



# User Experience Evaluation by ERM: Experience Recollection Method

Masaaki Kurosu<sup>1(✉)</sup>, Ayako Hashizume<sup>2</sup>, and Yuuki Ueno<sup>3</sup>

<sup>1</sup> The Open University of Japan, Chiba, Japan  
nigrumamet-s23@mbr.nifty.com

<sup>2</sup> Tokyo Metropolitan University, Tokyo, Japan

<sup>3</sup> Otsuka Business Service, Tokyo, Japan

**Abstract.** Evaluation of user experience (UX) is the key for improving the quality of artifact that people use in their everyday life and thus improve their quality of life. Two types of evaluation methods can be classified, i.e. the real-time method (e.g. ESM by Larson & Csikszentmihalyi) and memory-based method (e.g. UX curve by Kujala et al.). The ERM proposed here belongs to the latter and asks users to remind their past experience from the time when they had the expectation for targeted artifact and to rate the degree of satisfaction by  $-10$  to  $+10$  rating scale. Two ways of analysis are proposed, one is classifying the result by the aspect of experience and another is classifying the result on the time phase. Web-based application was developed so that the data collection will become efficient.

**Keywords:** User experience (UX) · Evaluation  
Experience recollection method (ERM) · Rating scale · Satisfaction

## 1 Introduction

Since 1998 when Norman proposed the concept of user experience [20], evaluating the user experience (UX) has become the key activity for improving the artifact, i.e. products and services, that people are using in their everyday life and thus improve the quality of life. In the website of AllAboutUx [1], total of 86 different methods are listed as of June 2017. They include field study methods, laboratory study methods, online study methods and questionnaires to investigate the UX. before usage, as the snapshot, as the episode and as the long-term UX.

This suggests that the UX evaluation methods can be classified into two groups including real-time methods and memory-based methods. The former includes some of field study methods and laboratory study methods in addition to online study methods and questionnaires. And the latter includes the rest of field study methods and laboratory study methods. The real-time method evaluates the experience at the time of evaluation. But by considering the fact that the value of UX may vary depending on the episode that the user may face, it would be better to evaluate the UX, at least, along with a certain period of time or for several times. In this respect, the memory-based method that can evaluate the UX covering a long period of time is

better than the real-time method while there are many reports that the memory sometimes lies and is incomplete since Munsterberg pointed it out in 1908 [19]. Because there are different phases of UX including the expectation before the actual usage, episodes at the time of purchase or short-time after the purchase, episodes that may occur during the long-time usage, and the current impression, the UX evaluation should cover the whole range of experiences for some artifact.

All about UX website [1] includes such developmental phases as the concept, early prototype, functional prototype as well as products on the market, authors think that the real UX can only be evaluated for products on the market and products in use. Foreseeing the UX would be the silver bullet for manufacturers for minimizing the risk of providing unwanted products or services that will give users the poor experience. But what can be called as the experience should be based on the real interaction with the artifact in the past in the real context of use by the real users.

Based on these ideas, the ERM, a new method for evaluating the UX, will be explained in this article.

### 1.1 Satisfaction as the Measure

In ISO9241-11:1998 [5], the concept of usability has three sub-concepts, i.e. the effectiveness, efficiency and satisfaction. In this 1998 version, it was defined as “freedom from discomfort, and positive attitudes towards the use of the product”. But during the discussion for its revision that was held in the spring of 2017, the trial definition of satisfaction was changed to “person’s perceptions and responses that result from the use of a system, product or service.” Interesting thing is that the definition of UX in the same version was “person’s perceptions and responses that result from the use and/or anticipated use of a system, product or service”. In essence, these definitions are almost the same. Although the trial definition of satisfaction was changed to a different sentence thereafter, this fact shows that the concept of satisfaction and UX are quite closely related with each other even among ISO standard professionals

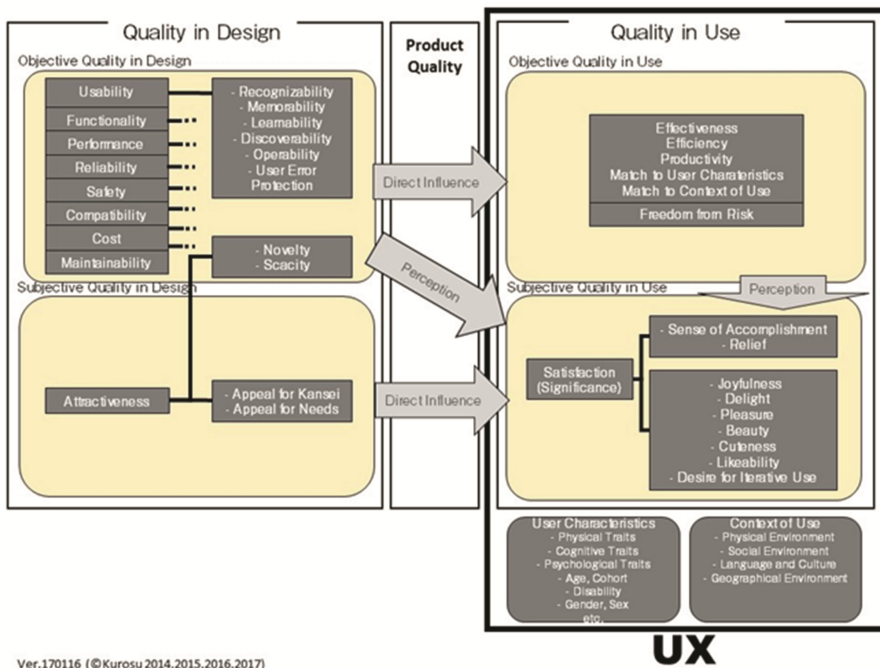
This is related to the idea that the satisfaction is a super concept of various hedonic aspects including the pleasure, joy, delight, beauty, cuteness, etc. The fact that the satisfaction is a topmost concept among hedonic descriptive word and other quality characteristics was also confirmed by the concept dependency analysis [13].

Based on this idea, the concept structure of quality characteristics as shown in Fig. 1 was proposed [14]. This figure is rather complex, but the basic structure is as follows: on the left, there are “qualities in design” of which ISO/IEC25010:2011 [7] referred to as “product quality”. The reason why it is not called the product quality is that the quality characteristics included in these two boxes are the qualities that should be considered and manipulated during the design activity. And the product quality is just the result of this activity. On the right, there are “qualities in use” that include quality characteristics while the real user is using the artifact in the real context of use. This is the reason why the UX is related to the quality in use.

On the upper side of the figure, there are qualities that are overt and can be measured objectively, thus they are named as objective qualities. On the lower side of the figure there are qualities that are covert because they occur in the human mind, thus are named

as subjective qualities. This differentiation is similar to the dichotomy of pragmatic attributes vs. hedonic attributes by Hassenzhal. Based on the combination of these two dimensions, the quality in design vs. quality in use and the objective quality vs. subjective quality, there are four regions of quality characteristics. As can be seen in the lower right area or the subjective quality in use, there is the satisfaction on top of such hedonic characteristics as joyfulness, delight, pleasure, etc.

One more thing that should be noted is that other three areas are linked to the subjective quality in use based on the direct influence or the perception. In other words, the satisfaction is the final and topmost quality characteristics of all. This is the reason why we decided to use the satisfaction as the measure of UX in the ERM.



**Fig. 1.** Four categories of quality characteristics that shows the relationship between the UX and satisfaction. Cited from [14].

### 1.2 Real-Time Methods and Memory-Based Methods

As was suggested before, we classified UX evaluation methods into two categories: real-time methods and memory-based methods.

The real-time methods include such methods as ESM, DRM [9] and TFD [15]. These methods have the merit of being able to measure the non-memory-biased and non-distorted impression regarding the UX. But, they have the demerit of invasiveness to the life of informant. To take an example of ESM, informants will have to stop their everyday

behavior while answering to the phone call. Because of this invasiveness, it is usually said that the adequate maximum span for ESM is 2–3 weeks in the field of clinical psychology.

On the other hand, the memory-based methods including CORPUS, Co-drawing of Chronological Table of Usage [2], iScale [10] and UX curve [12] can evaluate the UX for a long period of time. What is common to these methods is that it is a graph-based method where the abscissa is time and the ordinate is the scale value for evaluation such attributes as the attractiveness, ease of use, etc.

Because it is based on memory, there might be the influence of memory decay and distortion of memory. These deficiencies, however, was typically found in the field of forensic psychology in, for example, the work of Loftus [18]. The case of Prof. von Liszt cited in Munsterberg [19] is the same, though it was in an experimental situation.

But we will have to notice that there is a big difference in the forensic psychological situation and in the UX evaluation situation. In the former situation, what witnesses experience and remember is the external event that, at least initially, nothing to do with them. But in the latter UX situation what users experience and remember is the internal events that were initiated, in most cases, by their own will. In other words, in the UX evaluation, the degree of self-involvement is far stronger than in the case of the witness. Of course, there might be a loss of memory and some distortions. But the degree of this influence is expected to be smaller compared to the case of witness. Furthermore, from some viewpoint, what users report may not always be true. It is paradoxical, but when we obtain the recollection of past experiences of users, we should interpret it as the evidence of current reflection of users' integration of past events and it is the evaluation at the time of survey. In other words, it is the overview of the past events regarding experiences of users from the current standpoint.

With this in mind, we started to improve the memory-based method for evaluating the UX.

### 1.3 Temporal Model of UX

It can be observed every day that the evaluation of our experience on some artifact (product/service) is changing depending on what happens in the use of that artifact and how it happened. In other words, UX cannot be measured as a single value but should be measured as a series of dynamic fluctuation.

In this sense, the temporal model of UX is necessary for the purpose of evaluating the UX as a fundamental conceptual framework. Although UX White Paper [21] proposed the model of UX over time, the model describing the time spans of UX (Fig. 2 in [21]) looks inadequate because the combination of momentary UX during usage and the episodic UX after usage is not shown as a repetition, and because the concept of cumulative UX over time is not adequate considering the peak end rule proposed by Kahneman et al. [8]. Peak end rule tells that people judge the experience based on the most intense point (peak) and at the end of experience (end). That is, the UX is not just the result of simple accumulation of the total experience, hence the concept of cumulative UX is against this conceptual model. We then thought to evaluate the UX during the usage of artifact but not to summate it to get such a value as the cumulative UX.

Regarding the temporal sequence, another figure in the UX White Paper is suggestive. Because of inclusion of such inadequate keywords as cumulative UX and other UXs, we revised the figure to create a new one. Expectation is included in the total sequence of UX because it is the result of mental activity based on the external information such as the catalogue, information on the web, TV commercial, magazine article, information from the friend, etc. in addition to the internal information stored in the memory of user regarding the experience with the similar or previous version of the artifact. Then comes the phase of purchase. In some cases, people get the artifact by not purchasing at the retail store but by getting it from the friend or family member. Anyways, this is the change point when the consumer becomes the user. Consumer behavior theory such as the EBM model [3] usually deals with the phase up to this point. Generally, the experience value at the purchase is positive because of fulfillment of the purchase motivation. But in some case, for example, when the student entered the university that he rated as a second or third level, the satisfaction level may be a bit negative even at the beginning based on the law of level of aspiration [17].

After the purchase or obtaining the artifact, the phase changes to the initial use. In the case of person-to-person service, its lifecycle usually terminates here. During the initial use, some trouble may occur such as to experience the difficulty while installing the product in user's environment. After the initial use, the long-term use follows. During this phase, many types of episodes may occur including positive ones and negative ones. The duration of this phase will vary based on the type of products. Some products will only be used for months (e.g. battery), and some others will be used for a year or more than that (e.g. laptop computer, telephone, washing machine, university education, being hospitalized etc.).

When the performance of the product degrades, the user becomes bored of using it, or the new product catches the eye of the user, and the product will be wasted. This is the end of the product lifecycle.

The UX evaluation should usually be conducted 6 to 12 months after the installment as is described in ISO9241-210:2010 [6]. It is the time when most users get accustomed to the product and the location of that product may be fixed in the life of the user.

We thought that the memory-based UX evaluation should be conducted at 6 to 12 months after the user started to use the product. But sometimes, the length of time may be much shorter or longer after the start in some special cases.

## 2 Previous Trials Based on UX Curve

As the best convenient memory-based UX evaluation method that have been proposed, we focused on the UX curve [12] because of its simplicity and visual impact.

After using this method for more than a year, we made a revision of this method and gave it another name, UX graph [14]. The UX graph was proposed because there were many aspects in the UX curve that need to be improved as follows:

1. Abscissa representing the time is not uniform.

Although abscissa is assigned the time, the original UX curve did not show the unit of time scale and the uniformity of time is not guaranteed. In the UX graph, time scale

by the year is shown so that the arbitrary change of angle of inclination may not influence the wrong impression.

2. Curve was arbitrarily drawn.

Users could draw the curve in any way they liked in the UX curve. But to give more exactness to the graph, we asked users to draw the curve as exact as possible in the UX graph.

3. Coordinate of each episode is not exactly determined.

Because the curve was arbitrarily drawn, the coordinate of each episode point is not exact in the UX curve. For this reason, we asked users to give the value from +10 to -10 so that the coordinate may be more exact in the UX graph.

4. Drawing similar curves for 3 times is tiresome.

In the UX curve, users were asked to draw three similar curves in terms of attractiveness, ease of use, and utility. Another graph is about degree of usage but its structure is different from three graphs described above. Anyways, we found that three curves looked similar in many cases.

Considering the burden of drawing similar curves for three times, users were asked only to draw the graph regarding the level of satisfaction in the UX graph. The reason for this was written in the Sect. 1.1.

5. Expectation phase is not included.

Although the phase of expectation is included in the UX White Paper as one of the UX phases, the UX curve did not ask users to draw the curve from the expectation phase. In the UX graph, they were asked to draw the graph from the expectation phase.

Based on these considerations we proposed the UX graph and also provided the online version [22] that the users can enter the data from their laptops or smartphones.

In the UX graph method, users are asked to describe episodes they encountered in terms of the artifact they are using. Episodes are written in text and the rating scale value from +10 to -10 is also given. Then they are asked to plot the coordinate of episode on the graph sheet by the solid line using the time and scale value as the coordinate. Finally, they are optionally asked to draw the curve of frequency of use by the dashed line.

An example of UX graph on the use of smartphone by a male user aged 43, the whole graph can roughly be split into two parts. In the first part, the user purchased the smartphone for the first time with a certain degree of expectation and he became satisfied as he got accustomed to its use. But as he continued to use, he encountered some software/hardware troubles. As a result, the frequency of use lowered down a bit. And finally, he now thinks to buy another one.

We used the UX graph for about two years to get more than 200 samples and noticed following point: it might be less meaningful of asking users to specify the time when the episode occurred.

Because of the ambiguous nature of memory, it would be less meaningful to ask users to remind of the time when the episode regarding the artifact occurred. Even though the graph is seemingly exact, it should be called as the quasi-exactness for that reason.

And we decided to discard the idea of asking the exact time to informants. It seemed to be enough to roughly distinguish the phases.

Then we came up to the idea of the experience recollection method or simply the ERM.

### 3 Experience Recollection Method (ERM)

We have developed the ERM based on the above described idea. In the ERM, the user will be shown only the blank slots on the paper. The time scale is not shown by the year but by the rough phase including the expectation, impression at the time when the user started to use the artifact, experiential episodes along with the long phase of using the artifact, recent episodes and the prospect for the future. Every time when the user fills in the cell with an episode, they are also requested to give the satisfaction rating from +10 to -10.

The format with example is shown in Fig. 2. Because there is no curve nor the graph, it is difficult to grasp the general tendency at a glance. But when we use this data at the interview session afterwards, it will give us many suggestions and further research questions.

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 2016	<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>	-8
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 2017	<i>There are many restrictions.</i>	2
Estimation in the near future	<i>Would like to study more seriously.</i>	4

Fig. 2. An example of ERM regarding the university education. (translated from Japanese)

What we found in this paper version is: in the paper version, the space for writing the episode is limited to write up as many episodes as were reminded.

Because the first version was printed on the paper, there was not the flexibility for giving additional spaces for writing episodes. Paper version has a merit that the group

data can be obtained in a short period of time. But we had to think of the improvement regarding this space problem by using the software.

### 3.1 Development of On-Line Version

Developing the on-line version, as was in the UX graph, was thought to be the solution to the space issue. In the case of UX graph, the space issue was not much critical because the user can add any episodes after the last line regardless of the time sequence. In the on-line version, episodes are written in chronological order. But it is not necessary for the user to write down the episode in the order of occurrence.

In the ERM, the space issue was crucial because the area for episodes are blocked along with the time phase. On-line version will give us the solution to this issue. Based on the adequate programming, a new line will appear on the screen one after another as necessary. We first developed the Japanese version [23] and released it in the spring of 2017.

### 3.2 Various Application Field

Because the ERM can be applied to any kind of artifact, it can be applied, for example, to the education at university as a service activity. This kind of information will give university staffs the feedback informative for understanding problems that students may have and for improving the educational environment or the curriculum. In one example, the student is highly motivated for the study at the university. But he has complaints about the location of the university that is very far from his home and is suffering from the burden of the home task. Though the second point was rated as negative, it is not a big issue because he has a high level of motivation.

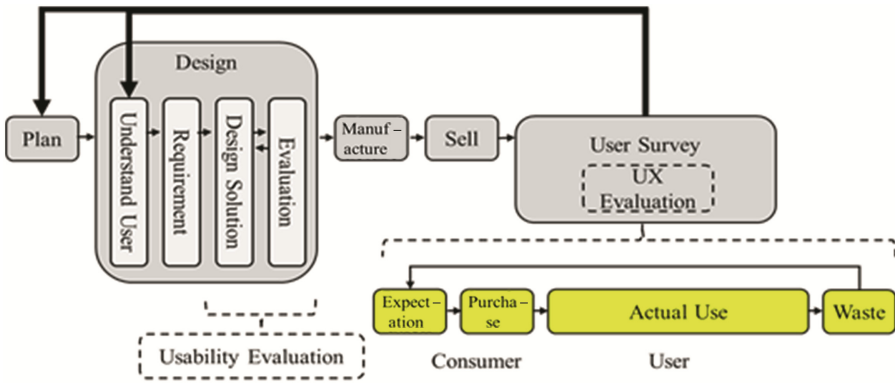
But there may exist some other cases where the students point out fundamental deficiency of the university system that may give university staffs a good opportunity to improve the education if they take the information seriously.

### 3.3 Feedback to the Development

After the ERM (and interview research) is conducted, the information obtained should be categorized in terms of the type of issue (e.g. battery life, waterproof, etc.) considering the phase of occurrence by using KJ method [11] or the affinity diagram. Summarized information should be fed back to the planning section or the technological section. For example, it may be revealed that prolonging the battery life of the smartphone while maintaining the current thickness and weight is technically very difficult, but its importance is quite high. Hence the development of the long-life battery may lead the company to the top position in the smartphone market.

Figure 3 shows this relationship. The upper flow is the industry development process starting from planning, designing, manufacturing and selling. During the designing stage, there are understanding user, specifying the requirement, designing the solution and evaluating the design included.





**Fig. 3.** Relationship of UX evaluation and planning. The upper flow represents industry process and the lower flow represents user experience process.

The user survey on the UX by using, e.g. the ERM, should be conducted after the release of the product/service. The lower flow is the user experience process starting from the expectation, then the purchase and actual use follow, and finally reach the waste. All these user experience process should be evaluated by the UX evaluation method. The information obtained at the UX evaluation should be fed back to the planning section or the design team (heavy line). This feedback is crucially important for reflecting the user experience to the development of next version of the product/service.

#### 4 Conclusion

Based on the consideration on the concept of satisfaction as an integrative measure for the UX, the consideration on the time phase of the UX, we developed the ERM or the Experience Recollection Method. This method is based on our experience with the UX curve and its revised version, UX graph. It was revealed that this method will give us much information on the UX in terms of various artifacts including products (e.g. smartphone) and services (e.g. university education). The result of this method is useful for further interview research and, then, will be effective for giving the adequate feedback to the planning activity of the next version of products/services in their development process.

#### References

1. All About UX. <http://www.allaboutux.org/all-methods>
2. Ando, M.: Dynamic Change of Long-term Usability – Change of Context of Use and Its Influence, pp. 31–45, Sokendai Bunka-kagaku Kenkyu (2007). (in Japanese)
3. Blackwell, R.D., Engel, J.F., Miniard, P.W.: Consumer Behaviour. South Western, Div of Thomson Learning International (2005)
4. Hashizume, A., Ueno, Y., Tomida, T., Suzuki, H., Kurosu, M.: Web tool for the UX graph. In: ISASE 2016 Proceedings (2016)

5. ISO9241-11:1998: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability
6. ISO 9241-210:2010: Ergonomics of Human-System Interaction - Human-Centred Design for Interactive Systems
7. ISO/IEC25010:2011: Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality Models
8. Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A.: When more pain is preferred to less: adding a better end. *Psychol. Sci.* **4**(6), 401–405 (1993)
9. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: User experience over time: an initial framework. In: *ACM SIGCHI 2009 Proceedings*, pp. 729–738 (2009)
10. Karapanos, E., Martens, J.B., Hassenzahl, M.: Reconstructing experiences with iScale. *Int. J. Hum. Comput. Stud.* **70**, 849–865 (2012)
11. Kawakita, J.: Idea Generation Method - for the Development of Creativity. Chuo-koron sha (1967). (in Japanese)
12. 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**(5), 473–483 (2011)
13. 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](https://doi.org/10.1007/978-3-319-20901-2_21)
14. Kurosu, M.: *Theory of User Engineering*. CRC Press, Boca Raton (2016)
15. Kurosu, M., Hashizume, A.: TFD (Time Frame Diary) – a new diary method for obtaining ethnographic information. In: *APCHI 2008 Proceedings* (2008)
16. Larson, R., Csikszentmihalyi, M.: The experience sampling method, In: Reis, H.T. (ed.) *Naturalistic Approaches to Studying Social Interaction, New Directions for Methodology of Social and Behavioral Sciences*, pp. 41–56. Jossey-Bass (1983)
17. Lewin, K., Dembo, T., Festinger, L., Sears, P.S.: Level of aspiration. In: Hunt, J.McV. (ed.) *Personality and the Behavioral Disorder*, vol. 1, pp. 333–378 (1944)
18. Loftus, E.F.: *Memory: Surprising New Insights into How We Remember and Why We Forget*. Addison-Wesley, Reading (1980)
19. Munsterberg, H.: *On the Witness Stand: Essays on Psychology and Crime*, Doubleday, Page & Co., Garden City (1908)
20. Norman, D.A.: *The Invisible Computer – Why Good Products Can Fail, the Personal Computer is So Complex, and Information Appliances are the Solution*, The MIT Press, Cambridge (1998)
21. Roto, V., Law, E.L.-C., Vermeeren, A., Hoonhout, J. (eds.): *User Experience White Paper – Bringing Clarity to the Concept of User Experience* (2011). <http://www.allaboutux.org/uxwhitepaper/>
22. Ueno, Y., Kurosu, M., Hashizume, A.: On-line version of UX graph (2016). (in English). <https://ux-graph.com/ver4/index.php?lang=en>
23. Ueno, Y., Kurosu, M., Hashizume, A.: On-line version of ERM (2017). (in Japanese). <https://ux-graph.com/erm/>