Towards the Multidimensional Measurement of Library Performance. User Experience Index

Markku Antero Laitinen[1], Jouni Henrik Juntumaa[2] and Svetlana Kirichenko[3]

1National Library of Finland. Email: markku.laitinen@helsinki.fi

2City of Helsinki, Culture and Leisure. Email: jouni.juntumaa@hel.fi

3University of Helsinki. Email: svetlana.kirichenko@helsinki.fi

Abstract: The User Experience Index (IUX) is a composite index aimed to measure the overall user experience of services supplied by a library. The user satisfaction is evaluated by calculating the sum of the selected attributes coming from the (scaled) responses to user surveys and weighted by the coefficients assigned by experts. The reason for using the composite index approach is to capture the aggregate effect of multiple factors influencing the overall user experience. The attributes of the IUX were derived from the responses to the National User Survey conducted in Finland over several years in the same format. Altogether seven attributes were selected. The weights of these attributes in relation to each other were defined in a library directors' workshop. The proposed composite index allows the libraries to follow a long-term development of the selected attributes through a single number encompassing their aggregate effect. At the same time, analysis of the individual attributes of the composite index can reveal components of the user experience that require special attention.

Keywords: Composite Index, User Experience Index (IUX), User Satisfaction

1. Introduction

1.1 Evaluation as Part of Knowledge Based Management

Today, the evaluation is perceived as natural part of management of the library and a need to justify the decisions with the help of facts is also widely perceived.

In successful management and decision-making based on knowledge, different points of view, alternatives and sufficient information are taken into consideration. The precondition of knowledge-based management is that the facts describing the phenomenon as well as possible are collected. These facts must be based on systematic analysis of information, and through this, the strengths and subjects of development will be found.

The concept Evidence Based Librarianship (EBL) is often mentioned when talking about information management or knowledge management. There are several definitions and reviews for EBL (Booth 2002, 2006, 2011; Crumley and Koufogiannakis 2002; Eldredge 2002). In these definitions, research derived evidence is emphasized.

As Booth (2002, 2006, and 2011) reported, several obstacles and barriers are encountered by librarians, preventing them from adopting the evaluation as everyday practice. On the other hand, there has been an ongoing discussion about the input of the libraries to the output of the universities and integration of libraries in institutional learning analytics (see e.g. Kiviniemi et al. 2009; Oakleaf 2010, 2018 and Dobbs 2017).

The performance and impact of the library may be estimated as a part of the evaluation of its frame organization (municipality, institute of higher education, research institute etc.). Being this desirable, the evaluation of the library has, of course, to be included in the evaluation system practiced by the frame organization.

So, the library must take an active role in the evaluation so that it, actively, will bring out the knowledge based management which is based on the showing of the impact of the library in both its frame organization and in the society.

In the evaluation of impact, there is the challenge that no specific indicator is assigned for this purpose but instead, the evaluators should find data collected from different sources to be used for the evaluation of the library's impact.

As defined in the International Standard ISO 16439 (Methods and procedures for assessing the impact of libraries), bringing qualitative data (the results of user surveys) next to conventional statistical data about the library, new ways of showing the impact and value of the library may be reached (ISO 16439(E):2014;5.6).

Hence, the International Standard encourages the libraries to develop new types of indicators in the hope of categorizing the qualitative data into classes that may be assigned numeric values (ISO 16439:2014(E);5.6.3) and of utilizing information about user satisfaction to identify the services that may have affected the skills, competence, behaviour or opinions of the users (ISO 16439:2014(E);6.4.3).
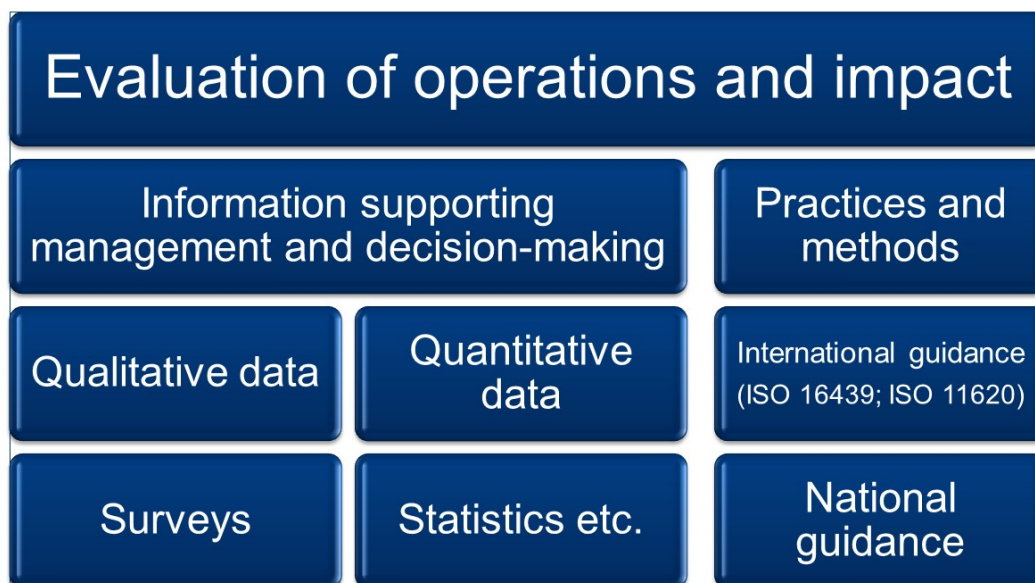

Figure 1 Theoretical framework of evaluation.

The above mentioned encouragement can be dressed as the theoretical framework of evaluation illustrated in Figure 1. Both qualitative and

quantitative data is collected to get information that supports management and decision-making. The outlines what kind of data should be collected for evaluation are given by international guidances like the standards and sometimes also by national instructions. The data collected should cover the analysis of user surveys (qualitative data) and statistical facts (quantitative data) about the issues of the user survey.

1.2 Composite Index for Evaluating a Single Library Service

1.2.1 The Idea of the Composite Index

Many matters to be measured, such as library users' experience (UX) are multidimensional and affected by many factors. Because of this, it is impossible to describe it comprehensively with the help of one simple indicator. A solution for this, might be the composite index. It is constructed by combining several variables, or attributes, together. The idea is to use a linear function for mapping n-dimensional vectors of the attributes, where the function coefficients are derived from experts' opinions on the importance of the attributes.

On the other hand, the presenting of several indicators simultaneously side by side does not necessarily give a clear overview of the topic and there is not always enough time to handle several indicators e.g. in connection with the budget negotiations or other context. Summing up of several indicators as a so-called sum index or composite index provides the opportunity to easily get the general view of the matter which is measured. So, a composite index collects the information of a certain topic area as sensible parcel the indicators of which are more comparable than basic statistics.

1.2.2 Weighting of the Indicators or Attributes

The fact that the composite index consists of a group of separate indicators or attributes intended to measure a multidimensional phenomenon by calculating the sum of the indicators or attributes gives the opportunity to give different weight values to them. Yet, even though not necessary, the weight value of the

individual attributes of a composite index may be different – which is more important when using e.g. a search portal: able to find the information the patron needs or that he enjoys using the service? Consequently, a need to rank the attributes by giving the individual attributes weight coefficients by proportioning each attribute to each other may arise.

Jia (et al. 1998) discussed the aspects of different weighting methods of individual attributes in a simulation study of different attribute weighting techniques. Also, Guitouni and Martel (1998) introduced a comprehensive review on multi criterion decision methods. Kao (2010) discussed the aspects of multiple criteria decision analysis (MCDA). Greco et al. (2019) introduced a recent comprehensive review of the methodological framework of composite indices.

2. Method

2.1 Composite Index – Key Definitions and Choices

Conducting a customer survey is a common way to evaluate library (and other) services. While detailed responses are required for clarifying what to improve in the services, it is also useful to aggregate the customer opinions into a single measure reflecting the overall picture and allowing to easily track how the library customer opinions change over time. We describe here one possible approach: a composite index.

A composite index can be used to measure the overall user experience of the services supplied by a library. To capture the aggregate effect of multiple factors influencing the user experience, we have to take into account:

- Weights of the selected attributes (i.e., evaluation criteria) assigned by experts
- Survey responses by customers

Suppose we conducted a survey with $N > 0$ respondents, and there are m options for answering a question they were asked. Assume the values of the m answers

to the question {x1, x2, …, xm} belong to the interval between Smin and Smax, where Smin is the minimum possible score and Smax is the maximum one.

Let n1, n2, n3, …, nm be the numbers of the respondents who chose the options 1, …, m respectively,

so   N = n1 + n2 + n3 + … + nm   or $N = \sum_{(i=1)}^{m} n_i$

$$(1)$$

Then the Weighted Sum is defined as

$$\text{Weighted Sum} = \sum_{i=1}^{m} x_i * n_i \qquad (2)$$

Let's compute the Weighted Mean S (essentially the arithmetic mean of the responses):

S = Weighted Sum / N $\qquad\qquad\qquad (3)$

In our case, S belongs to the interval [Smin, Smax], and the range L of the responses is:

$$L = S_{max} - S_{min}. \qquad\qquad\qquad (4)$$

So far, we considered only one question from the survey, and usually surveys consist of several questions. To combine the responses to all the survey questions into a single value – composite index – we need to make the response values comparable. Since the response ranges for different questions can be different, we have to normalize them, that is, 'fit' them into the segment [0, 1]:

Normalized Score = (Weighted Mean – minimum score) / range

$$\textbf{Normalized Score} = (S - S_{min}) / L \qquad\qquad (5)$$

Assume our survey consists of k questions, which can be considered k criteria evaluating essential aspects of the library services. The importance of these criteria, that is, their weights, are usually defined by library service experts. There are several popular methods for assigning weights. For example, some of those are presented in [Jia et al. 1998]: "The weighting methods considered include equal weighting of all attributes, two methods for using judgments about the rank ordering of weights, and a method for using judgments about the ratios of weights. The question addressed is: How well does each method perform

when based on judgments of attribute weights that are unbiased but subject to random error?

In our work, we chose the method based on the ratios of weights, as we believe it nicely combines simplicity and flexibility and avoids over-emphasizing attributes with high weights and neglecting attributes with low weights assigned by the experts. We normalize the expert-assigned weights by dividing each of them by their sum, thus, preserving their ratios and making the total weight be equal to 1.

The normalized weights of the attributes and the results from the user survey are combined into a composite index in a natural way: multiplying the normalized score for each question by the corresponding normalized attribute weight and summing all the products together. The Weighted Score for a specific attribute is computed as follows:

$$S\_w = W\_R * (S - S\_min)/L \text{ , where} \tag{6}$$

Sw is the Weighted Score,

WR is the Normalized Weight,

S is the Weighted Mean,

Smin is the Minimum Score

and L is the Range Length.

Then the composite index of user experience I_UX is computed as the sum of the Weighted Scores for all the attributes:

$$I_{UX} = \sum_{j=1}^{k} \left(S_{w_j} \right) \tag{7}$$

where k is the number of the attributes.

As long as the set of the evaluation criteria and their weights stay the same, we can track the composite indexes over time and study the user experience evolution (in a time series analysis fashion).

2.2 Applying the composite index approach to National User Survey

The attributes of the User Experience Index (IUX) were derived from the National User Survey conducted online in the same format in Finland within the period of 3-4 weeks in the years 2016-2021. Altogether seven attributes were included in the survey (Table 1). The weights of these attributes in relation to each other were defined by library directors in the workshop "Building Indicators for the Future" organized for them in order to form a point of view to the weighting of the attributes of the IUX.

Each library director was asked to weight the attributes on a scale from 1 to 100 and mean of the weightings was calculated. The mean weights were then adjusted giving the value 100 to the attribute with highest mean weight and the other attributes were weighted relative to it (Table 1).

Table 1. Attributes of user experience index: the weights given by the library directors (mean weight) and the relative weighs of the attributes scaled from the mean weights.

Name of the attribute	Mean weight	Relative weight
[Service] is useful	90.8	100.00
I'm able to find what I'm looking for on [Service]	90.0	99.1
[Service] is easy to use	78.9	86.9
Using [Service] saves time	74.2	81.7
On a scale from 0-10, how likely is it that you would recommend [Service] to others? (0 being the lowest and 10 being the highest)	71.3	78.4
Using [Service] is enjoyable	56.7	62.4
On a scale from 0-10, how would you rate [Service]? (0 being the lowest and 10 being the highest)	56.3	61.9

The weights of the attributes and the results from the user surveys were then scaled to build a composite index that takes into account both the weight values defined by the library professionals and the library users' experience of the service they received.

3 Results

As a result, a new type of index measuring the users' experience was reached. This composite index allows the libraries to follow a long-term development of the wholeness of several attributes.

Altogether four user surveys were carried out during the examination periods in 2016, 2018, 2019 and 2021. The respondents were asked to take a stand on the claims on given scale that varied depending on the claim. The scores given by the users and numbers of respondents are shown in Table 2.

Table 2. Score averages given by the respondents. Mean values of all user groups and Finna interfaces.

	2016	**2018**	**2019**	**2021**
Useful[1]	4.39	4.42	4.42	4.56
I'm able to find…[1]	3.91	3.94	3.99	4.08
Easy to use[1]	3.92	3.96	4.01	4.06
Saves time[1]	3.96	4.05	4.12	4.24
Want recommend[2]	29.94	43.92	45.62	56.42
Enjoyable[1]	3.87	3.90	3.95	4.08
Rating[3]	8.02	8.10	8.19	8.30
Respondents	**14 520**	**33 443**	**47 687**	**28 395**

[1] Likert scale 1 to 5, [2]Net Promoter Score (NPS), [3]Scale 0 to 10.

The IUX calculation results during the follow-up period are shown in Figure 2. We detected a positive development in all the IUX components of the users' experience related to the Finna search service.
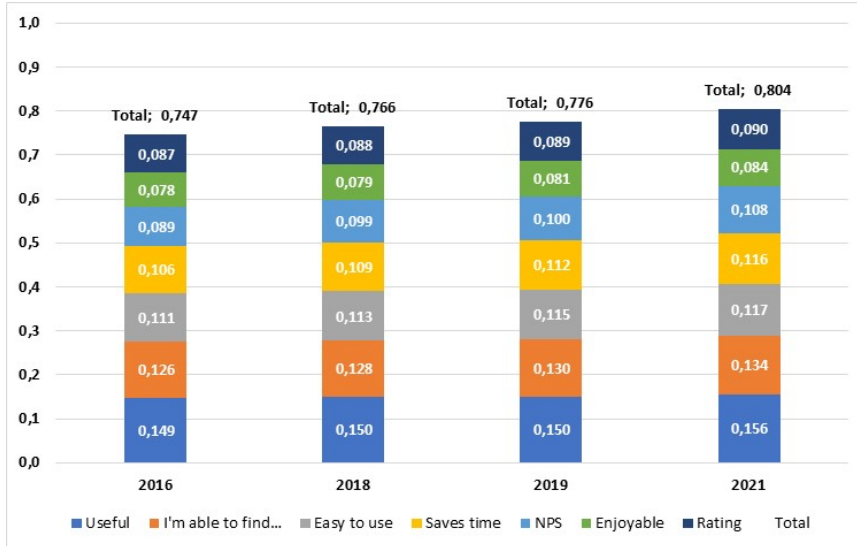
 Figure 2 Finna Search Service's IUX 2016-2021. (All User Groups)

A further analysis using the base index with year 2016 as the base period (i.e., the values in 2016 are taken as 100%) revealed strong growth of willingness to recommend the service to a friend or colleague (NPS). Also the respondents' experience of time saving and ease of finding information as well as experienced comfort of use increased during the follow-up period (Figure 3).
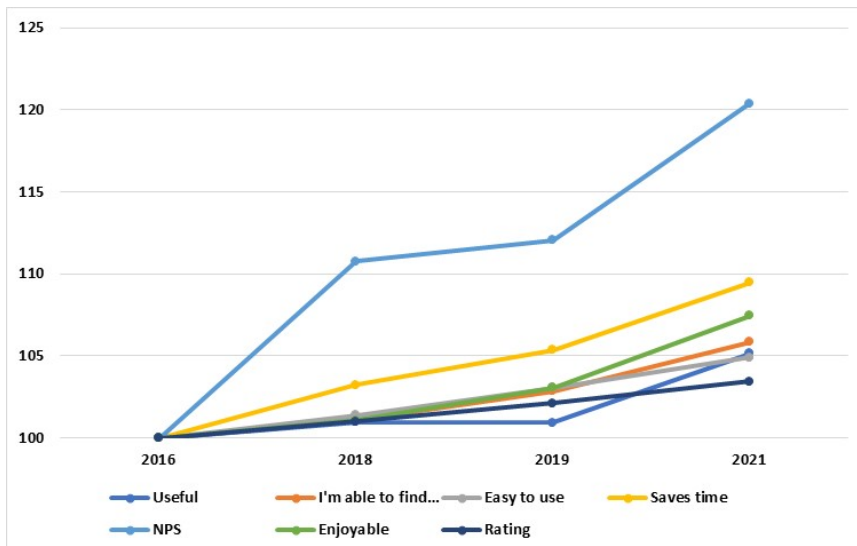
Figure 3 Finna Search Service's IUX in 2016-2021 (All User Groups; base index, the values in 2016 = 100%).

As the Figure 3 indicates, the perceived usefulness and information availability have noticeably risen, the use of the service has also been perceived as more pleasant and the willingness to spread a positive message to others (NPS) has increased.

4. Discussion

A positive development of user experience during the follow-up period was clearly detected. Significantly, the level of any monitored attribute that affects the customer experience has not fallen below the starting point of the measurement during the monitoring period.

The strong growth of the willingness to recommend the service to others (NPS) from 2016 to 2018 may be result of many usability improvements made during the years 2017 and 2018, before the survey of 2018 (Laitinen 2019).

Other strong growths of the components of IUX were detected from 2019 to 2021: NPS, usefulness and enjoyable of use of the service perceived by the users. Also, the perceived or real saving of time seems to have increased during the follow-up period.

In this research, the contribution of learning to use the interface was not investigated, so it cannot be totally ruled out. Whether the significance of growing accustomed to the use the Finna service had an impact on the respondents' estimates for the service, a deeper analysis might be needed.

Being the attributes of the composite index carefully selected, changes in the values of private attributes may reveal the components of user experience needing special attention. For comparability, the weights of the attributes shall be kept constant during several cycles, but they must also be re-evaluated regularly. For the future development and application of the IUX, it may be

appropriate to consider the weights of the attributes of the composite index by each sector of the cultural heritage organizations (institutes of higher education, public libraries, museums and archives).

We have demonstrated the applicability of the composite index to the evaluation of one service wholeness. In our study, we computed the IUX on the top level only, on all user groups together but the results encourage us to a deeper analyse of different user groups. Also, further analysis about the reasons for the strong changes of the focal attributes will be needed by analysing the open answers.

The strength of the IUX is to simultaneously show the effect of several factors affecting the wholeness. The composite index is at its best when applied to evaluate a certain service or a certain library as a whole and long-time follow-up of change.

We believe that the idea of the composite index can be extended also to other functions, such as the measuring the co-effect of different factors of input (economy, staff, premisses) to the library.

References

Booth, Andrew. (2002). From EBM to EBL: Two steps forward or one step back? Medical Reference Services Quarterly, 21(3), 51-64.

Booth, Andrew. (2006). Counting what counts: Performance measurement and evidence-based practice. Performance Measurement and Metrics, 7(2), 63-74.

Booth, Andrew. (2011). Barriers and facilitators to evidence-based library and information practice: An international perspective. Perspectives in international librarianship 2011:1. Retrieved from doi: 10.5339/ pil.2011.1 [accessed 2021-02-05].

Crumley, Ellen and Koufogiannakis, Denise. (2002). Developing evidence-based librarianship: practical steps for implementation. Health Information and Libraries Journal 19, 61-70.

Dobbs, Aaron W. (ed. 2017). The library assessment cookbook. Association of College and Research Libraries, a division of the American Library Association. 15 + 192 p. Chicago.

Eldredge, Jonathan D. (2002). Evidence-based librarianship: What might we expect in the years ahead? Health Information and Libraries Journal, 19, 71-77.

Greco, Salvatore, Ishizaka, Alessio, Tasiou, Menelaos and Torrisi, Gianpiero. (2019): On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. Soc Indic Res., 141, 61–94. https://doi.org/10.1007/s11205-017-1832-9 [accessed 2021-02-05].

Guitouni, Adel and Martel, Jean-Marc. (1998). Tentative guidelines to help choosing an appropriate MCDA method. European Journal of Operational Research, 109, 501-521.

ISO 16439:2014(E) (2014). Information and documentation – Methods and procedures for assessing the impact of libraries. International Standard. 1st ed. International Organization for Standardization (ISO). 82 p. Geneva.

Jia, J., Fischer, G.W. and Dyer, J.S. (1998). Attribute Weighting Methods and Decision Quality in the Presence of Response Error: A Simulation Study. Journal of Behavioral Decision Making 11, 85-105.

Kao, C. (2010): Weight determination for consistently ranking alternatives in multiple criteria decision analysis. Applied Mathematical Modelling 34, 1779–1787.

Kiviniemi, Vesa & Laitinen, Markku & Saarti, Jarmo (2009). Are libraries worth investing in?: Finnish University libraries and their effect on the national economy. Liber 19(1):[8 pp.]. https://www.liberquarterly.eu/articles/10.18352/lq.7952/ [accessed 2021-04-15].

Laitinen, M.A. (2019): Fix, Develop, Keep! Net Promoter Score as a Tool for Improving Customer Experience. Qualitative and Quantitative Methods in Libraries (QQML) 8(2):147-158. Available from: http://www.qqml.net/index.php/qqml/article/view/577 [accessed 2021-04-13].

Oakleaf, Megan (2010). The value of academic libraries: a comprehensive research review and report. Association of College and Research Libraries, a division of the American Library Association. Chicago, Il. 182 p.

Oakleaf, Megan (2018). Library Integration in Institutional Learning Analytics. Syracuse University. 110 p.