, or positive, result. Although statistically significant, due to the current exploratory phase, it is critical to evaluate the significance of the change itself and consider its importance. While an increase in the session daily frequency suggests users find value and joy in the services provided through the app, increased utilization could also suggest the app has become addictive and potentially detracting from users' mindfulness when not on the app. That being said, however, a statistically significant improvement in users' moods over time in cell B compared to that of cell A, as well, would support the decision to implement the change. Conversely, an unexpected or flat result in the users' moods over time would support the decision to question the focus of the experiment and our understanding of the user group (King 2017).

If we see a statistically significant decrease in the session daily frequency in cell B compared to that of cell A, then our experiment yielded an unexpected, or negative, result. In the case of an unexpected result, we will reevaluate our understanding of cell B as our hypothesis did not resonate with them. Moreover, a statistically significant improvement in users' moods over time in cell B compared to that of cell A suggests that while our original hypothesis was off point, something positive occurred, thus redefining the focal point of the experiment would serve future utility. Critical questions in our reevaluation will include: are users utilizing the feature as intended, do users care about other app features more, and does our focus only appeal to a narrow group of users? On the contrary, if we have a statistically significant decline in users' moods over time in cell B compared to that of cell A, then paired with the overarching unexpected result, the data suggest that this design should not be pursued any further (King 2017).

Lastly, if we see no statistically significant increase nor decrease in the session daily frequency in cell B compared to that of cell A, then our experiment yielded a flat, or inconclusive, result. A flat result could be the product of our experiment requiring more power to actually see any significant result or that our test design does not influence the user behavior enough to be seen at scale. Either way, an analysis on sample group and size, test duration, and any external factors will be conducted to determine if the hypothesis has any merit. Additionally, we will analyze for statistical significance in the users moods over time in cell B compared to that of cell A. If there is either a statistically significant improvement or decline in cell B users' moods compared to that of cell A, then we will further investigate and analyze potential insight on the user behavior that could help explain our flat result (King 2017).

Design Implications and Potential Design Failures

Initially, this design could fail if users do not like the addition of gamification in the app. Users could potentially find the addition unnecessary and potentially implement a competitive feel that they may not desire. Any of these challenges may catalyze change aversion and trigger users to not return to the app. Which could then yield less than the desired of results for this design plan. Alternatively, users could love engaging with the gamification and become addicted to that side of the app, instead of focusing on the meditation aspect. This could affect their mental health, potentially skewing the results of this design experiment. The goal is not to take away from the main feature of the app, meditation, thus a major design failure would be that the gamification overshadows the meditation and relaxation features of the app.

There is also an opportunity for the emergence of a novelty effect to appear with users. The novelty effect can be defined as increased positive results from users due to a change but over time as excitement fades around the feature so does usage and engagement. As we have not chosen to focus on a design cell of new users versus existing users there is a chance that this could impact our results.

Additionally, we chose to more deeply focus on the impact of gamification on young adults which could further exacerbate the impact of novelty due to a shortened attention span from younger users. However, six months should be an adequate time frame to combat this.

Ultimately, potential lessons learned are as follows:

- There could be a vaster set of design cells to ensure that we are getting a full view of impact per demographic breakout.
- There should be a consideration around lengthening testing times to ensure that we are taking change aversion and novelty effect into account.
- Secondary research gathering is of utmost importance to contribute to proof of concept and supporting reasoning to pursue a certain hypothesis.
- It is significant to evaluate other methods beyond gamification to ensure that there is a vastness in scope of testing methodologies for increased usage.
- It is significant to ensure that metrics are actually aligned with needs of the hypothesis and that it is possible to think outside of the box for how to best measure impact.

Conclusion

Recap

By introducing gamification into apps used by young adults, we will be able to see whether these issues benefit this Cell B for the better. We will do this by focusing on collecting quantitative data and qualitative data of our primary and secondary metrics, session daily frequency and daily mood, and analyzing the longitudinal data for statistical significance by using a t-test. Through performing this experiment, we hope to find whether gamification on the Insight Timer app will lead to increased use, as well as benefit user's mental health.

Future Goals

In the future, we would perform this experiment on users of different age groups. This way, we will be able to see if the addition of gamification in other age groups will create an increase of attendance and use all around, as well as an improved mental health in other age groups. Additionally, we will collect data and perform experiments to test sustained engagement, expanding on our testing of increased engagement.

Contributions

The findings of this experiment will contribute insight on user reactions and behavior for both problems motivating this design, Insight Timer's user engagement and the negative impact on the mental health of young adults as a result of the COVID-19 pandemic. First, as Looyestyn et al found in their review of studies, gamification increased adult users' online engagement, however, the forms and application of gamification varied throughout the studies compared (2017). The results of this experiment will contribute findings specific to the context of our hypothesis to support the question of how best to apply gamification for effective engagement. Next, as Stecher et al found in their study, app-based meditation habits were associated with positive mental health benefits for the routine users (2021). With the goal of increasing user utilization, the findings of this experiment will further explore the impact of a specific gamification experience design on user engagement and routine use of meditation and mindfulness apps. Furthermore, all else equal, the secondary metrics of this experiment, daily mood, will

provide insight on the importance of user engagement and utilization of the app on the users achieving the mindfulness the app sets out to offer. Altogether, this experiment will contribute insights to continue the design of a potential solution for the prominent mental heal crisis amongst young adults.

References

Dwyer, K. (2022). *Why therapists are worried about mental health in America right now*: Times insider.

King, Churchill, E. F., & Tan, C. (2017). *Designing with data: improving the user experience with A/B testing* (First edition.). O'Reilly.

Looyestyn, Jemma, et al. (31 Mar 2017*). Does gamification increase engagement with online programs? A systematic review.* PLoS ONE, vol. 12, no. 3, p. e0173403. Gale Academic OneFile, link.gale.com/apps/doc/A487976425/AONE?u=drexel_main&sid=bookmark-AONE&xid=2442e047. Accessed 1 June 2022.

Stecher C, Berardi V, Fowers R, Christ J, Chung Y, Huberty J. (2021, November 23). *Identifying App-Based Meditation Habits and the Associated Mental Health Benefits: Longitudinal Observational Study*. J Med Internet Res. Retrieved April 10, 2022, from https://www.jmir.org/2021/11/e27282 DOI: 10.2196/27282

Panchal, N., Kamal, R., Cox, C., & Garfield, R. (2021, February 10). *The implications of COVID-19 for mental health and substance use*. KFF. Retrieved April 10, 2022, from https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/

Zhang, Renwen, Natalya N. Bazarova, Madhu Reddy. (2021). *Distress Disclosure across Social Media Platforms during the COVID-19 Pandemic: Untangling the Effects of Platforms, Affordances, and Audiences*. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 644, 1–15. https://doi.org/10.1145/3411764.3445134