Improving Usability, Safety and Patient Outcomes with Health Information Technology
F. Lau et al. (Eds.)

doi:10.3233/978-1-61499-951-5-375

Thought Spot: Embedding Usability Testing into the Development Cycle

Shehab SENNAH[a,*], Jenny SHI[b,*], Elisa HOLLENBERG[b], Andrew JOHNSON[b], Genevieve FERGUSON[b], Alexxa ABI-JAOUDÉ[b] and David WILJER[a,b,c,1]
[a] *University of Toronto*
[b] *Centre for Addiction and Mental Health*
[c] *University Health Network*

Abstract. Usability testing is a vital component in the development of any digital innovation. Thought Spot, a mental health and wellness mobile application designed for and by transition-aged youth, underwent three distinct phases of usability testing (lab testing, field testing and heuristic evaluations). Testing highlighted that participants generally had a positive experience with the platform. Although some app functions were initially difficult for users, positive trends in learnability were observed. The key lesson learned from this process is the need for iterative testing timelines, concurrent with app development.

Keywords. usability testing, mHealth, student mental health

1. Introduction

Mental illness and substance use disorders are the leading cause of disability and total disease burden for youth in high-income countries [1]. In fact, about 20% of Canadian youth aged 15 to 24 have reported experiencing symptoms of mental illness and 8% have reported substance use concerns [2]. Unfortunately, only 36% receive the mental health or addictions support they need, leaving transition-aged youth at risk of long-term illness, unemployment, youth justice involvement or self-medication [3]. In addition, system access and navigation barriers along with stigma and confidentiality concerns make help-seeking difficult for this age group [4].

Thought Spot is a co-created and iteratively designed online and mobile platform to help reduce access-to-service barriers and confidentiality concerns for post-secondary youth [5, 6]. Its core function is to allow users to find and share health and wellness resource (spots) while providing a private space for them to track their thoughts and moods (thoughts).Using an interactive and crowdsourced map, users are able to search for relevant resources through geo-location and search filters/tags. Given that youth use the internet regularly [7] and often seek mental health information via web-based resources [8], Thought Spot has the potential to support this population in identifying useful services while promoting their mental health. The platform is currently undergoing

*Co-first Authors

[1] Corresponding and lead author, David Wiljer, Executive Director, Education Technology Innovation, University Health Network, Toronto General Hospital, R. Fraser Elliot Building, 3-411, 190 Elizabeth St., Toronto, ON M5G 2C4; E-mail: David.Wiljer@uhn.ca.

an RCT to evaluate its impact on self-efficacy for mental health help seeking and health literacy amongst transition-aged youth [9].

Usability testing is widely used in the design of digital health interventions and more recently in the development of m-health applications [10]. App usability is one of the main factors determining app success, where failure to meet user demands can decrease effectiveness, efficiency, satisfaction and task performance [11]. End user feedback is critical for understanding what works, what doesn't work and where there are technical or user-interface gaps that might affect app performance or satisfaction.

The purpose of the present paper is to discuss the usability testing of Thought Spot, a multi-platform application, conducted under time- and resource-limited conditions.

2. Methods

Thought Spot was developed through a participatory design research process involving end-user consultations throughout its conceptualization and design. Transition aged youth played an active role in developing the content, the structure/functioning and the look and feel of the app through a range of co-creation methods including: a youth-led development team (Thought Spot Student Group), crowdsourcing/data workshops, co-design workshops and a hack-a-thon [5]. As a part of the iterative and participatory evaluation process, Thought Spot underwent distinctive phases of usability testing (lab, field and heuristic evaluations) over the course of three weeks. A total of 15 post-secondary students from the Greater Toronto Area (GTA) and two technical experts were recruited to complete usability testing across three platforms (desktop, iOS and Android). Following the REB protocol for Phase 1 of the Thought Spot study, numerous methods were used to recruit students for usability testing from three GTA schools (George Brown College, Ryerson University and the University of Toronto). Information for usability testing session was promoted through social media accounts (Facebook, Instagram and Twitter) of various academic departments and student organizations. Members of the research team also promoted usability testing sessions directly to students on campus.

2.1. Usability Design and Procedures

2.1.1. Lab Testing

Each platform (desktop, iOS and Android) was assigned five testers. Over the course of 90 minutes, each participant completed a demographic and technology use questionnaire, a series of 12 "think aloud" usability tasks, the Single Ease Question (SEQ) questionnaire [12], a Post-Study System Usability Questionnaire (PSSUQ) and a debrief interview [12]. Sessions were video recorded and detailed observations and notes were taken during the session by a second facilitator.

2.1.2. Field Testing

At the end of each lab testing session, participants were offered the opportunity to participate in a field test. Participants continued to use the same platform they were originally assigned during lab testing (five desktop, four iOS and three Android), and were asked to complete a list of tasks over the course of one week from their lab test date. At the end of field testing, participants completed a semi-structured interview (in-person or phone) and the PSSUQ questionnaire.

2.1.3. Heuristic Testing

Experts received detailed descriptions of the app and were asked to evaluate it using Nielsen's 10 usability heuristic principles and severity scale [13].

3. Results and Analysis

A total of 15 end user participants completed lab testing (10 females and 5 males), 12 completed field testing and one external subject expert provided incomplete heuristic results. The mean age was 20.9 years (SD = 2.66). In terms of technology use, 93.3% of participants rated their use as highly connected and 86.7% rated their comfort with technology as high.

The SEQ provides insight into user satisfaction and usability issues by asking users to rate the difficulty of a task [13]. On a 7-point Likert scale, with higher scores indicating a lower difficulty, the average score across all tasks was 5.2 (SD = 0.82). To determine which tasks were most difficult relative to each participant, individual SEQ scores were analyzed for their differences from a participant's average score across all tasks (Figure 1). The core features of 'find spot' and 'add spot' represented the most difficult tasks when completed for the first time, scoring 1.1 (SD = 0.6) and 1.2 (SD = 1.4) below average respectively. The third core feature 'add thought' scored 0.06 (SD = 0.5) below average when completed for the first time. Overall trends for SEQ scores suggest 'learnability' for difficult tasks (e.g., find spot, add spot and add a thought), with participants scoring higher during subsequent attempts for these tasks.

Grouping the results by platform-type provided some insight into the high variability of the scores. Desktop users had a significantly more difficult initial experience with the 'add spot' feature compared to iOS and Android. Further investigation during the debrief interviews highlighted that the workflow required to complete the task was unexpected for desktop users. Participants noted that they were using the search function, instead of the intended 'add spot' button, due to the similarity of the Thought Spot user interface with other map-based apps.

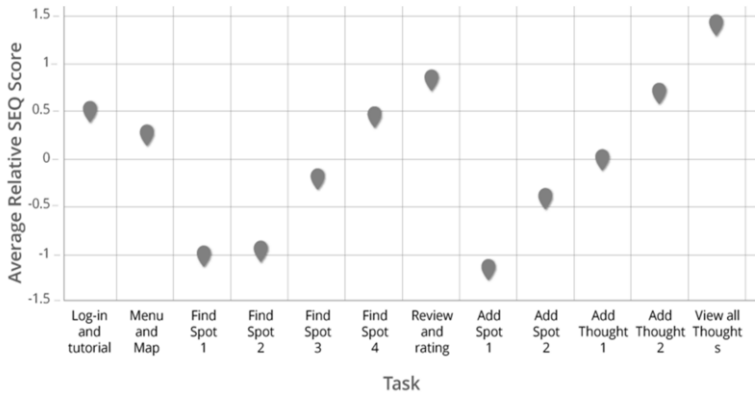

Figure 1. Average relative SEQ scores for post-task experiences are determined by taking the difference between each participant's individual scores from the average score given on all tasks.

Usability scores for the PSSUQ are presented in Figure 2 for both lab and field testing. Specifically designed for scenario-based usability studies, the PSSUQ provides insight into system usefulness, information quality and interface quality [12]. The lowest scores were given to the PSSUQ statement "The system gave error messages that clearly told me how to fix problems" with an average of 2.2 (SD = 1.8) on a 7-point Likert-scale after lab testing, and 2.5 (SD = 1.4) after field-testing." These scores highlighted a need for built-in error messages to streamline user experience. Lower field testing scores compared with lab testing scores suggest decreased satisfaction and lower overall acceptability of the platform after extended usage. However, it is possible that lab testing scores were inflated due to observer bias, and functional and user interface issues may have hindered satisfaction during field testing.

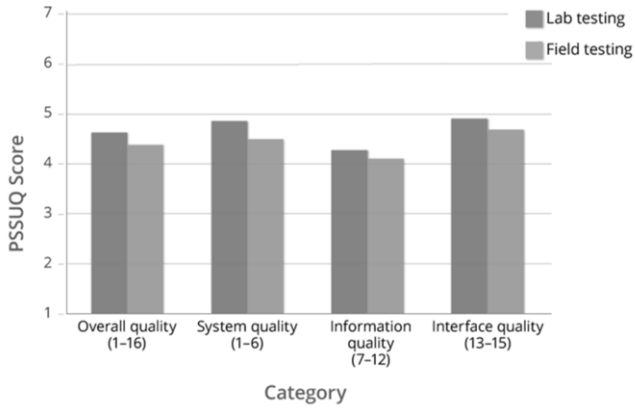

Figure 2. Average PSSUQ scores for lab and field testing.

Despite the above issues, participants were generally pleased with Thought Spot's interface quality. The only cosmetic issue concerned the spacing and overlay of icons on the timeline menu.

4. Discussion

Usability testing is a key area of app development because usability can seriously affect user satisfaction and adoption rates [11]. During our experience of this key step, we encountered unexpected constraints. Unforeseen complications during app development pushed usability testing past its intended timelines and affected planned availabilities of content-expert project staff for leading the testing process. The result was a decision to move forward with testing some features that were considered functionally complete while other features had known unresolved functionality issues. In addition, based on the vital feedback from initial usability testing, Thought Spot underwent a documented functional update two-thirds of the way into the process. The update was a necessary step, despite the known limitations, given the importance of seeking usability feedback on key app functionalities. Notably, caution is necessary when examining usability results with

or without changes to the product during testing. For instance, it is unclear if the differences between field and lab testing scores in Figure 2 are due to functionality issues hindering user satisfaction during field testing or observer bias inflating scores during lab testing.

The real-world challenges with the usability testing of Thought Spot resulted in interim rather than final usability testing of the app during its development. However, the interim testing afforded a greater opportunity for iterative design and made it possible for developmental changes. Exploring features that were confusing to users or that users interpreted differently than intended enabled developers to revise the design. In fact, knowing about functionality issues during testing gave developers greater flexibility to change course based on user feedback. For example, insight into the demanding process of adding a spot, a key feature of the crowdsourcing component of Thought Spot, encouraged the addition of autocomplete features. Moreover, interim testing highlighted the need for further development of key features such as the search function, which helped focus resources to optimize data organization and hygiene.

Given that Thought Spot is a multi-platform app, interim usability testing also allowed us to compare desktop and mobile platforms and to understand how to navigate trade-offs between these formats. For example, we found that the desktop platform had much greater stability compared with the mobile platforms, while the mobile platforms better use the application's portability and geolocation capabilities. The difference in stability between desktop and mobile platforms may have been a result of the development method where a single version of Thought Spot was created as a hybrid-app, making use of a"wrapper" for different operating systems, rather than creating multiple native apps. Along with stability, usability may have been affected by this choice, as users of different operating systems have different expectations in application workflow. Although developing a hybrid-app is less resource intensive, native apps offer a better user experience [14].

Several months of additional app development and testing (e.g., informal internal and external testing, user acceptance testing and stability testing) were necessary based on the usability testing findings before the app could be launched for a randomized control trial (RCT) to evaluate the efficacy of Thought Spot [9].

5. Next Steps and Recommendations

Consistent with the iterative approach to Thought Spot's development, there is an intention to continue evaluating the usability of the app as an endpoint to the RCT. An adaptation of the USE questionnaire [15] and open-ended usability questions are a part of the final survey study design. In addition, a purposeful sample of participants will be selected for semi-structured interviews to gain a deeper understanding of app usability for future iterations of Thought Spot.

The usability testing of Thought Spot demonstrated real-world challenges of resource constraints that arose due to unforeseen complications during app development. While testing during the development cycle may not be the best use of time as known errors and bugs are often reported, our experience resulted in valuable insight into the usability and design of the app, more opportunities for end-user engagement, the ability to utilize content-expert project staff resources under time constraints and provide developers with more opportunities for essential functional/design changes to the platform.

Based on Thought Spot's interim testing, we recommend in circumstances of limited resourcing to focus on testing features that are functionally complete but do not avoid testing features still in development. We also encourage concurrent app development along with key functional or usability updates during testing. Finally, detailed version logs of these updates should be maintained to overcome the instability of the in-development environment. The usability process was an essential component of the development of Thought Spot, and the continued partnership with end users will drive research and implementation of this application to improve student mental health.

Acknowledgments

We thank our funders: Ontario Ministry of Training, Colleges and Universities (Health Innovation Fund), CIHR (eHealth Innovations Partnership Program), and CAMH's Slaight Family Centre for Youth in Transition. We also thank the Thought Spot Research Team and all of the students who have contributed to the project. Special thanks to Annie Hart for UX design support and to Hema Zbogar for editing this manuscript.

References

[1] H.E. Erskine, T.E. Moffitt, W.E. Copeland et al., A heavy burden on young minds: the global burden of Mental and substance use disorders in children and youth, *Psychological Medicine* **45** (2015), 1551-1563.

[2] T. Peter, L.W. Roberts, J. Dengate, Flourishing in life: An empirical test of the dual continua model of mental health and mental illness among Canadian university students, *International Journal of Mental Health Promotion* **13** (2011), 13–22.

[3] J. M. Carver, S. Capelli, W. Davidson, M.-A. Caldwell, M. Bélair, M. Vloet, *Taking the Next Step Forward: Building a Responsive Mental Health and Addictions System for Emerging Adults*, Mental Health Commission of Canada, Ottawa, 2015. Available from: https://www.mentalhealthcommission.ca/sites/default/files/Taking%252520the%252520Next%252520Step%252520Forward_0.pdf

[4] S. Davidson, M. Capelli, M.A. Vloet, *We've Got Growing Up to Do: Transitioning Youth from Child and Adolescent Mental Health Services to Adult Mental Health Services*, Ontario Centre of Excellence for Child and Youth Mental Health, Ottawa, 2011. Available from: http://www.excellenceforchildandyouth.ca/sites/default/files/policy_growing_up_to_do.pdf

[5] N. VanHeerwaarden, G. Ferguson, A. Abi-Jaoude et al., The optimization of an eHealth solution (Thought Spot) with transition-aged youth in postsecondary settings: Participatory design research, *Journal of Medical Internet Research* **20**: 3 (2018), e79.

[6] D. Wiljer, A. Johnson, E. McDiarmid et al., Thought Spot: Co-creating mental health solutions with postsecondary students, *Studies in Health Technology and Informatics* **234** (2017), 370–375.

[7] J.M. Burns, T.A. Davenport, L.A. Durkin, G.M. Luscombe, I.B. Hickie, The internet as a setting for mental health service utilisation by young people, *Medical Journal of Australia* **192** (2010), S22–6.

[8] F.M. Wetterlin, M.Y. Mar, E.K. Neilson, G.R. Werker, M. Krausz, eMental health experiences and expectations: A survey of youths' Web-based resource preferences in Canada, *Journal of Medical Internet Research* **16**: 12 (2014), e293.

[9] D.Wiljer, A. Albi-Jaoude, A. Johnson et al., Enhancing self-efficacy for help-seeking among transition-aged youth in postsecondary settings with mental health and/or substance use concerns, using crowdsourced online and mobile technologies: the Thought Spot protocol, *Journal of Medical Internet Research* **5: 4** (2015), e201.

[10] D. Zhang, B. Adipat, Challenges, methodologies, and issues in the usability testing of mobile application, *International Journal of Human–Computer Interaction* **18** (2005), 293–308.

[11] B.C. Zapata, J.L. Fernández-Alemán, A. Idri, A. Toval, Empirical studies on usability of mHealth apps: A systematic literature review, *Journal of Medical Systems* **39**: 2 (2015), 1.

[12] J. Sauro, J.R. Lewis, *Quantifying the User Experience: Practice Statistics for User Research*, Morgan Kaufmann, Cambridge, MA, 2012.

[13] H. Monkamn, J. Griffith, A.W. Kushniruk, Evidence-based heuristics for evaluating demands on eHealth literacy and usability in a mobile consumer health application, *Studies in Health Technology and Informatics* **216** (2015), 358–362.
[14] R. Jabangwe, H. Edison, A.N. Duc, Software engineering process models for mobile app development: A systematic literature review. *Journal of Systems and Software* **145** (2018), 98-111.
[15] A.M. Lund, Measuring usability with the USE questionnaire, *Usability Interface* **8**: 2 (2001), 3–6.