Apps with Benefits: Using Benefits and Burdens to Predict Mobile App Usage

Katherine M. Cheng
Stanford University
Stanford, CA 94305, USA
katiemc@stanford.edu

Veronica J. Lin
Stanford University
Stanford, CA 94305, USA
vronlin@stanford.edu

Kush Nijhawan
Stanford University
Stanford, CA 94305, USA
kushn@stanford.edu

Ashley Westhem
Stanford University
Stanford, CA 94305, USA
awesthem@stanford.edu

Michael S. Bernstein
Stanford University
Stanford, CA 94305, USA
msb@cs.stanford.edu

CHI'17 Extended Abstracts, *May 06-11, 2017, Denver, CO, USA*
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053276

Abstract
How do mobile apps keep users coming back? Suh et al. [15] proposed that the level of burden placed on a user has a negative effect on user retention. They developed the User Burden Scale, and showed that computing systems still in use had lower burdens than those that were abandoned. What is not captured is how the added benefits a system provides increases user retention. We hypothesize that both benefits and burdens of a mobile app predict usage. To show this, we design and validate a User Benefit Scale to complement the User Burden Scale, for the evaluation of benefits of mobile apps. Our scale consists of four constructs: if an app is 1) useful and informational, 2) enjoyable and enables pursuit of interests, 3) enables social interaction, and 4) has good usability and visual/ interaction design. We administered the benefit and burden scales to 347 participants. Our results suggest that benefit is more predictive of mobile app usage than burden, and our model of app usage includes constructs from both the benefit and burden scales.

Author Keywords
Mobile apps; acceptance; user retention; user experience; measurement; prediction; benefit; burden

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

User Benefit Construct Definitions

Useful and Informational: Help users with planning and logistics, give users timely and relevant information, and assist users in getting through the day more efficiently. This term is used in reference to the function or purpose of the app. Ex: Venmo, Calendar.

Enjoyable and enables pursuit of interests: Offer a way for users to pursue their interests or hobbies and express themselves. Ex: ESPN, Candy Crush.

Social interaction: Allow users to stay in touch and communicate with friends and family. Ex: Facebook, Snapchat.

Usability and visual/ interaction design: Characterized by an appealing and intuitive user interface and by the ease and convenience of use. This term is in reference to the user's experience with the app. Ex: Chrome, WhatsApp.

Introduction

Why do apps fail? In addition to timing, marketing, money, and competition, the benefits and burdens of mobile apps may predict continued app usage. Prior work in HCI has focused on minimizing burdens [15]; however, maximizing benefits may also serve to improve the initial adoption, retention, and overall user experience. This paper aims to develop a model for user benefit in order to predict user's behavior based on benefit and burden ratings for mobile apps.

Related Work

To create the User Benefit Scale, we reviewed literature to examine why some products capture user's attention [6], what factors lead to users' acceptance or rejection of information technologies [5], and the usability features of media technologies [1], which guided our understanding of the characteristics and dimensions of user benefit (Table 1). In addition, we adhered closely to the methodology used to develop the User Burden Scale [15]. Although previous studies and theoretical frameworks provide reasons why users might choose to use a mobile app, there does not yet exist a systematic method for categorizing and measuring user benefits.

Themes	Evidence of theme
Task completion	[1], [2], [4], [5], [6], [7], [9], [10], [12], [16]
Storing and sharing information	[7]
Social connectedness	[7], [16]
Enjoyment	[3], [4], [9], [10], [12], [16]
Not burden	[1], [3], [10], [13], [14]
Self-expression	[3], [13]

Table 1. Related literature about user benefit.

Preliminary Interviews & Pilot Survey

Informed by the themes from our literature review, we conducted nine 30-minute interviews to determine possible categories of benefits. A sample question from our interview protocol asked: "If your phone ran out of storage, which 3 apps would you keep, and why?" We used results from the interviews to expand and refine our benefit categories: useful tool, provides information, social, entertainment value, usability, frequent app updates, and other.

Next, we created and deployed a 27-item pilot survey, containing similar wording to that of the User Burden Scale. We used convenience sampling within the authors' social networks, and received 106 responses to our online survey. Participants were asked to rate 2 mobile apps (one they currently use, and one they once used but have since abandoned), which they chose from a list of 26 apps. The list of 26 apps was compiled based on apps referenced by participants in the preliminary interviews. (Since this list of 26 apps was inherently limited, in the final Benefit survey described later on, we devised a method that elicited greater variance in the apps selected by users.)

With the data from this pilot survey, we conducted a principal component analysis to extract factors. Results showed that only 4 factors had eigenvalues greater than 1, contrary to our 7 hypothesized categories of benefit. These 4 constructs, accounting for 56.3% of the total variance in our survey responses, were retained for our user benefit scale. These constructs are defined in the column to the left. Through the PCA, the informational and useful benefit categories were combined into one construct, as were enjoyment and self-expression. We used these results from the pilot survey to select 20 items for the final Benefit survey,

Currently used app	# of responses
WhatsApp	31
Facebook	18
Instagram	15
YouTube	13
Chrome	13
FB Messenger	9
Twitter	9
TOTAL	347 responses (170 distinct apps)

Abandoned app	# of responses
Facebook	21
Twitter	21
Candy Crush	10
Hikes	10
Instagram	10
Tinder	9
Skype	8
TOTAL	347 responses (190 distinct apps)

Table 2. List of most frequently rated currently used and abandoned apps.

selecting those items that explained the greatest variance in each construct.

Final User Benefit Scale and Question Set

Data Collection

We deployed the final Benefit survey on Amazon's Mechanical Turk (mTurk), and collected survey responses from 347 subjects. Our online survey was estimated to take about 7 minutes to complete. Mechanical Turk participants were compensated $0.25 for their task.

One of the primary modifications we made to the final Benefit survey was to prompt participants to rate a more varied sample of mobile apps. Instead of having participants select from a predetermined list of 26 mobile apps, we asked them to list several apps they currently used or had abandoned. To avoid the potential pitfall that most respondents would choose to rate the most popular apps, we introduced an element of randomization. Users were asked to list a few apps, and to use the last app they had listed for the survey. This strategy was successful; we received responses for 170 distinct apps for the currently used apps, and 190 distinct apps for the abandoned apps, as opposed to the 26 apps rated in the pilot surveys. The most frequently rated apps are listed in Table 2.

The final benefit and burden survey incorporated a few other small modifications from the pilot survey. For example, we randomized the order of the survey items and incorporated catch questions to ensure respondents' attention. The two catch questions were: "If you're still alive, click 'Very Often'" and "Select 'A little bit of the time' for this statement."

Results

We proposed in our hypotheses that mobile app usage could be predicted by the relative benefits and burdens an app provides. We started by looking at whether benefit alone, or burden alone, was predictive. We later examined the relationship between benefits and burdens, and whether the relative benefits and burdens are predictive of mobile app usage.

Do used and abandoned apps differ in average benefits and burdens? Indeed, the total benefit of used apps ($m=42.39$) exceeded that of abandoned apps ($m=26.55$) ($t(691)=13.68$, $p<.001$). On the other hand, the total burden of used apps ($m=16.18$) was less than that of abandoned apps ($m=18.88$) ($t(689)=-2.38$, $p<.05$). We plot the mean benefits and burdens in Figure 1. Our statistical tests, and this plot, appear to show that benefits are more predictive of whether an app is used or abandoned than burdens.

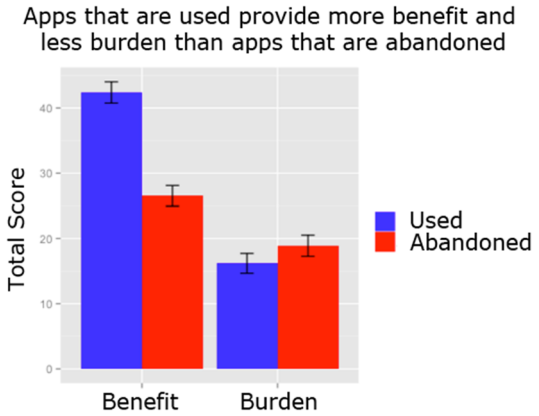

Figure 1. The mean total benefits and total burdens of abandoned and used mobile apps (both scales out of 80).

We then looked at which constructs within the benefit and burden scales were most predictive of whether an

app was used or abandoned. Subscale scores were calculated to be the average rating across all items in the construct, so the range of possible scores for each subscale was 0 to 4. The differences in used and abandoned apps by subscale are presented in Figure 2. The subscales that were predictive of whether an app was used or abandoned were the benefit subscales measuring whether an app was enjoyable, useful, usable, and social, as well as the burden subscales for whether an app was difficult to use or posed a burden on privacy.

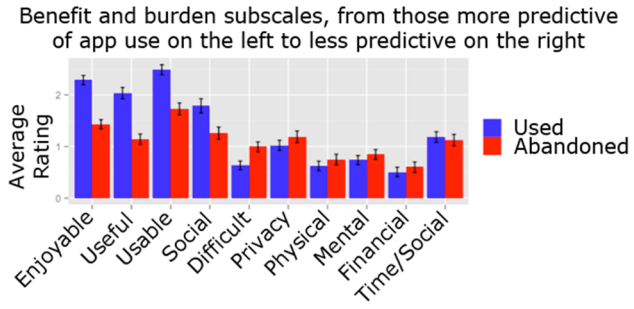

Figure 2. The mean benefits and burdens of used and abandoned mobile apps, by subscale construct.

Next, we considered the relationship between benefit and burden. We looked at patterns across all 694 pairs of benefit and burden ratings (Figure 3). Had our benefit scale simply been measuring the opposite construct from the burden scale, we'd expect to see a perfect negative correlation between total benefit and total burden. However, we see that this is not the case. In fact, there is a positive correlation ($r=.29$, $p<.001$) between total benefit and total burden, such that as benefit increases, burden likewise increases. This makes intuitive sense - as an app expands its features, it becomes more and more difficult to present those features well. As a general rule, the higher the ceiling, the higher the threshold. We also see a clear gap in benefit between used apps (blue) and abandoned app (red). At every level of total burden, used apps have higher total benefits than abandoned apps.

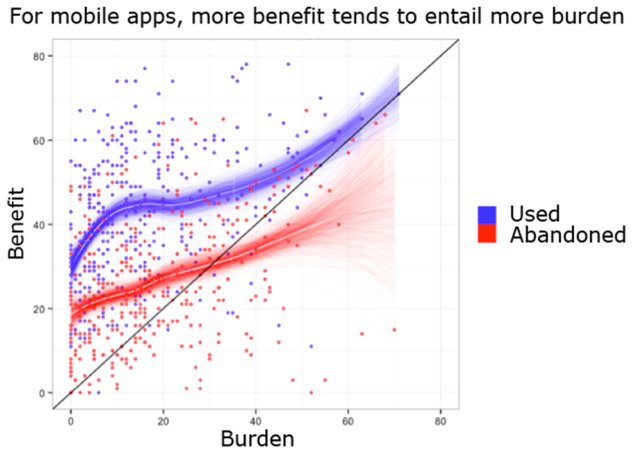

Figure 3. Total burdens plotted against total benefits.

We then looked at correlations by benefit and burden construct. All benefit subscales were highly positively correlated with one another, as were all burden subscales. Correlations between most benefit subscales and burden subscales were also positively correlated, to a lesser extent. Usability was not correlated with most of the burden subscales; plotting these relationships indicated that apps that were rated as usable, no matter how usable, were rated as having low burden across the board.

Finally, we created a predictive model of app usage. We entered all of the subscale constructs that were individually predictive of app usage (all benefit subscales, and the difficulty and privacy burden subscales) and used a backward and forward/backward stepwise regression to determine which constructs should remain in the model. Both stepwise regressions converged on the same model (Table 3).

	B	SE(B)	z	p
(Intercept)	-2.08	0.23	-8.91	<.001
Benefit: Useful	0.67	0.11	6.04	<.001
Benefit: Enjoyable	0.99	0.13	7.41	<.001
Burden: Difficult	-1.00	0.12	-8.17	<.001

Table 3. A logistic regression on whether an app is currently used or has been abandoned.

Our model fits significantly better than an empty model ($X^2(3) = 255.76$, $p<.001$). We also fit the overall model using McFadden's pseudo r-squared (p^2), and found $p^2=0.27$, where McFadden suggested p^2 values of between 0.2 and 0.4 indicate a very good model fit [11]. Both of these measures indicate that our model is predictive of app usage.

Interpreting the predictor variables entered into the model, we see that for every 1-unit change in usefulness, the log odds of using the app increase by 0.66. For every 1-unit change in enjoyment, the log odds of using the app increase by 0.99. And, for every 1-unit change in difficulty, the log odds of using the app changes by -1.00. This indicates that the more beneficial an app is in terms of usefulness and enjoyment, and the less burdensome an app is in terms of difficulty of use, the more likely it is that the app is used rather than abandoned.

Though this result may seem contradictory to our earlier finding that benefits and burdens are positively correlated, it is possible to reconcile these results. While our correlation results show that burdens tend to increase along with benefits of a mobile app, our predictive model shows that those apps that manage to maximize benefits and minimize burdens (in spite of the positive correlation), are the apps that are more likely to be used.

Limitations and Future Directions

Our results verify that used and abandoned apps differ in average benefits and burdens; however, we acknowledge some limitations to our research. First, while our survey is able to capture a meaningful component of the benefit of mobile apps, it is only a 20-item survey and simply does not capture the full complexity of benefit.

Furthermore, even if the User Benefit and User Burden Scales perfectly captured the constructs, we acknowledge that other reasons for app usage exist outside of the benefits and burdens of the app itself; for example, one would also have to consider what other competitors exist in the app market. Or, for social apps for instance, the benefits of the app may not be inherent to the design of the app itself, rather it would matter whether the app had gained sufficient adopters to be useful. What we set out to do was to consider some important attributes of apps applicable across many types of apps and users. However, input from mobile app designers, as well as insight into their design and development process, may inform new

changes to make the tool more usable and useful for them.

In addition, our findings are based on 347 responses on over 300 distinct mobile apps, but we expect that results could differ with a larger sample of people and apps. Also, although the user population on mTurk is relatively diverse, it is not representative of all populations – for example, we observed a majority of international respondents. To claim applicability of our findings in specific user populations, we would need to conduct tests in more clearly specified participant samples based on features such as age, country of citizenship, and technology expertise.

In future work, rather than measuring users' ratings of the benefits and burdens of mobile apps after an arbitrary period of use, it could be informative to measure perceived benefits just after the users download mobile apps or after a specified period of use. Also, because the User Burden Scale was designed and validated for computing systems more generally, it could be worthwhile to explore an adaptation of the scale that is applied to mobile apps only. Some of the burdens, especially in the physical burden category, were less applicable to mobile apps.

Finally, in this paper, we prioritized interpretability of the model over accuracy in prediction, so we used logistic regression to create a predictive model. Alternately, we could have used machine learning to create a predictive model, using cross validation to train and test the weights on the predictors. It would be interesting to know with what percent accuracy a model could predict new instances of app usage.

Conclusion

In this paper, we described the design and validation of a four-construct user benefit scale for assessing user benefits in mobile apps: if an app is 1) useful and informational, 2) enjoyable and enables pursuit of interests, 3) enables social interaction, and 4) has good usability and visual/interaction design.

Combining our User Benefit Scale and Suh et al.'s [15] User Burden Scale, we found that benefits are more predictive of whether an app is used or abandoned than burdens. Our results also indicate that the most predictive subscales were the benefit subscales measuring whether an app was enjoyable, useful, usable, and social, as well as the burden subscales for whether an app was difficult to use or posed a privacy burden. Though apps that are more beneficial tend to also be more burdensome, our predictive model shows that the more beneficial an app is in terms of usefulness and enjoyment, and the less burdensome an app is in terms of difficulty of use, the more likely it is that the app is used rather than abandoned.

These results can help mobile app designers and developers better understand trade-offs in design and functionality, thus allowing them to appropriately prioritize what matters most to their users. We propose that mobile app designers should focus on maximizing benefits to the user, and maintain keeping burdens low as an important but secondary concern.

Acknowledgements

We thank Rob Semmens, Kesler Tanner, and the Stanford Institutional Review Board (IRB) for their assistance and guidance in this research.

References

[1] Babbar, S., Behara, R., White, E. (2002). Mapping Product Usability. International Journal of Operations and Production Management, 22(10): 1071-1089, 2002.

[2] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 194: 4-7.

[3] Chou, C., Chiu, C., Ho, C., Lee, J. (2013). Understanding Mobile Apps Continuance Usage Behavior. Proceedings of Pacific Asia Conference on Information Systems 2013.

[4] Davis, F.D., Bagozzi, R.P. & Warshaw, P.R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. Journal of Applied Social Psychology, 22(14), 1111-1132.

[5] Davis, F. (1993). User Acceptance of Information Technology: System Characteristics, User Perceptions and Behavioral Impacts. International Journal Man-Machine Studies, 38(3): 475-487, 2003.

[6] Eyal, N. (2014). Hooked: How to Build Habit-Forming Products. Canada: Penguin.

[7] Fogg, B.J., Cuellar, G., and Danielson, D. (2002). Motivating, influencing, and persuading users. In J. Jacko and A. Sears (eds.), The human-computer interaction handbook.

[8] Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70.

[9] Kim, Y., Kim, D., Wachter, K. (2013). A study of mobile user engagement (MoEN): Engagement motivations, perceived values, satisfaction, and continued engagement intention. Decision Support Systems, 56, 361-370.

[10] Lin, C.S., Wu, S. and Tsai, R.J. (2005). Integrating perceived playfulness into expectation confirmation model for web portal context. Information & Management, 42(5), 683-693.

[11] Louviere, J., Hensher, D., & Swait, J. (2000). Stated choice methods: analysis and applications. Cambridge University Press.

[12] Moon, J. & Kim, Y. (2001). Extending the TAM for a World -Wide-Web context. Information & Management, 38(4), 217-230.

[13] Park, J., Snell, W., Ha, S. & Chung, T.L. (2011). Consumers' post-adoption of m-services: interest in future m-services based on consumer evaluations of current m-services. Journal of Electronic Commerce Research, 12(3), 165-175.

[14] Steel, C. (2012). Fitting learning into life: Language students' perspectives on benefits of using mobile apps. Ascilite Conference 2012.

[15] Suh, H., Shahriaree, N., Hekler, E., & Kientz, J. (2016). Developing and Validating the User Burden Scale: A Tool for Assessing User Burden in Computing Systems. Proc. CHI '16, 2016.

[16] Yarosh, S., Markopoulos, P., Abowd, G. (2014). Towards a Questionnaire for Measuring Affective Benefits and Costs of Communication Technologies. CSCW '14, 2014.