



Can UX Over Time Be Reliably Evaluated? - Verifying the Reliability of ERM

Masaaki Kurosu^{1(✉)} and Ayako Hashizume²

¹ The Open University of Japan, Tokyo, Japan
nigrumamet-s23@mbr.nifty.com

² Hosei University, Tokyo, Japan

Abstract. Experience Recollection Method (ERM) was announced at HCII 2018 conference by the author and is a research method for the UX (User Experience) that evaluates the contents and degrees of UX. It is a memory-based measurement method for the UX where the informants are asked about the content and degree of episodes concerning the use of targeted artifact (product, system or service) along with the rough time scale. Although the method requests episodes on the artifact usage, it doesn't ask the time information in detail considering the vagueness of memory, hence it does not generate the visual representation of UX curve/graph in such methods as CORPUS, iScale, UX Curve and UX graph.

This presentation is based on the data that was obtained in FY 2017 for the same informants once in September and another in January. We compared two data and checked if the information obtained by ERM has a certain degree of reliability, i.e. the episode in the first survey is kept in the second survey and if the rating scale value are almost the same. This presentation is based on the comparison of two datasets. Generally speaking, ERM was confirmed to have the high reliability and provide a reliable information on the UX.

Keywords: UX evaluation · Reliability · Memory-based method

1 Evaluation of UX

1.1 Previous Approach

The UX (User Experience) can be defined as “person’s perceptions and responses resulting from the use and/or anticipated use of a product, system or service (ISO9241-11:2010) [1].” As an extension of this definition, Kurosu [2–5] proposed that the quality in design including the usability is the basis for the quality in use and the latter is related to the UX. Thus, the UX should be evaluated or measured in relation to wider range of quality characteristics including usability, functionality, performance, reliability, safety, attractiveness, etc.

From this perspective, evaluation methods that have been proposed until now will be screened to a small set of methods, although there are so many so-called “UX evaluation methods” have been proposed as listed in the website of All About UX [6].

UX evaluation methods can be classified into two categories; real-time UX evaluation, and memory-based UX evaluation.

The real-time UX evaluation methods including the questionnaire can obtain the information in-situ and just in time. But, because of their invasive nature, i.e. informants are requested to answer the question during their everyday life and it is sometimes an obstruction for them, it is not recommended to repeat the survey for more than a few weeks. In other words, real-time UX evaluation methods can be repeatedly applied only for a limited time range.

On the contrary, memory-based UX evaluation methods do not have such temporal limitations. They can be applied to the UX over the long-term period. But they do have limitations originated from the nature of human memory. People forget many events even though they were important, and people may also edit or change the contents without any ill will. Hence the validity and reliability are important in terms of the memory-based methods.

Anyways, real-time evaluation methods include the questionnaires and methods for evaluating emotion and other methods. The questionnaire includes SUS [7], SUPR-Q [8], Product Reaction Card [9] and AttrakDiff [10]. The evaluation methods for emotion include 3E [11] and Emo2 [12], and the other real-time evaluation methods include ESM [13], and diary methods such as DRM [14, 15] and TFD [16].

Memory-based methods include CORPUS [17], iScale [18], UX curve [19], UX graph [20] and ERM [5, 21].

1.2 ERM

ERM was proposed by Kurosu et al. [5, 21] based on the reflection of advantages and disadvantages of memory-based methods until now. Similar to previous methods, informants are asked past events (episodes) and the rating for them. But the curve or graph will not be drawn in ERM, because of the idea that the memory will not have such preciseness as can be represented as the coordinate on time scale. As can be seen in Fig. 1, informants are given only 7 rough time zones: expectation, purchase, early use, use, recent use, present time and near future that include all phases of experience for an artifact (product, system or service).

Each time zone means;

- Expectation: estimation of UX before the purchase
- Purchase: evaluation of UX at the purchasing or obtaining the artifact
- Early use: evaluation of UX just after the purchase (around a few weeks to a few months)
- Use: evaluation of UX after the early use until the recent use. This may range from a few weeks to several years depending on the time of purchase and the time of survey
- Recent use: evaluation of UX just before the time of survey (around a few weeks to a few months)
- Present time: evaluation of UX at the time of survey
- Near future: expectation and/or estimation of UX after the time of survey

Although ERM uses the letter-sized paper and the number of the row is limited, informants are allowed to write more than one episode in a row. Every episode should accompany the rating on the feeling from positive (+10) to negative (−10), i.e. 21 points scale is used.

Recording Sheet for ERM: Experience Recollection Method Target Item: University Male/Female Age: 18

1. Write what you experienced at each phase and fill in the evaluation by +10 to -10 rating.

Phase	What you experienced	Evaluation
Expectation before the purchase	<i>Expecting that I can learn what I could not learn at high school.</i>	3
Evaluation at the time of start of usage	Year: <u>2016</u> <i>May be able to learn various subjects.</i>	4
Evaluation at early days from the start of usage	<i>It's located very far from home</i>	-7
	<i>Interesting to study, but teachers don't let me know how to study.</i>	-3
Evaluation during the use	<i>Science and technology are interesting</i>	2
	<i>Could get many good friends.</i>	4
	<i>Very interesting to go to the university.</i>	6
	<i>Too much home tasks and have little time to do them.</i>	-5
Recent evaluation	<i>There are many interesting lectures.</i>	1
	<i>There is a limit of the number of lectures that I can take.</i>	-2
Present evaluation	Year: <u>2017</u> <i>There are many restrictions.</i>	2
Estimation in the near future	<i>Would like to study more seriously.</i>	4

Fig. 1. An example of ERM regarding the university education. (translated from Japanese and cited from Kurosu et al. 2018)

1.3 Reliability of ERM

Similar to the psychological test, UX evaluation methods should possess a certain level of validity and reliability. Regarding the validity especially the content validity, there will be no problem because informants are asked about their own experience regarding the use of an artifact they've been using. But the reliability is not yet confirmed and there seems to be no researches conducted to investigate this issue. This is the reason why this paper deals with the reliability issue of the memory-based UX evaluation method.

2 Verifying the Reliability of ERM

2.1 Method

The basic idea of the verification of reliability of ERM is to compare the result of two surveys for the same group of informants, i.e. re-test method. Luckily, I had a class at the graduate school and thought that I should ask students to collaborate for the survey twice, once at the first lecture and another at the last lecture of the semester. And the first survey was conducted on Sep. 25, 2017 and the second survey on Jan. 22, 2018. There were 119 days between two surveys. It would have been almost impossible for the informants to remember what they answered at the first survey when they were subject to the second survey.

Informants

There were 26 students registered to my class and the attendants at the lecture on Sep. 25, 2017 and Jan. 22, 2018 were shown in Table 1. Because of the absence at each lecture, total of 23 students attended either of the lecture and, from among them, 17 students attended both lectures resulting 17 data available for the analysis. Unfortunately, it became clear that 3 of them purchased a new model during the analysis and

were excluded from the analysis that followed. Finally, 14 data were actually used in the total analysis. To our regret, all the students were male perhaps because my class was opened at the engineering department.

Table 1. Attendants of the class on Sep. 25, 2017 and Jan. 22, 2018

		Jan. 22, 2018		
		Data that can be used	Data that can't be used	Sum
Sep. 25, 2017	Data that can be used	17	1	18
	Data that can't be used	5	0	5
	Sum	22	1	23

Targeted Device

Because all of the students were using the smartphone (not the cellphone), it was decided to be the targeted device. Because it is the multi-purpose device and almost all users are using it daily, or more to say, many times during the day, its user must have various experiences from a positive one to a negative one.

Procedure

ERM sheet was delivered to each of the informants, then a brief instruction was given, and 30 min were allowed to write down their experiences. During the instruction, informants were told that

- This is to ask you to write down your personal experience on the smartphone
- First, write down your university ID, sex, age, and description of the smartphone
- There are seven time-zones including the expectation, purchase ... near future (with the explanation of each time range)
- You are requested to write down what you experienced in the episode slot and the degree of satisfaction vs. dissatisfaction or positive feeling vs. negative feeling from +10 through 0 to -10 depending on your subjective impression in the rating slot. Please use the integer
- You may begin with the expectation to the near future, but it depends on your feeling which time zone you would write. You can go back to the previous time zone that you have already filled
- Although there are limited number of slots, you can write two or more episodes and ratings in one slot if you need more than one
- If you don't remember what you experienced at any time zone, you can skip it

Obtained Data

Handwritten ERM sheets were obtained and were input to Excel, then were translated into English. Appendix shows all the raw data.

3 Results

3.1 Rating Results

Rating values correspond to the vertical position of the curve/graph in CORPUS, iScale and UX curve. But unlike those methods, ERM only separates rough time zones corresponding to the quasi-continuous horizontal position in them.

One point that should be warned for the use of the time zone during the reliability verification is the meaning of each time zone in the first survey and in the second survey. As shown in the imaginative data of Fig. 2, each time zone shifts bit by bit depending on the displacement of the time when the surveys were conducted. For example, to an imaginative informant in this figure, the smartphone was purchased, of course, at the same time. But the following time zone represents a bit displaced physical time depending on when the survey was conducted. For example, the present time at the first survey was Sep. 2017 while the present time at the second survey was Jan. 2018. This displacement will become larger as the span between the purchase and the survey becomes shorter.

Fortunately, informants who purchased their smartphone in 2017 (informant C, F, H, I, L and M) showed similar ups and downs of the rating value for the first survey and second survey as can be seen in the following sections. This may mean that the time zones for the informants were not exact but rough, hence horizontal position in curves/graphs in CORPUS, iScale and UX curve were not exact based on the equal time unit and thus they should be called quasi-continuous.

	2016		2017				2018	
				Sep. 25, 2017	119 days	Jan. 22, 2018		
First survey	expectation	purchase	early use	use	recent use	present	near future	
Second survey	expectation	purchase	early use		use	recent use	present	near future

Fig. 2. Chronological table showing the difference of meaning of the time zone. In this imaginary data, the same informant purchased the smartphone in early 2017, but the time zones from early use to near future slip out of place in the first survey and the second survey depending on the time when the survey was conducted (2017 and 2018).

3.2 Reliability Measure

Usually, the reliability (ρ) is represented by the correlation coefficients (r). In this study, Kendall’s coefficient of concordance (W) was also calculated. These values were calculated based on the average rating for 7 time zones. The distribution of r is shown in Fig. 3 and that of W is shown in Fig. 4. These graphs show rather high reliabilities.

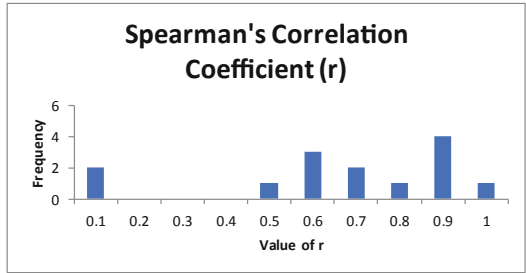


Fig. 3. Distribution of Spearman's correlation coefficients for 14 data.

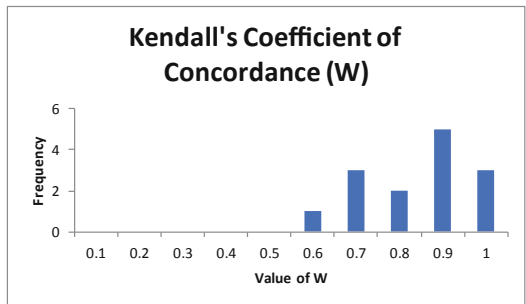


Fig. 4. Distribution of Kendall's coefficient of concordance for 14 data.

3.3 Episode Results

Episodes were verbal in nature, thus will be analyzed one by one in the next chapter.

4 Analysis of Each Data

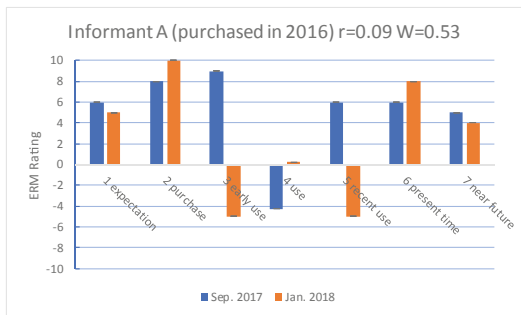


Fig. 5. Ratings by informant A

ID of each informant was randomly assigned. Episode is assigned its ID as <informant ID> episode number><year> , e.g. A4-2017. Please refer to the Appendix.

Informant A generally gave positive ratings except for the size of device that was negatively rated (A12-2017 and A13-2018) and its weight that was also negatively rated (A12-2017 and A8-2018) (Fig. 5).

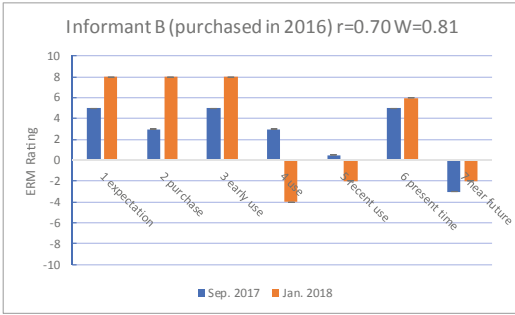


Fig. 6. Ratings by informant B

Informant B generally gave positive ratings especially in 2017. But he rated negatively regarding the future (B10-2017, B10-2018) that the device may not be able to correspond to the new applications (Fig. 6).

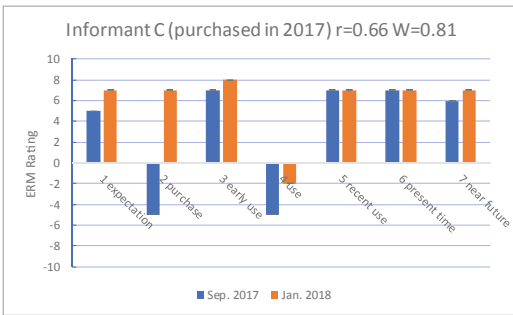


Fig. 7. Ratings by informant C

Informant C gave strange ratings to the same aspect of the device that the specification is almost the same with the previous cellphone, one negatively and another positively (C2-2017, C2-2018) (Fig. 7).

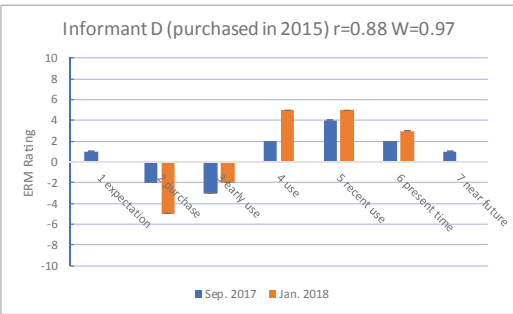


Fig. 8. Ratings by informant D

In the beginning of use, informant D felt a negative impression on the trace of finger print on the touch-panel (D2-2017, D4-2018), but his evaluation gradually became positive during the usage (Fig. 8).

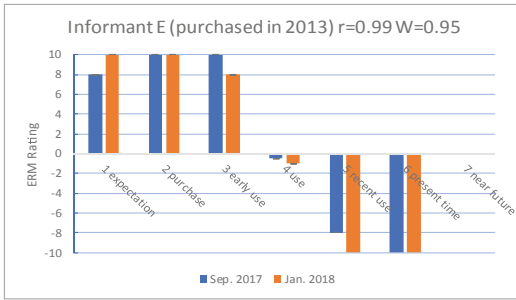


Fig. 9. Ratings by informant E

Informant E showed a drastic change of ratings that was quite high in early days and changed into negative during the usage in terms of the battery life (E9-2017, E9-2018), and the wifi connection (E10-2017, E10-2018) (Fig. 9).

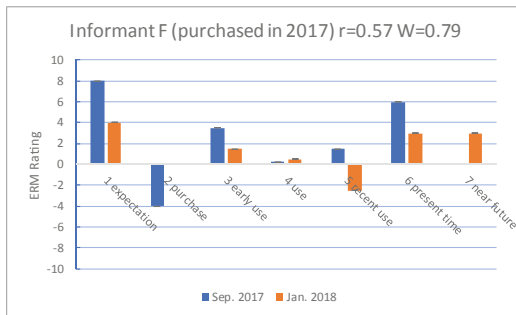


Fig. 10. Ratings by informant F

Informant F wrote episodes differently for 2017 and 2018. But there are some common episodes such as the speed (F4-2017, F5-2018), the quality of photograph (F7-2017, F6-2018), the convenience of second screen (F11-2017, F8-2018), etc. (Fig. 10).

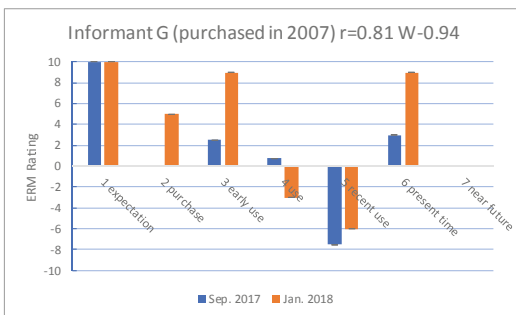


Fig. 11. Ratings by informant G

Informant G wrote about the joy of accessing internet (G1-2017, G1-2018) and that of using net contents (G5-2017, G3&G4-2018) positively (Fig. 11).

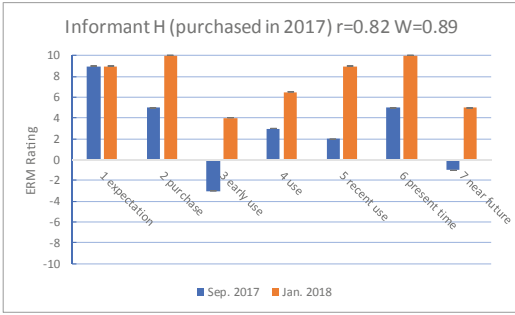


Fig. 12. Ratings by informant H

Informant H wrote about the high expectation (H1-2017, H1-2018) and the screen quality (H2-2017, H2-2018). Generally his ratings are higher in 2018 (Fig. 12).

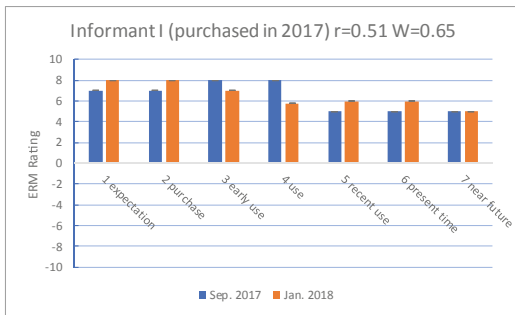


Fig. 13. Ratings by informant I

Informant I gave no negative ratings. Positive evaluations are for the processing speed during the expectation (I1-2017, I1-2018) and the early use (I3-2017, I3-2018). Strangely he rated the fast battery loss positively (I10-2017, I9-2018) and he might have misunderstood the instruction (Fig. 13).

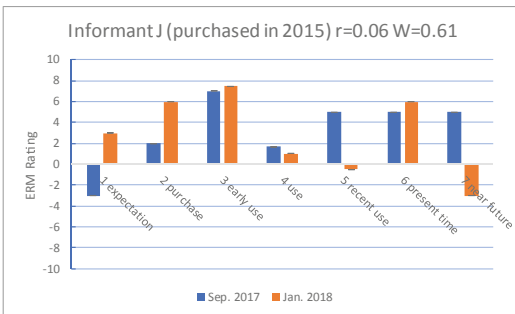


Fig. 14. Ratings by informant J

Informant J had different expectation one negatively for the poor operability (J1-2017) and another positively for the good performance (J1-2018). This informant did not give the consistent episodes except for the present time evaluation (J12-2017, J12-2018) (Fig. 14).

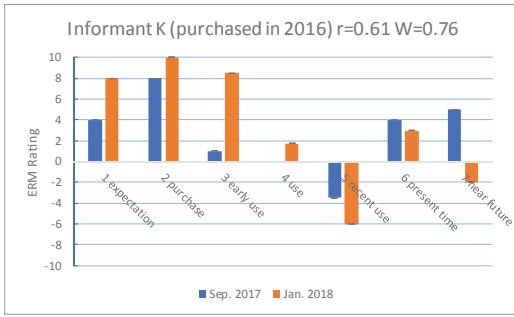


Fig. 15. Ratings by informant K

Informant K generally gave positive ratings except recently for the finger print recognition (K9-2017, K9-2018) and the unexpected break down (K10-2017, K10-2018) (Fig. 15).

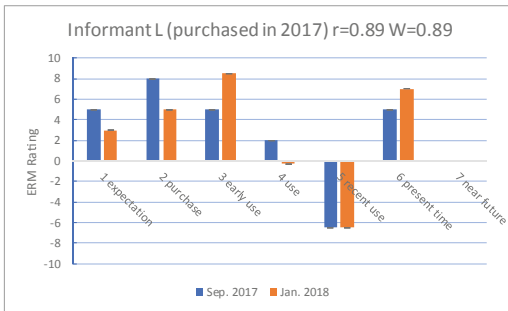


Fig. 16. Ratings by informant L

Informant L gave negative ratings only recently for the lack of storage (L9-2017, L9-2018). Another negative evaluation was given differently, one to the heat (L10-2017) and another to the system freeze (L10-2018) (Fig. 16).

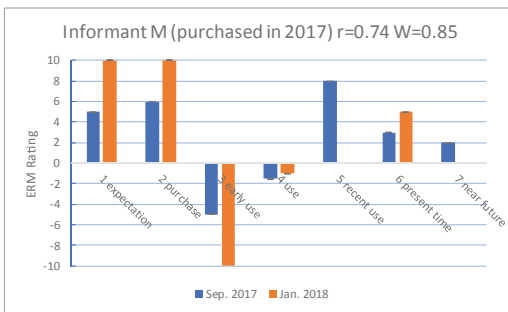
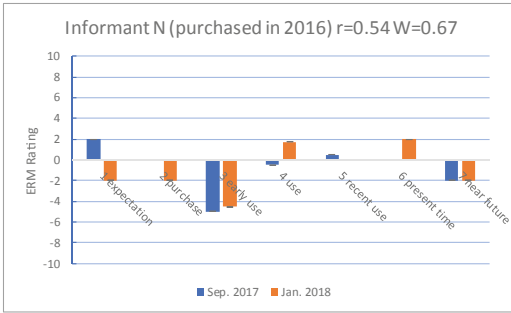


Fig. 17. Ratings by informant M

Informant M complained for the same problem of gyro sensor that occurred during the early use (M3-2017, M3-2018). He still pointed out that problem at the present time (M7-2017, M7-2018) (Fig. 17).



Informant N pointed out the usability of the plastic case cover (N2&N3&N4&N9&N10-2017, N3&N9-2018) generally a bit negatively (Fig. 18).

Fig. 18. Ratings by informant N

5 Conclusion

The reliability of ERM was tested in terms of the smartphone using the re-test method. Two surveys for the same 14 informants who continued to use the same model were conducted one on Sep. 2017 and another on Jan. 2018. By the use of ERM, episodes and subjective ratings were obtained for 7 time zones including the expectation, purchase, early use, use, recent use, present and near future. Two reliability measures (one is the correlation coefficient and another is the coefficient of concordance) were calculated and relatively high reliability was confirmed.

Based on the content analysis for each informant, the same episodes were found around the same time zone and were rated in the same way. This also confirmed the high reliability of ERM

Appendix

Raw data (episodes and ratings) of ERM at Sep. 2017 and Jan. 2018 for all 14 informants are shown in following tables.

ID	Age Sex	Device	When	Sequential ID	Episode	Rating	Jan. 2018	Rating				
A	M 24	Iphone 2	expectation	A1	is big and will be easy to use	6	Good reputation on the internet and among friends made me feel much expectation.	5				
				A2	Cost effective than the smartphone	6						
				A3	May be used in many ways	6						
			purchase	A4	Screen is indeed big	8						
				A5	Body is also big than expected	8						
				A6	Beautiful, good designed	8						
			early use	A7	Cool	8						
				A8	Can be used for wider range of purpose than the smartphone	9						
				A9	Screen is big and is easy to see. Less eye strains	9						
			use	A10	Some applications do not work on iPad	-3						
				A11	Consumes large amount of communication fee	-4						
				A12	Because of its size, it's bulky and heavy	-5						
				A13	Photo-taking is not easy	-5						
				A14	Needs both hands for using it	-6						
				A15	Slow decrease of battery (good battery life)	8						
			recent use	A16	I can focus on my main job while using this device	4						
				present time	A17	Though big and heavy, it's no importance when I adapted it. It's easy to use.			6			
					A18	If possible, I'd have a device one size smaller. But, for the moment, I'm satisfied with this			5			
B	M 23	Iphone SE	expectation	B1	I was expecting the handy size and better performance than iPhone5S I used to use	5	Good performance will allow me to use without irritation	8				
				B2	Size and shape were same with the previous one and I could accept it easily	3						
			purchase	B3	I was astonished to see the high performance than expected	7						
				B4	Because the size and shape were same with iPhone5S, I could use its case to the new one	3						
			early use	B5	No sudden shut down occurred and I can use it feeling relieved, and I had no complaints	3						
				B6								
			use	B7	Some recent applications require more powerful specification and my device cannot correspond to them. A bit of dissatisfaction	-2						
				B8	As it worked smoothly on a new OS, I felt relieved	3						
										Overall performance is good and tough than other phones	8	
										Will become lighter, with better performance and resolution will be better	4	
					Processing is fast and could use it without irritation	8						
					Installed various applications and found that most of them run fast	8						
					Smartphone that I used before didn't crack the screen but iPhoneSE cracked instantly when I dropped it	-5						
					When applications that was released later were installed, I felt the speed slowed down	-3						
					A bit more frequently froze	-2						

		present time	2017	D12			
E M X 23 p e r i a S o - o 2 E		near future expectation purchase early use use recent use	2017	D12	For me, my device is for game applications as well as the communication	2	3
				D13	Will use the device mainly for game applications	1	
				E1	Expected so much because my device was the forefront model at the time of purchase	8	10
				E2	Was astonished for the specification that is far better than the previous model	10	10
				E3	Very convenient for using SNS	10	8
				E4	Very pleasant to use because of the latest 3G circuit	10	8
				E5	No dissatisfactions for about 2 years after the purchase	9	6
				E6	As the advent of 4G circuit, I started to feel the lack of power	-3	-2
				E7	Power button was broken and I could not off the power supply	-4	-3
				E8	Because of the lack of power, I don't enjoy games on this device	-4	-5
F M P 23 R G O L - e o t J		near future expectation purchase early use use	2017	E9	Sometimes, the battery residual goes down to 0% suddenly. Very unpleasant	-8	-10
				E10	No WiFi connections and am worried so much	-8	-10
				E11	Degradation over time and lack of power make me feel disastrous	-10	-10
				E12	Will purchase a new model whenever I can afford	0	0
				F1	Expecting its high power based on the catalog specification compared to the previous model	8	4
				F2	Because the power button is not located on the side but on the back of body, it is difficult to use	-4	0
				F3	Made much mis-touches	-4	
				F4	Performance is good. Applications that didn't work on the previous model works well	6	-2
				F5	HIFI sound is good. Pleased so much because I was not expecting it	1	5
				F6	Because it uses TypeC cable, I had to buy some peripherals anew	-3	4
J o n i c s v		use		F7	Performance of the camera was found to be very nice. Though I'm using it in automatic mode now, it would be wonderful if I get used to it.	2	-1
				F8	Decrease of battery is a bit faster	-2	1
				F9	Charging speed is fast	4	-2
				F10	Pleased that the smartphone VR can be used	4	-4

I	M22	X P e r i a Z 5 S O - 0 I H	use	H4	Inconvenient that the reboot of application frequently occurs. Perhaps it is for protecting from the overload	-3			
				H5	Accustomed to it now and can operate it easily	3	At last, I could understand how to use the leather case, but it has the risk of slipping down from hand		
				H6			8	Longtime use made the leather case soft and easy to use	
				H7	Positive point is that it accepts heavy processes by good specification against the high load.	2	Inserted the student ID to the leather case and found it convenient. But the LSI of ID card sometimes fails to be read by the gate machine		
				H8	2017 present time near future	5	Very easy to use	10	Performance has not yet degraded and can be used satisfactorily
				H9		-1	What will happen when all the storages are filled with data?	5	Performance is still good but peripheral devices have become degraded
				I1		7	Processing is fast	8	Will operate smoothly
				I2	2017	7		8	Screen is large
				I3	early use	8	Runs smoothly, adequate weight and size	7	Screen images move smoothly
				I4				7	Battery works for a long time
				I5		8	It is a major and popular model so that there are many covers available, furthermore my friend has the same model.	5	During the game playing, the area around NFC becomes hot
				I6	use			5	Fingerprint lock cannot always be released easily when finger tips are moist
				I7				6	Works smoothly without time lag
				I8				7	Can be used under the rain as it is life waterproof
				I9				6	Battery seems to be lowered than before
J	M22	i P h o n e S e 6 S	recent use	I10	Battery consumption is fast. Found a small scratch. Get hot while using it	5	Has a small scratch		
				I11		5	Moving without any troubles		
				I12	I will change to a new model	5	Battery will be lowered faster than now		
				J1	Worried about the difficulty of one hand operation as I heard that the screen size is larger than previous model	-3	Expecting the high efficiency by the high powered CPU compared to the current model (iPhone4S)		
				J2			3	Anxious about the less easy holding of the device because of the large screen	
				J3	2015	2	Screen size is larger and it's easy to read text on the screen	6	Large screen allows me to look letters and images easily. Satisfied
				J4				6	Difficult to put the device into the small pocket of the pants
				J5		8	At first, I used it by both hands because of the large screen, but now, I can operate it by one hand	8	Feeling satisfaction and the sense of superiority based on the high performance compared to other devices

Year	Model	Usage	Year	Model	Expectation	Actual Experience	Rating
2017	K12	early use	J6	J6	Specifications are far better than the previous one and I'm satisfied	Specifications are far better than the previous one and I'm satisfied	6
			J7	J7	Processing speed is faster than other smartphones. I'm satisfied	Processing speed is faster than other smartphones. I'm satisfied	7
			J8	J8	A problem occurred that the system shuts down suddenly maybe because of the battery malfunction	A problem occurred that the system shuts down suddenly maybe because of the battery malfunction	-5
			J9	J9	Though there are problems that applications stop suddenly for previous model 1 or 2 generations before, this model doesn't have such problems	Though there are problems that applications stop suddenly for previous model 1 or 2 generations before, this model doesn't have such problems	3
			J10	J10	New models are appearing one after another in the market, but I don't think my model lacks the power	New models are appearing one after another in the market, but I don't think my model lacks the power	5
			J11	J11	My model has sufficient power for the ordinary use	My model has sufficient power for the ordinary use	5
			J12	J12	Recent smartphones have over specifications and have some inconvenience such that there is no earphone jack	Recent smartphones have over specifications and have some inconvenience such that there is no earphone jack	5
			J13	J13	Purchased for an inexpensive price and the high performance cannot be expected. But because it was the latest model, I expected its performance to a certain extent	Purchased for an inexpensive price and the high performance cannot be expected. But because it was the latest model, I expected its performance to a certain extent	4
			K1	K1	Appearance was better than the photograph in catalogue and the size was good to hold	Appearance was better than the photograph in catalogue and the size was good to hold	8
			K2	K2	It was stressful because of the low communication speed perhaps due to the low price	It was stressful because of the low communication speed perhaps due to the low price	-4
			K3	K3	I could use it freely without any instability for any latest applications	I could use it freely without any instability for any latest applications	6
			K4	K4	Though it has the sufficient specification for AR and VR applications, it didn't accept such applications	Though it has the sufficient specification for AR and VR applications, it didn't accept such applications	-8
K5	K5	Functions for which I decided to but this model are now convenient tool for everyday use	Functions for which I decided to but this model are now convenient tool for everyday use	5			
K6	K6	Malfunctions of calendar occurs many times	Malfunctions of calendar occurs many times	-3			
K7	K7	It was customizable in detail for my own purpose	It was customizable in detail for my own purpose	6			
K8	K8	Finger print verification does not always work properly	Finger print verification does not always work properly	-4			
K9	K9	Unexpected things occur frequently	Unexpected things occur frequently	-3			
K10	K10	Compared to other models, it has not big advantages. But it doesn't have any deficiencies considering the price	Compared to other models, it has not big advantages. But it doesn't have any deficiencies considering the price	4			
K11	K11	I will use this model until when the next generation model will appear with latest functions	I will use this model until when the next generation model will appear with latest functions	5			
K12	K12						
2016	K11	recent use	J6	J6	Screen size that seemed to be too large is now felt just right to my hand	Screen size that seemed to be too large is now felt just right to my hand	7
			J7	J7	New applications appeared in the market for which other devices cannot correspond	New applications appeared in the market for which other devices cannot correspond	5
			J8	J8	Initial failure of the battery was found	Initial failure of the battery was found	-7
			J9	J9	No scratches on the screen after the long time use	No scratches on the screen after the long time use	5
			J10	J10	New model has some degradation such as the abolition of input jack	New model has some degradation such as the abolition of input jack	3
			J11	J11	Performance degradation by updating iOS was now revealed	Performance degradation by updating iOS was now revealed	-4
			J12	J12	Latest applications run without any problems. Am satisfied on the whole	Latest applications run without any problems. Am satisfied on the whole	6
			J13	J13	CPU won't be able to catch up the application. Battery consumption	CPU won't be able to catch up the application. Battery consumption	-3
			K1	K1	It was revealed in the specification document that the performance will be better than the smartphone that I have	It was revealed in the specification document that the performance will be better than the smartphone that I have	8
			K2	K2	Appearance is more beautiful than in the photograph	Appearance is more beautiful than in the photograph	10
			K3	K3	Applications that require high specification so that I could not install to the previous device could easily be installed	Applications that require high specification so that I could not install to the previous device could easily be installed	9
			K4	K4	Camera function is far better than I expected	Camera function is far better than I expected	8
K5	K5	Dual SIM function worked well when I went to foreign countries	Dual SIM function worked well when I went to foreign countries	5			
K6	K6	Age verification of LINE application needed the carrier line and I could not confirm it	Age verification of LINE application needed the carrier line and I could not confirm it	-5			
K7	K7	UI changed by updating the OS and the device has become less usable	UI changed by updating the OS and the device has become less usable	-2			
K8	K8	UI could be changed rather freely and I customized it	UI could be changed rather freely and I customized it	9			
K9	K9	Precision of finger print verification has become worse	Precision of finger print verification has become worse	-6			
K10	K10	Applications stop suddenly frequently when many applications are active - maybe due to the amount of memory	Applications stop suddenly frequently when many applications are active - maybe due to the amount of memory	-6			
K11	K11	Any applications can be used, but sometimes the speed becomes slow	Any applications can be used, but sometimes the speed becomes slow	3			
K12	K12	Does not accept all the latest applications (AR, VR). I will change it to the new one	Does not accept all the latest applications (AR, VR). I will change it to the new one	-2			

L	M24	Xper	expectation	L1	Expected for the graded up quality of photo software compared to the previous model	5	Purchased at the renewal after 2 years term, I had not big expectation	3
				L2	Screen/glass was a bit curved and the feeling when I touched it was good	8	Camera, music playback and finger print verification were new to me. No unsatisfactory points.	5
				L3	Battery life is far better than the previous model	10	Lock release by finger print verification was much more convenient than was expected	7
				L4	Usability of automatic camera function is not so much better than the previous model	0	Battery life was much longer than previous versions	10
				L5	5.2 GHz can be used because the WiFi corresponded to IEEE80211	5	Regretful that some softwares that I've used cannot be used because of the lack of compatibility. Difficult to find the alternative software	-5
				L6	Noticed that there are some convenient applications among pre-installed software provided by Sony	3	Not unsatisfactory for functions that I used everyday	3
				L7	Finger print can be used for unlock the device and it was unexpectedly comfortable to use	8	Smartphone case was officially available that was useful and good designed	8
				L8	Google applications frequently appear at unnecessary moments because of the updated OS	-8	Finger print verification did not work when the fingertip is moist. It was a new unsatisfactory point.	-5
				L9	Because the size of storage of main body was limited, I had to move picture files and other files to SD memory	-5	Shortage of storage in the main body and I had to delete some of the data	-5
				L10	Sometimes camera did not work because of the high temperature when I keep continue taking many pictures	-8	Freezing that requires the reboot sometimes occurred maybe because of the operating system	-8
M	M24	Xperia	expectation	L11	Not many negative aspects but not many positive aspects. Battery life is better than before and I will use this model for the moment	5	Though rarely inconvenient, it is generally easy to use and I'm satisfied on the whole	7
				L12	I will use this model until the contract with the carrier will end up as long as there will be no breakdown or damage	0	I will use this as before for one year term until the next contract update	0
				M1	Because XperiaZ3 had a problem that my voice will be disconnected randomly during the voice communication, I expected that this bug should have been improved	5	I've been using XperiaZ3 and was annoyed by noise during the call, I had much expectation for the new version XperiaXZ	10
				M2	Felt secured as the bug mentioned above was fixed	6	Noise during the call was settled down and I was pleased	10
				M3	Malfunctioning occurred during the use of Google map that the direction is not correctly shown and it irritated me	-5	A bug of unknown origin existed regarding the gyro sensor and the direction on Google map was not correctly shown	-10
				M4	While the direction malfunctioning occurred even for Z3, this bug was fixed by the procedure printed on the official help. But for XZ, bug was not fixed and I couldn't use Google map effectively	-10	High resolution music source of which Z3 could not replay is now played on the device and I purchased some music. But I couldn't distinguish the difference. Anyways, it is a pleasure that high resolution music can be played.	2
				M5	Music replay application called "music" that is only installed for Xperia series can be used as when I was using Z3. I'm satisfied with its usability and the sound quality	1	A social game was released of which I was expecting to play. I realized that the play area is larger on the tablet with 4:3 screen. It is small on XZ with 16:9 screen.	-4

N M 23 u o s - S 0 6 H	recent use	M6	The latest social game of which Z3 could not work smoothly because of the lack of power can now be working good. Because I had a strong feeling to this game, I was satisfied so much	8	
	present time	M7	This model was perfect if there is no malfunctioning on Google map	3	5
	near future	M8	I will change the device when a new series will appear	2	0
	expectation	N1	Renewal of smartphone was exciting but, at the same time, a bit tiresome to learn the new system	2	-2
		N2	Purchased the black body because the blue body was sold out, but it's not a big issue. It's the first time that I purchase the case of purse shaped	0	-2
	early use	N3	Function to shut down the smartphone by closing the case is not important for me	0	-5
		N4	Outside of case was partly broken and the vinyl stuck to other stuff	-10	-4
		N5	There's no drastic change in the UI compared to the past models	-1	2
		N6	Shortcuts of applications can be created as a UI function, but I don't use the shortcut frequently	-1	5
		N7	Sometimes, rainbow colored line appears on the right edge of screen and I can see nothing on the screen	-5	2
		N8	Data capacity was enlarged	5	-2
	recent use	N9	As I peeled off all scratches on the case, there's no problem now	2	2
N10		When I tried to let the smartphone to stand up by using the case, it did not stand straight up because the case was flexible	-1	-2	
N11		Though several problems exist, I accept my device as such	0	2	
N12		At the timing of model change, I don't think it is not important to change to a new model	-2	-2	

References

1. ISO 9241-210:2010. Ergonomics of Human-System Interaction - Human-Centred Design for Interactive Systems (2010)
2. Kurosu, M.: Re-considering the concept of usability. In: Keynote Speech at APCHI2014 Conference (2014)
3. Kurosu, M.: Usability, quality in use and the model of quality characteristics. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9169, pp. 227–237. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20901-2_21
4. Kurosu, M.: Theory of User Engineering. CRC Press, Boca Raton (2016)
5. Kurosu, M., Hashizume, A., Ueno, Y.: User experience evaluation by ERM: experience recollection method. In: Kurosu, M. (ed.) HCI 2018. LNCS, vol. 10901, pp. 138–147. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91238-7_12
6. All About UX. All UX Evaluation Methods. <https://www.allaboutux.org/all-methods>
7. Brooke, J.: SUS: A “quick and dirty” usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996)
8. Sauro, J.: SUPR-Q: a comprehensive measure of the quality of the website user experience. *J. Usability Stud.* **10**(2), 68–86 (2015). <http://uxpajournal.org/supr-q-a-comprehensive-measure-of-the-quality-of-the-website-user-experience/>
9. Benedek, J., Miner, T.: Measuring desirability: new methods for evaluating desirability in a usability lab setting. In: Proceedings of Usability Professional Association (2002)
10. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: a questionnaire to measure perceived hedonic and pragmatic quality (in German). In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer. B.G. Teubner (2003)
11. Tahiti, M., Arhippainen, L.: A proposal of collecting emotions and experiences. In: Interactive Experiences in HCI, vol. 2, pp. 195–198 (2004)
12. Laurans, G., Desmet, P.M.A.: Using self-confrontation to study user experience: a new approach to the dynamic measurement of emotions while interacting with products. In: Desmet, P.M.A., van Erp, J., Karlsson, M. (eds.) Design & Emotion Moves. Cambridge Scholars Publishing (2006)
13. Larson, R., Csikszentmihalyi, M.: The experience sampling method. *New Dir. Methodol. Soc. Behav. Sci.* **15**, 41–56 (1983)
14. Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A.: Method for characterizing daily life experience: the day reconstruction method. In: American Association for the Advancement of Science, pp. 1776–1780 (2004)
15. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: User experience over time: an initial framework. In: CHI 2009 Proceedings, pp. 729–738. ACM (2009)
16. Kurosu, M., Hashizume, A.: TFD (Time Frame Diary)—a new method for obtaining ethnographic information. In: APCHI 2008 Proceedings (2008)
17. von Wilamowitz-Moellendorff, M., Hassenzahl, M., Platz, A.: Dynamics of user experience: how the perceived quality of mobile phones changes over time. In: UX WS NordiCHI 2006, pp. 74–78 (2006)
18. Karapanos, E., Martens, J.-B., Hassenzahl, M.: Reconstructing experiences with iScale. *Int. J. Hum Comput Stud.* **70**, 1–17 (2012)

19. Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Karapanos, E., Sinnela, A.: UX curve: a method for evaluating long-term user experience. *Interact. Comput.* **23**, 473–483 (2011)
20. Kurosu, M.: Is the satisfaction evaluation by UX graph, a cumulative one or recency-based one? (in Japanese). In: Japan Kansei Engineering Society Spring Conference Proceedings (2015)
21. Kurosu, M., Hashizume, A., Ueno, Y., Tomida, T., Suzuki, H.: UX graph and ERM as tools for measuring Kansei experience. In: Kurosu, M. (ed.) *HCI 2016. LNCS*, vol. 9731, pp. 331–339. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39510-4_31