Designing and Testing User-Centric Systems with both User Experience and Design Science Research Principles

Emergent Research Forum papers

Soussan Djamasbi
Worcester Polytechnic Institute
djamasbi@wpi.edu

E. Vance Wilson
Worcester Polytechnic Institute
vwilson@wpi.edu

Diane M. Strong
Worcester Polytechnic Institute
dstrong@wpi.edu

Carolina Ruiz
Worcester Polytechnic Institute
cruiz@wpi.edu

Abstract

The User Experience (UX) and Design Science Research (DSR) paradigms have much in common; they both contribute to Information Systems (IS) research by providing guidelines for designing successful Information Technology (IT) systems. We are working toward a research paradigm that combines the best elements of DSR and UX for designing user-centric IT systems with an outstanding user experience. To achieve this goal, we are jointly applying these two paradigms to develop an IT artifact (a sleep app). We will examine what we have learned from applying DSR and the UX principles and explore how these two paradigms individually and jointly can strengthen the design and development process for user-centric systems. Our initial results indicate that jointly using these two paradigms can strengthen the design and development process for user-centric systems and can be of great value to theory and practice.

Keywords

Design Science, User Experience, User-Centric, Sleep Health, IT artifact

Introduction

Both the User Experience (UX) research paradigm and Design Science Research (DSR) paradigm provide value to Information Systems (IS) researchers as they design and develop Information Technology (IT) systems for use by a variety of people. These two paradigms have much in common, but each has weaknesses that could benefit from the strengths of the other. We argue that UX research could benefit from the formal structure of DSR (e.g., the seven design principles) to better communicate its findings and contribution to theory and practice. We also argue that DSR could benefit from UX principles that provide specific guidelines, practices, and metrics for measuring the development progress of IT systems designed for a variety of users. Thus, we are working toward a research paradigm that combines the best elements of DSR and UX for designing and developing IT systems for which an excellent user experience is critical. We refer to such IT systems as user-centric systems.

To achieve this goal, we are jointly applying these two paradigms as we conduct a research project addressing a specific problem, sleep health, by developing an IT artifact, an Android app. At each stage of design, development, testing, and user studies, we are examining what we have learned through applying the lens of DSR and the UX to explore how these two paradigms individually and jointly can strengthen the design and development process for user-centric systems. In this paper, we report our initial research results, specifically the results gained from initial user testing of the app prototype as guided by UX and DSR principles. Longer term, we also expect to extend related behavioral theories as we learn more about users and their behavior through their app usage.

Background

User Experience (UX) refers to the subjective experience of an individual in his/her encounters with a technology. UX principles specifically acknowledge that different people are likely to have different experiences when encountering the same technology and that their experiences may differ from what designers of that technology had planned (Hassenzahl, 2003; Hassenzahl et. al., 2015). Therefore evaluating users' experiences with a technology is an important and critical step in designing systems that will be successful (Djamasbi, 2014; Albert & Tullis, 2013). Such UX evaluations help ensure that intended design aspects are communicated adequately to users and are well received by them. This is an iterative process, with the results obtained from UX testing serving as a guide for the next design cycle.

The DSR paradigm has served as an excellent framework for guiding information systems research that involves developing IT artifacts (Gregor et. al., 2013; Hevner et. al., 2004). It covers a broad range of information systems, broader than the user-centric focus of UX design principles.

While they have a somewhat different focus, the UX and DSR design paradigms have a great deal of overlap. Similar to DSR, UX research requires an artifact. In UX research, understanding and solving a design problem from a user's point of view can only be attained through developing and iterative testing of prototypes or artifacts. Furthermore, the technological solutions must go far beyond satisfying user expectations to provide the required positive and competitive users' experiences (Hassenzahl, 2003; Hassenzahl et. al., 2015). In doing so, the development of competitive UX artifacts by themselves often contribute to knowledge by either solving unsolved problems or solving known problems more effectively and/or efficiently, which are goals advocated by the DSR paradigm. Additionally, similar to the DSR process, the UX design process enables a search for novel solutions in a desired problem space.

These characteristics of the UX research paradigm are identical to core DSR guidelines proposed by Hevner et al. (2004). What a UX design process lacks can be found in DSR in its guidelines for communicating contributions to research in a more comprehensive and systematic manner. A DSR process, when applied to user-centered systems, can benefit from advances in UX theories that extend beyond considering only utilitarian needs of consumers but also take into consideration the psychological hedonic needs that often are a stronger driver of behavior (Hassenzahl, 2003; Hassenzahl et. al., 2015).

App Design

Our study's design problem is in the health domain, in particular sleep health. Sleep deprivation is a common unmet public health problem with many adverse effects on people, including accidental injury from driving while sleepy, poor performance, and difficulty in remembering or concentrating ("Insufficient Sleep Is a Public Health Problem," n.d.). We apply persuasive behavioral theories to motivate the forming, changing, and reinforcing of users' attitudes and behaviors (Fogg 1999). Through our research, we intend to refine and extend these theories.

The initial version of the app is a basic prototype (Figure 1), which provides functionality for users to (1) set a sleep goal (I like to sleep 8 hours per night), (2) track their sleep time using a manual toggle button (I am going to sleep, I am waking up), and (3) view their sleep history. The app also includes (4) an alarm clock that can be set to play users' favorite music when going to sleep and/or when waking up.

App Testing Methods Informed by UX and DSR Principles

Both DSR and UX call for iterative testing. While DSR does not provide instructions for testing, the practice of UX design provides detailed guidelines for various methods of conducting user studies based on the goals of the research. For example, Albert and Tullis (2013), grounded in empirical evidence, recommend formative studies (i.e., frequent tests with small number of users) at the early stages of development. Formative studies, particularly the initial sets with only a handful of user (4 to 6 users), typically involve qualitative research because they yield rich data sets (Albert and Tullis 2013).

Using these guidelines, we conducted two formative user studies with a total of 10 participants (n1=4 and n2=6), during which we observed users completing several core tasks with the app (Figure 2). Performance on core tasks provides an important first step in formative studies (Albert and Tullis 2013).
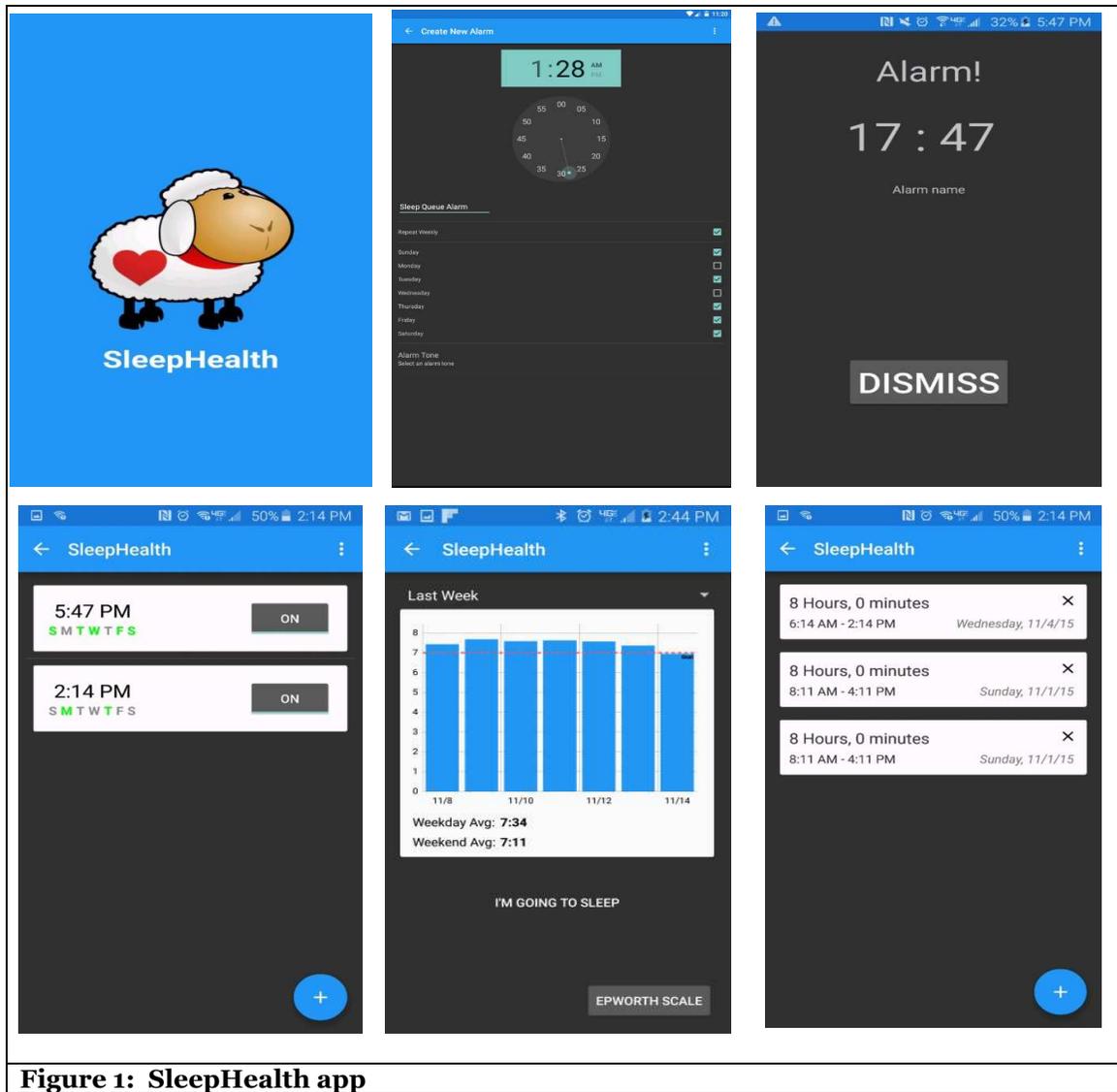
Figure 1: SleepHealth app

1. Download and install the app off of "server name"
2. Proceed through the initial setup procedure of the app
3. Set a time for an alarm
4. Set music for an alarm
5. View the graph of sleep data
6. Take the Epworth Sleepiness Scale Survey (Johns, 1991)
7. Opt out of notifications

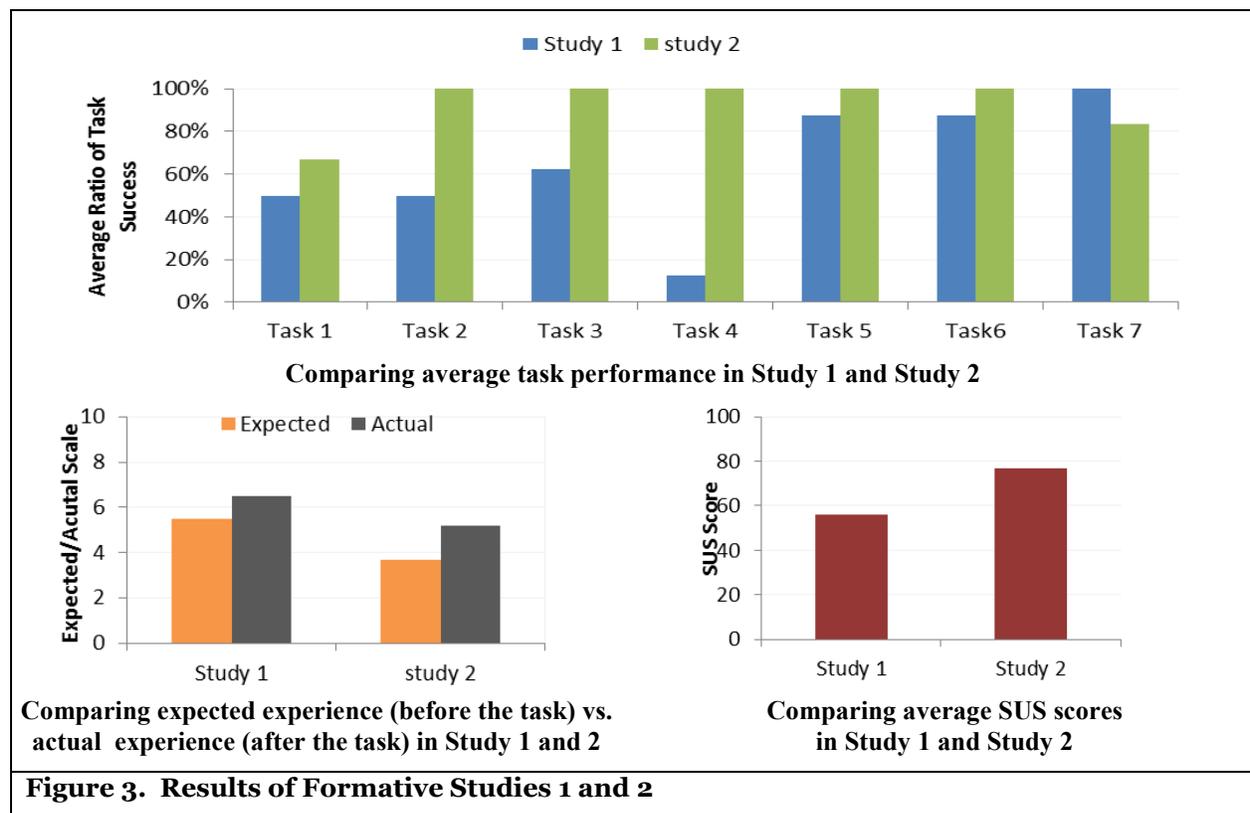**Figure 2. Core Tasks in Studies 1 and 2**

In each user study, before completing these tasks, we asked users to rate on a 1-10 scale how easy/difficult they expect it would be to complete each of these tasks using the app. After completing all the tasks, we asked users to rate how easy/difficult it actually was to complete each of these tasks using the app on the same 1-10 scale. This method, developed by Albert and Dixon (2003), is used in industry research to

prioritize development resources because it gauges a user's perception of a technology against his/her expectation of that technology. For example, if a user gives a low score to actual experience of the technology but indicate that he/she expects that technology to be easy to use, there will be an immediate need to make design improvements. In contrast, when the user expects the technology to be hard to use and gives a low score for actual experience, design changes can be deferred until more important issues are resolved. In cases where users expect the technology to be hard to use but find it easy to use, the technology provides an experience that exceeds user expectations and designers must make sure that the design is kept intact in the next iterations.

We employed the widely-used System Usability Scale (SUS) to track our design process (Albert and Tullis, 2013). SUS is a 10-item survey, designed by John Brooke (1996). From the survey item results, a single SUS score between 0-100 is calculated. SUS scores with a value below 50 indicate poor design; between 50 and 70, an acceptable design but with some usability problems; between 70 and 85, a good design; and above 85, an excellent design (Bangor et al. 2008, 2009).

Results

The results of our user studies show that we were able to learn from the first formative study; that is, the results of the second study, done after improving the app based on the results of the first study, were better than the results of the first study (Figure 3). For example, the revised app provided clearer instructions for signing up – making it easier to enter information during the sign up process, for setting the alarm and for selecting music for the alarm. We also removed some bugs encountered when downloading the app. As shown in Figure 3, these improvements increased task performance for six of the seven tasks. SUS scores also improved from a poor range for Study 1 (SUS= 56.25, below 70) to an acceptable range for Study 2 (SUS= 76.66, above 70). Despite these improvements, users' actual experience with the app remained harder than they expected in Study 2. These results indicate that measuring user expectations, in addition to traditional SUS usability scores and objective task performance measures, can provide a more comprehensive picture of user experience, and as such adds a new dimension to novelty measurement as required by the DSR process.

Comparing average task performance in Study 1 and Study 2

Comparing expected experience (before the task) vs. actual experience (after the task) in Study 1 and 2

Comparing average SUS scores in Study 1 and Study 2

Figure 3. Results of Formative Studies 1 and 2

Discussion
While we do not yet have the behavioral data needed to explore behavioral change theories, our results suggest that the DSR guideline in regard to novelty could be refined for user-centered IS studies to include user perception of novelty in terms of exceeding expectations. Similarly, UX principles can be extended by exploring various ways in which novelty, as a relationship between user perception and user expectation, can be captured and assessed. Research provides ample evidence that user perceptions can have a significant impact on driving behaviors (e.g., Davis 1989) and as such are crucial in designing competitive systems (Djamasbi 2014). Our results show that while we were able to improve task performance we did not create an experience that exceeded users' expectation (Figure 3). These results provide evidence for the importance of including both subjective and objective performance measures in a DSR study that involves IS system design.

Conclusion
Currently, we are in the process of using the results of the second formative study to improve the app. At this stage of our research, our artifact is not yet ready for testing behavioral change theories. We expect several more design, development, and formative study iterations before the app is ready for testing behavioral change theories. We expect that, as we continue our iterative development process, we will learn more about the connection between DSR and UX research principles and will be able to compare the combination of these two research paradigms for assessing progress in earlier and later stages of development.

ACKNOWLEDGEMENTS

REFERENCES
Albert, W., & Dixon, E. (2003). Is this what you expected? The use of expectation measures in usability testing. Proceedings of Usability Professionals Association 2003 Conference, Scottsdale, AZ.

Albert, W., & Tullis, T. (2013). Measuring the user experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann.

Bangor, A., Kortum, P., & Miller, J. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574–594. doi:10.1080/10447310802205776

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS Scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114–123.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry.* London: Taylor and

Djamasbi, S. (2014) Eye Tracking and Web Experience. AIS Transactions on Human-Computer Interaction (6) 2, pp. 37-54.

Fogg, B. J. (1999) Persuasive technologies: Introduction. Comm. of the ACM, 42(5), 26–29.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37(2), 337-355.

Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), Funology: From Usability to Enjoyment (pp. 31-41). Dordrecht: Kluwer.

Hassenzahl, M., Wiklund-Engblom, A., Bengs, A., Hagglund, S., & Diefenbach, S. (2015). Experience oriented and product-oriented evaluation: psychological need fulfillment, positive affect, and product perception. International Journal of Human-Computer Interaction, 31(8), 530-544.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75-105.

Insufficient Sleep Is a Public Health Problem (n.d.). Retrieved from http://www.cdc.gov/features/dssleep/

Johns MW (1991). A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. Sleep; 14(6):540-5.