Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1111–1118
https://doi.org/10.54941/ahfe1008341

An Exploratory Study of Weekly Satisfaction Tracking for Long-Term UX Evaluation

Toshihisa Doi

Osaka Metropolitan University, Osaka, 5588585, Japan

ABSTRACT

This study explored what kinds of data can be obtained when satisfaction is tracked over a long period using weekly reports. Ten participants used a Fitbit Inspire 3 for eight weeks, and weekly overall satisfaction and retrospectively reported daily satisfaction ratings were collected. The data were analyzed using a linear mixed-effects model. No clear uniform temporal trend in satisfaction was observed across the eight weeks. Approximately 30% of the variability in weekly satisfaction was attributable to differences between users. Among the summary measures derived from retrospectively reported daily satisfaction, Mean and Min were significantly associated with weekly overall satisfaction, whereas Max and End were not. These findings suggest that weekly reporting may provide a relatively simple approach for tracking individual satisfaction trajectories during long-term product use. However, because daily satisfaction was retrospectively reported at the end of each week, the relationships between daily and weekly satisfaction should be interpreted cautiously. Although this was an exploratory study with a small sample and a single product, the findings provide a basis for further examining weekly reporting as an approach to long-term UX monitoring.

Keywords: User experience, UX monitoring, Long-term UX, UX evaluation, Satisfaction

INTRODUCTION

User experience (UX) develops over time through repeated interactions with products and services. Experiences before use, during initial use, and throughout continued use may contribute to users' overall evaluations. Therefore, understanding UX requires not only short-term usability evaluation but also examination of how experiences and evaluations change during long-term product use.

Several approaches have been used to investigate UX over time. The UX Curve retrospectively captures changes in experience over an extended period (Kujala et al., 2011), whereas the experience sampling method (ESM) and day reconstruction method (DRM) capture experiences at shorter temporal intervals (Csikszentmihalyi and Larson, 1987; Kahneman et al., 2004). Previous longitudinal studies have also examined changes in UX evaluations at multiple time points during product use (Kujala and Miron-Shatz, 2015; Doi et al., 2022; Doi, 2026). These studies demonstrate the importance of

Received September 8, 2026; Revised October 8, 2026; Accepted October 13, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

considering the temporal dimension of UX. However, repeatedly assessing UX over extended periods requires consideration of both temporal resolution and participant burden. Weekly reporting may offer a practical compromise, allowing repeated tracking of individual users while requiring less frequent reporting than daily or momentary assessment.

This exploratory study examined satisfaction reported weekly over eight weeks of product use. Participants reported weekly overall satisfaction together with daily satisfaction and usage episodes at the end of each week. The study examined the following three research questions:

RQ1: How does weekly overall satisfaction vary across eight weeks?
RQ2: To what extent is variability in weekly satisfaction attributable to differences between users?
RQ3: How are daily satisfaction ratings related to weekly overall satisfaction?

METHOD

Participants

Ten university students (male: 5, female: 5, mean age: 20.6 years, SD = 1.35) participated in the study. None of the participants had previous experience with the product. They were recruited via e-mail announcements at Osaka Metropolitan University, and received 5,000 yen as compensation.

The purpose and content of the experiment were explained before the experiment, and written informed consent was obtained in advance. Approval was obtained from the Ethics Committee of the Graduate School of Human Life and Ecology of Osaka Metropolitan University (approval number: 23-64).

Apparatus

The Fitbit Inspire 3 was used as the product to be investigated in this study. The product was selected because it could be used easily in everyday life, allowing repeated assessment of satisfaction and usage episodes. This product also offers services linked to smartphone applications.

Procedure

In this study, the Fitbit Inspire 3 and corresponding applications were used freely by each participant for eight weeks. The eight-week duration was chosen to capture UX over several weeks of continued product use. To allow naturalistic use, participants were not required to use the product daily and were not given instructions regarding specific functions or tasks. Participants were, however, encouraged to use the product sufficiently to engage with it during the study period.

During the eight weeks of use, the participants were asked to complete a questionnaire using MS Forms at the end of each week. The questionnaire items consisted of (1) overall satisfaction for the week (answered on a scale of –10 to 10), (2) a description of the episodes of use on each day, and (3) satisfaction for each day (answered on a scale of –10 to 10).

Measures

To examine the relationship between daily and weekly satisfaction, the maximum (Max), minimum (Min), mean (Mean), and satisfaction on the final day (End) of each week were calculated from the daily satisfaction ratings reported in the weekly questionnaire. In addition to these indicators, the week of use (Week) was also included as an independent variable.

Data Analysis

A linear mixed-effects model was estimated using restricted maximum likelihood (REML), with weekly overall satisfaction as the dependent variable. The fixed effects included Week (categorical), Max, Min, Mean, and End, while individual differences among participants were treated as random intercepts. The study comprised 80 possible observations from 10 participants over eight weeks; one missing weekly response resulted in 79 observations being included in the linear mixed-effects model. Linear mixed-effects models were estimated using the lme4 and lmerTest packages in R (ver. 4.5.2), with statistical significance assessed using the Satterthwaite approximation ($\alpha = .05$). This exploratory study aimed to examine the characteristics of data obtainable through weekly UX reporting and to inform future studies with larger samples.

RESULTS

Changes in Satisfaction Over Eight Weeks

The weekly changes in overall satisfaction for each of the ten participants are shown in Figure 1. As shown in Figure 1, although variability among participants can be observed in satisfaction trends, no clear overall pattern across participants—such as a consistent increase or decrease over the eight-week period—was observed. The transition of the mean weekly satisfaction across the ten participants is shown in Figure 2. As shown in Figure 2, the mean satisfaction remained around six throughout the period, and no large fluctuations were observed. Table 1 presents the correlation coefficients among the indicators calculated from the daily satisfaction ratings. High correlations were observed among some of these indicators.

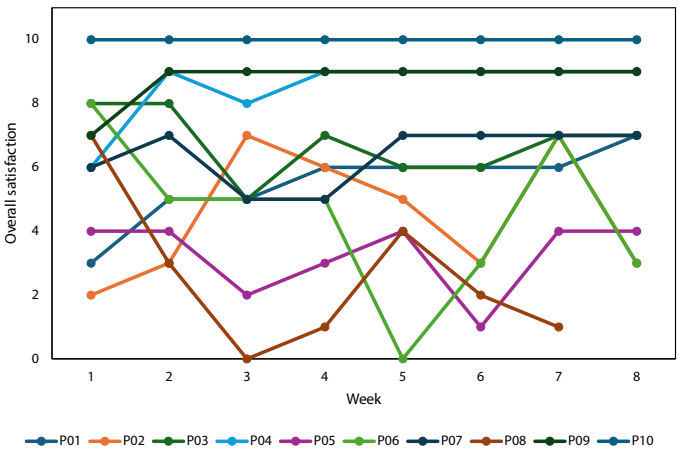


Figure 1: Satisfaction trends over eight weeks for each participant.

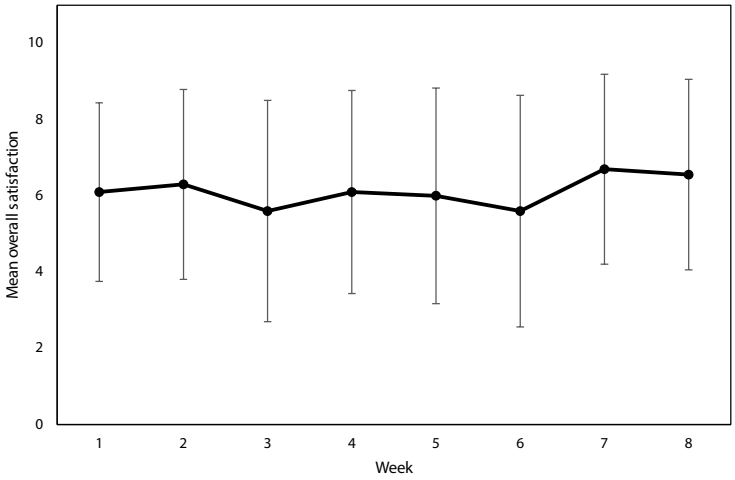


Figure 2: Mean weekly satisfaction across participants.

Table 1: Correlation coefficients among the indicators calculated from daily satisfaction ratings.

	Max	Min	Mean	End
Max	-	0.37	0.83	0.51
Min	**	-	0.68	0.51
Mean	**	**	-	0.72
End	**	**	**	-

Variability Between and Within Users

A linear mixed-effects model was estimated, yielding a conditional R^2 of 0.76 and a marginal R^2 of 0.66. The fixed effects are shown in Table 2, and the variance inflation factors (VIFs) are presented in Table 3. As shown in Table 3, all VIF values were below 6, indicating no severe multicollinearity. Regarding the random effects, the variance attributable to individual differences among participants was 0.689, and the residual variance was 1.579. Based on these values, the intraclass correlation coefficient (ICC) was calculated to be 0.30. This indicates that approximately 30% of the variability in weekly satisfaction was attributable to individual differences among participants.

Table 2: Fixed-effects of the estimated linear mixed-effects model.

Predictor	β	SE	95% CI		t
			Low	High	
Week2	0.37	0.57	–0.78	1.51	0.64
Week3	–0.40	0.58	–1.56	0.75	–0.70
Week4	0.24	0.57	–0.90	1.39	0.43
Week5	0.52	0.58	–0.63	1.67	0.90

(Continued)

Table 2: Continued.

Predictor	β	SE	95% CI		t	
			Low	High		
Week6	0.72	0.59	–0.46	1.89	1.22	
Week7	0.42	0.59	–0.76	1.60	0.72	
Week8	0.50	0.59	–0.69	1.68	0.84	
Max	0.20	0.15	–0.10	0.51	1.36	
Min	–0.19	0.09	–0.36	–0.02	–2.23	*
Mean	0.61	0.18	0.25	0.97	3.41	**
End	0.09	0.06	–0.03	0.22	1.51	

Table 3: Variance inflation factors (VIFs) for the fixed-effect variables.

Predictor	VIF
Week	1.37
Max	3.31
Min	2.24
Mean	5.77
End	1.47

Relationship Between Daily And Weekly Satisfaction

As shown in Table 2, none of the Week coefficients were significant, and no clear uniform temporal trend in satisfaction was observed over the eight-week period. In addition, among the indicators calculated from the daily satisfaction ratings, the Mean ($p < 0.01$) and Min ($p < 0.05$) values were significant. In contrast, Max and End were not significant.

Characteristics of Usage Episodes

All participants provided episode descriptions for each day across the eight weeks, resulting in a total of 553 episodes entries with no missing responses. Table 4 summarizes the usage patterns and characteristic episodes for each participant. Considerable variation was observed in how participants engaged with the device. Based on these descriptions, participants were descriptively grouped into three broad patterns of use. A first group (P04, P09, P10; mean overall satisfaction = 9.1, SD = 1.0) showed consistently high satisfaction and engaged actively with multiple features of the device, including sleep monitoring, heart rate tracking, and step counting. Their episodes frequently described feature discovery and successful integration of the device into daily routines. A second group (P01, P03, P06, P07; mean = 5.8, SD = 1.7) showed

moderate satisfaction levels and more varied engagement, using the device selectively—primarily as a watch or timer—with occasional exploration of additional features. Their episodes also included instances of forgetting to wear the device or encountering minor difficulties. A third group (P02, P05, P08; mean = 3.5, SD = 2.0) showed consistently lower satisfaction, characterized by limited engagement, frequent non-use, and in some cases specific negative incidents such as device unavailability or unmet feature expectations.

Table 4: Summary of usage patterns and characteristic episodes by participant.

Participant	Overall Sat. Mean (Sd)	Primary Usage Pattern	Characteristic Episodes	Notable Incidents
P01	5.5 (1.2)	Watch, step counting; gradually habituated	Feature discovery (timer, heart rate); occasional non-use	Unexplained vibrations; charging issues
P02	4.5 (2.0)	Step counting only; limited engagement	Motivated by high step counts on active days	Left device in another city (Week 8); sat. dropped to –5
P03	6.8 (1.0)	Watch and timer; stable selective use	Timer used extensively in laboratory work	Frequent non-use days (33% zero days)
P04	8.5 (1.1)	Active multi-feature use	Sleep, heart rate, social sharing	Occasional non-use due to sports/ work
P05	3.2 (1.2)	Minimal; watch and timer only	Use gradually became routine and disengaged	Interest declined over weeks
P06	4.5 (2.5)	Notification-focused; variable	Found notification feature valuable	COVID-19 isolation caused one week of non-use
P07	6.4 (0.9)	Sleep and health monitoring	Interested in heart rate and sleep quality	Repeated data recording failures (Week 6)
P08	2.6 (2.4)	Cycling-focused; frustrated	GPS sharing attempted repeatedly	GPS limitations caused early dissatisfaction
P09	8.8 (0.7)	Health monitoring; highly engaged	Sleep, steps, heart rate tracked consistently	Minimal non-use throughout
P10	10.0 (0.0)	Watch only; fully habituated	Consistent time-checking across all activities	No negative incidents reported

DISCUSSION

Although fluctuations were observed within individuals, no significant differences between weeks or uniform temporal trend across participants were observed over the eight-week period. These results imply that when

monitoring UX over a long period, a uniform trend across all participants may not necessarily be observed. Therefore, it is important to interpret the results not only by focusing on overall trends but also by considering each individual's usage context and episodes. From a practical standpoint, this suggests that weekly reporting can serve as a useful monitoring tool for detecting individual-level fluctuations that would be invisible in aggregated or single-point assessments.

The ICC of 0.30 indicated that individual differences accounted for approximately 30% of the variability in weekly satisfaction. This suggests that individual differences should be considered when tracking UX longitudinally.

The relationship between daily satisfaction and weekly overall satisfaction was examined. Mean and Min were significantly associated with weekly overall satisfaction, whereas Max and End were not. In particular, Mean showed a relatively large positive coefficient, suggesting that weekly overall satisfaction was associated with the average level of daily satisfaction during that week. Although Max and End were included as indicators corresponding to the peak–end rule, neither was significantly associated with weekly overall satisfaction in the present model, suggesting that the pattern observed here did not clearly conform to the peak–end rule (Kahneman et al., 1993).

The findings suggest that weekly satisfaction reporting may provide a relatively simple approach to tracking individual satisfaction trajectories over extended periods of product use. Such repeated reporting may help reveal individual-level fluctuations that are difficult to capture through single-point assessments. For example, designers could ask users to report weekly satisfaction during early product use and examine how experiences change over several weeks. Such information may help identify periods of unstable experience, negative episodes, or individual differences that are difficult to capture through conventional short-term usability testing.

Several limitations should be noted. All participants were Japanese university students, and the study focused on a single product, limiting the generalizability of the findings. In addition, the present approach was not directly compared with other long-term UX evaluation methods. Finally, daily satisfaction was retrospectively reported at the end of each week, which may have introduced retrospective bias. Future studies should examine these issues using more diverse samples and products and alternative data collection designs.

CONCLUSION

The results showed no clear uniform temporal trend in satisfaction over the eight-week period. Approximately 30% of the variability in satisfaction was attributable to individual differences between users. In addition, weekly overall satisfaction was significantly associated with the mean and the minimum values of daily satisfaction ratings, whereas the maximum and end-of-week values were not significant. These findings suggest that even a method based on recording satisfaction on a weekly basis can capture certain characteristics of changes and individual differences in UX during long-term product use. Although this study was an exploratory investigation based on

a small sample and a single product and therefore should be interpreted cautiously in terms of generalization, weekly reporting for UX monitoring may serve as a relatively simple approach for understanding long-term user experience in practice. Future studies should further examine the effectiveness of this approach by applying it to a wider variety of products and participants and by comparing it with other long-term UX evaluation methods.

ACKNOWLEDGMENT

This research was supported by the Salon de K Foundation's research grant program for 2024 and JSPS KAKENHI Grant Number 22K18140, 26K15511.

REFERENCES

Csikszentmihalyi, M., Larson, R. (1987). "Validity and reliability of the experience-sampling method", Journal of Nervous and Mental Disease, 175(9), pp. 526–536.

Doi, T. (2026). "Effects of anticipated UX on post-use satisfaction: The relationship between users' previous expectations and attitudes prior to using a product and evaluation of satisfaction after using the product", Theoretical Issues in Ergonomics Science, 27(3), pp. 259–270.

Doi, T., Doi, S., Yamaoka, T. (2022). "The peak–end rule in evaluating product user experience: The chronological evaluation of past impressive episodes on overall satisfaction", Human Factors and Ergonomics in Manufacturing & Service Industries, 32(3), pp. 256–267.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A. (1993). "When more pain is preferred to less: Adding a better end", Psychological Science, 4, pp. 401–405.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A. (2004). "A survey method for characterizing daily life experience: The Day Reconstruction Method (DRM)", Science, 306, pp. 1776–1780.

Kujala, S., Miron-Shatz, T. (2015). "The evolving role of expectations in long-term user experience", Proceedings of the 19th International Academic Mindtrek Conference, ACM.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., Sinnelä, A. (2011). "UX Curve: A method for evaluating long-term user experience", Interacting with Computers, 23(5), pp. 473–483.