CQFAU: Cascading Questionnaire for Feature-Oriented Assessment of Usability

Jakub Swacha
University of Szczecin
Szczecin, Poland *jakub.swacha@usz.edu.pl*

Abstract

The paper proposes CQFAU, a new questionnaire for usability assessment, designed for simplicity yet capable of obtaining information on improvement directions suggested by respondents. It differs from the available questionnaires having similar purpose in its scope (focusing at specific function of the software rather than its general impression), question form (using a set of binary questions followed by a single open one), and the cascading structure (presenting further questions only to respondents who are considered knowledgeable to answer them). The paper describes the instrument, explains how it can be used to identify flaws in software usability and for benchmarking usability, and reports the results of its preliminary evaluation.

Keywords: usability evaluation, usability measurement, usability questionnaire

1. Introduction

1.1. Motivation

Already in 1983, Bailey and Pearson observed the connection between organizational performance and the level of the organization's users' satisfaction with their software systems, proposing an instrument for assessing the latter [1]. At present, with the widespread use of software systems in different areas of life and business, the importance of their usability, and, therefore, also of instruments for its assessment, has grown considerably. While many such instruments have already been proposed, they are often hampered by being difficult to apply, or overly dependent upon evaluators' expertise [2, p. 267]. An even more striking problem with most existing evaluation methods is that they do not specifically identify usability problems, which may result in designers guessing at solutions [3, p. 373]. There is still a place for new usability assessment instruments that would both be simple to apply and provide results usable in practice.

1.2. Problem Setting

Usability evaluation can be based on both subjective (users' opinions or perceptions) and objective data (measuring users' performance, e.g. scenario completion time or rate) [4, p. 1]. The objective evaluation may be the only proper way of evaluating technical components, where test measurements may capture all important aspects of their performance or reliability, but in the case of systems involving humans, their subjective opinion may be more important. For instance, it may matter less that software A performs function X 10% faster than software B if users perceive software A as "slow" and software B as "fast", because the former freezes its whole user interface displaying a static hourglass for the entire execution time of function X, whereas the latter floods users with a constant flow of distracting progress status messages and/or allows users to perform minor operations meanwhile. Moreover, although the answers to "what" and "when" type questions could be obtained in various ways, there is no other way to get an answer to "why" type question than to ask users about their goals and/or motives.

1.3. Approach

The chosen approach is characterized by two properties: the orientation on specific

system features and the use of a cascading questionnaire. While providing an overall usability assessment of a software system is essential for its fair benchmarking, there are several reasons for which focusing the assessment on a specific feature of a system could be useful. For the experience of users, whose job makes them routinely use the system only to perform functions A, B, and C, it is irrelevant how good or bad the system is at functions D, E, and F. The development team that is responsible for functions A, B, and C should not be bothered with usability issues with functions D, E, and F designed by another team (of course, assuming these functions are used independently). Besides, if usability of all key features of a system is high, its overall usability will also be high.

Cascading questions are an efficient way to get more detailed information by filtering down through a hierarchy of questions, thus making it easier to get to the heart of root cause [5]. They allow to avoid asking many irrelevant questions, which not only waste respondents' time, but may also result in their failing to complete a questionnaire. Cascading questionnaires are used in surveys performed in various fields, in Information Systems, they are used, e.g., in the Information Security Awareness Capability Model (ISACM), where they allow to assess the Level 3 situation awareness (projection) only if the Level 2 (comprehension) has been attained, and to the Level 2 situation awareness only if the Level 1 (perception) has been attained [6].

1.4. Contribution

The paper proposes Cascading Questionnaire for Feature-oriented Assessment of Usability (CQFAU, suggested pronunciation: *kfo*) which, unlike the many questionnaires for usability evaluation in use (see section 2), allows to put the usability assessment in the context of a specific feature of a software system and the type of user. For the sake of keeping the survey quick and easy, it is based on cascading binary questions. For practical purposes, the assessment results can be analyzed in their raw form to help identify usability flaws. For research purposes, the synthetic measures ***ruCQFAU*** and ***wcCQFAU*** are defined that can be calculated to help benchmark systems' usability limited to specific features only. The development of CQFAU is still a work in progress, however, the presented preliminary evaluation results confirm its potential.

2. Related Work

As there is an extensive literature on survey-based usability assessment, for the sake of limited space, only the key properties of the ten best-known questionnaires are synthesized in Table 1. Its respective columns provide the acronym under which the given tool is known, its full name, year of introduction (note that some of the tools were updated later, for some, shorter versions have been developed), reference to the primary source introducing the questionnaire with the number of its citations according to Google Scholar (as of 12th April 2024) used as the sort key for the table rows, number of items included in the questionnaire, used measurement scale, and the measured subdimensions.

As Table 1 indicates, there are large differences among the questionnaires in the number of items and, as a consequence, in their level of precision: while short questionnaires consist only general usability questions, e.g., "I found the system unnecessarily complex" in SUS [9] or "The interface of this system is pleasant" in CSUQ [4], the long ones inquiry about more specific aspects of usability, e.g. "The way that system information is presented is clear and understandable" in SUMI [13] or "Is the label location consistent?" in PUTQ [2]. While the answers to the general questions can only provide a bird's eye view of the usability of a specific software, the detailed questions may be non-applicable to systems in which the elements they refer to do not exist (e.g., "Are users allowed to customize windows?" in PUTQ [2]). Nonetheless, the problem with all questionnaires listed in Table 1 is that they do not indicate the reason of specific answer, in particular, they do not point to the feature of the system interacting with which resulted in such a good or bad opinion of the surveyed user. As such, they can be effectively used for benchmarking systems against their prior releases or other systems, yet they are of little assistance in developing a list of concrete improvements to

the system. To address this weakness, not only the selection of questions, but the whole structure of the questionnaire must be changed.

Table 1. Selected existing questionnaires for assessment of usability/UI/UX.

Acronym	Full Name	Year	Source (Citations)	Items	Scale	Dimensions
PUEU	Perceived Usefulness and Ease of Use	1989	[7] (86843)	12	7-point Likert	2: Perceived Usefulness and Perceived Ease of Use
NAU	Nielsen's Attributes of Usability	1993	[8] (26706)	5	5-point Likert	-
SUS	System Usability Scale	1996	[9] (17965)	10	5-point Likert	-
CSUQ	Computer System Usability Questionnaire	1995	[4] (3291)	19	7-point Likert	-
QUIS	Questionnaire for User Interface Satisfaction	1988	[10] (2475)	27	10-point Semantic Differential	5: Overall Reaction to the Software, Screen, Terminology and System Information, Learning, and System Capabilities
UEQ	User Experience Questionnaire	2008	[11] (2259)	26	7-point Semantic Differential	6: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty
USE	Usefulness, Satisfaction, and Ease of Use Questionnaire	2001	[12] (1942)	30	7-point Likert	4: Usefulness, Ease of Use, Ease of Learning, and Satisfaction
SUMI	Software Usability Measurement Inventory	1993	[13] (1190)	50	3-point Likert	5: Efficiency, Affect, Helpfulness, Control, and Learnability
UMUX	Usability Metric for User Experience	2010	[14] (761)	4	7-point Likert	-
PUTQ	Purdue Usability Testing Questionnaire	1997	[2] (552)	100	7-point Likert	8: Compatibility, Consistency, Flexibility, Learnability, Minimal Action, Minimal Memory Load, Perceptual Limitation, User Guidance

3. CQFAU Questionnaire Design

3.1. Development of Questions

As the purpose of the CQFAU is the evaluation of software usability, its key questions aim to obtain respondents' assessment in this regard. Although most of existing questionnaires operationalize this construct with multiple items (from 10 in SUS to 100 in PUTQ, see Table 1), the positive reports on short questionnaires, in particular 4-item UMUX and its even shorter 2-item version UMUX-LITE, showing their ability to predict SUS score with a very high accuracy [15], indicate there is no true need to include many questions.

Similar simplification can be applied to the measurement scale, which ranges in the existing questionnaires from 3 (in SUMI) to 10 (in QUIS). Considering the results of prior work [16], showing that forced binary questions lead to equally reliable results as Likert-scale questions, while being perceived as simpler and saving respondents' time, we decided to use the binary form for 5 out of 6 CQFAU questions.

As the proposed questionnaire is not only to be used for usability benchmarking but also to help identify usability flaws, we decided to include one open-text question allowing the respondents to report their usability improvement ideas (indirectly, the response would reveal the usability issues causing the need for improvement). The open-text question takes less respondent's time to fill in than 100 detailed closed questions (as it is done in PUTQ), of which most are usually irrelevant to the software in question, and

it allows not only to obtain information on issues beyond the predefined list, but also hints on solving the issues, which would not be feasible with a series of closed questions.

An important factor in survey-based research is the respondents' knowledge and their ability to recall the information needed to answer the questions [17, p. 20]. Taking this into consideration, CQFAU begins with four questions (see Table 2) that, respectively, verify the respondent's knowledge of the software (Q1), its user interface (Q2), the function whose usability is assessed (Q3), and measure their level of familiarity with that function (Q4). The actual usability assessment is done with question Q5, whereas question Q6 lets the respondent provide explanation for the assessment.

Table 2. The CQFAU Questions.

	Question	Type	Measures
Q1	I believe there is a function for <action> in <software>	Binary	The user's knowledge of the software
Q2	I can find the function for <action> in <software>	Binary	The user's knowledge of the UI and clarity of the UI
Q3	I know how to use the function for <action> in <software>	Binary	The user's knowledge of the function and the software ease of use
Q4	I used the function for <action> in <software> many times	Binary	How much experience the user had with the function
Q5	I can use the function for <action> in <software> fast and effectively	Binary	How much satisfied the user is with the function usability
Q6	I have the following ideas on how to make the function for <action> in <software> work better	Text	What issues the user has with the software and what are the improvements suggested by the user

3.2. Cascading the Questions

The questionnaire is designed to be administered online. The questions are revealed one by one. The first three questions listed in Table 1 are ordered in such a way that a negative response to one question makes it pointless to ask those that follow it, therefore it immediately ends the survey. Specifically,

- a negative response to Q1 means the user hardly knows the software;
- a negative response to Q2 means the user hardly knows the UI of the software;
- a negative response to Q3 means the user has hardly used the function.

These questions, along with Q4, also allow to segment the respondents into the following classes:

- NON-USERS: those not knowing the software (responded negatively to Q1) – their responses are to be ignored;
- BEGINNER: those knowing the software to some extent but not knowing the function (responded negatively to Q2 or Q3) – their responses can, however, help in discovering problems with the function's visibility in the UI, its intuitiveness and ease of use;
- CASUAL: those using the function rarely (responded negatively to Q4) – their responses can help in discovering problems with the function's usability for occasional users;
- REGULAR: those using the function often (responded positively to Q4) – their responses can help in discovering problems with the function's usability for habitual users.

3.3. Using the Questionnaire for Identifying Usability Flaws

The primary purpose of CQFAU is to identify usability flaws degrading the experience of users using a specific function (or a set of functions) of a software system. The procedure for using CQFAU for this purpose can include the following steps:

1. The general function usability evaluation based on the ratio of positive answers to Q5 only among those respondents who positively answered Q3. Interpretation: A high ratio implies high usability.
2. The regular user function usability evaluation based on the ratio of positive answers to Q5 only among those respondents who positively answered to Q4.

Interpretation: If the result is worse than the general evaluation, it implies that the function does not provide extra capabilities needed only by experienced users.

3. The beginner user function usability evaluation based on the ratio of positive answers to Q3 only among those respondents who positively answered to Q2 and negatively answered to Q4. Interpretation: A low ratio may indicate a barrier for using the function by new users. The problem may be addressed twofold: by changing the way the function operates or by providing hints and/or tutorial guiding the new user step by step on the function use.
4. The UI evaluation based on the ratio of positive answers to Q2 only among those respondents who positively answered to Q1. Interpretation: A low ratio may indicate a problem with UI making the function unnoticeable to new users. The problem may be addressed twofold: by changing the UI layout or by providing hints and/or tutorial guiding the new user step by step on using the UI.
5. The analysis of possible improvements based on the answers to Q6. The users' suggestions should be arranged into four baskets:
    a. Novel ideas for improving the function (not inspired by any obvious usability flaw of its current implementation).
    b. Detected usability flaws (possibly with suggestions for how they should be addressed). These should be considered in the context of the answer to Q4: for instance, an automatic set-up helping casual users can be an annoying obstacle for regular users, whereas a customization needed by the latter can be an unnecessary complication for the former.
    c. Misunderstandings (the specific feature is already implemented in the software, but the user does not know about that or does not know how to make use of it) – if reported by respondents who positively answered to Q4, they clearly indicate a problem with the UI rather than the lack of users' familiarity with the system.
    d. Comments (none of the above, usually to be ignored).

3.4. Using the Questionnaire for Usability Benchmarking

Although not conceived for usability benchmarking, CQFAU can be used for this purpose at the level of a specific function (or a set of them). Thanks to it, one can compare the usability of a given function in various software systems (or various versions of the same system).

Benchmarking needs a synthetic measure. We propose two of them: the Regular-User CQFAU index (ruCQFAU) and the Weighted Compound CQFAU index (wcCQFAU). The ruCQFAU is easier to calculate as it simply measures the ratio of positive answers to Q5 among those respondents who positively answered to Q4 (with Qx^+ denoting the number of positive answers to question Qx and $Qx^+|Qy^+$ denoting the number of positive answers to question Qx from respondents who also answered positively to Qy):

$$ruCQFAU = 10 \cdot Q5^+|Q4^+ / Q4^+ \tag{1}$$

The interpretation of ruCQFAU is how usable the function is for its regular users. The range of values for both indicators spans from 0 (the worst) to 10 (the best).

The second indicator, wcCQFAU, strives to capture all kinds of usability imperfections observed by any respondent class (though with preference for answers obtained from the better informed, as indicated by weights). Its formula is given below:

$$wcCQFAU = 4 \cdot Q5^+|Q4^+ / Q4^+ + 3 \cdot Q5^+|Q4^- / Q4^- + 2 \cdot Q3^+ / Q2^+ + Q2^+ / Q1^+ \tag{2}$$

Note that, in the above formula, we need not to explicitly exclude those respondents who negatively answered to Q2 or Q1 as the cascading nature of the questionnaire prevents them from answering further questions (unlike Q4).

The measurements of wcCQFAU should be compared between evaluations

performed by groups of users having similar levels of knowledgeability. Large differences in $Q3^+ / Q2^+$ and $Q2^+ / Q1^+$ may indicate the opposite is true. In anticipation of such a case, extra questions that measure the knowledge of the software and its UI could be added to the questionnaire to distinguish those that barely know it from those who have issues with finding and/or using its given function (see [17, p. 68]).

4. Preliminary Evaluation

As observed by Kitchenham and Pfleeger, creating a set of questions is only the start of instrument construction, as it is then essential to evaluate it [18, p. 21]. Questionnaire evaluation may cover several different elements, such as checking the questions are understandable, evaluating the reliability and validity of the instrument, or ensuring the responses match the envisaged data analysis techniques [18, p. 21]. Consequently, a full-fledged evaluation requires a lot of time and effort. For the sake of introducing CQFAU, we opted for a preliminary evaluation, limited in scope and the number of involved respondents, so that the results could be obtained quickly. Further evaluation of CQFAU comprises our future work and will be described later.

The preliminary evaluation consisted in a survey asking the evaluators to assess the following aspects of CQFAU with a 5-point Likert scale:

1. Question form quality – their understandability, unambiguity, and precision, with the following evaluation survey items:
    a. The questions are easy to understand
    b. The questions are unambiguous
2. Content validity:
    a. The questions have the right level of detail
    b. No question is superfluous
    c. No question is missing
3. Questionnaire structure – important due to its cascading character:
    a. The order of questions is proper
    b. I like it that the questions are cascading
4. Envisaged areas of use:
    a. Can be useful for designing new software
    b. Can be useful for testing software UI/UX
    c. Can be useful for evaluating existing software UI/UX
    d. Can be useful for finding specific UI/UX flaws of existing software.

Three groups of evaluators were invited to the survey: full-time graduate Business Management students (15 responses received), part-time undergraduate Computer Science students (9 responses received), and part-time undergraduate Information Technology and Econometrics students (6 responses received). Two of the respondents have self-identified themselves as UI/UX Designers, and two others as Project Managers. We can thus consider that the evaluators are mostly target users (users of software to be assessed with CQFAU) with some subject matter experts (knowledgeable in software usability) which is consistent with the recommendations of Kitchenham and Pfleeger [18, p. 22]. The evaluation survey was preceded with two CQFAU-based assessments of functions of software systems well known to the respondents (sorting tables in Microsoft Excel and opening link in a new tab in Google Chrome) so that they could familiarize themselves with the instrument. The evaluation survey was administered online.

The results for the first three evaluation aspects are shown in Figure 1. For all seven questions it covers, there was a visible prevalence of positive responses over the negative ones: the highest for “the order of questions is proper” (67% vs. 10%), the lowest for “the questions are unambiguous” and “no question is superfluous” (43% vs. 20%). As for the extreme answers, 33% strongly agreed that “the questions are easy to understand”, whereas 7% strongly disagreed that “the questions are unambiguous” and “no question is missing”. The problem with ambiguity is probably due to the simplicity of functions assessed in the exemplary assessments as for them, finding a function is essentially

knowing how to use it. In future evaluation, more complicated functions will be considered (such as setting up a pivot table in Microsoft Excel). The simple functions for preliminary evaluation were chosen on purpose, to ensure most respondents proceed to the last question. Most respondents liked the cascading form of the questionnaire (57% vs. 10% thinking otherwise).

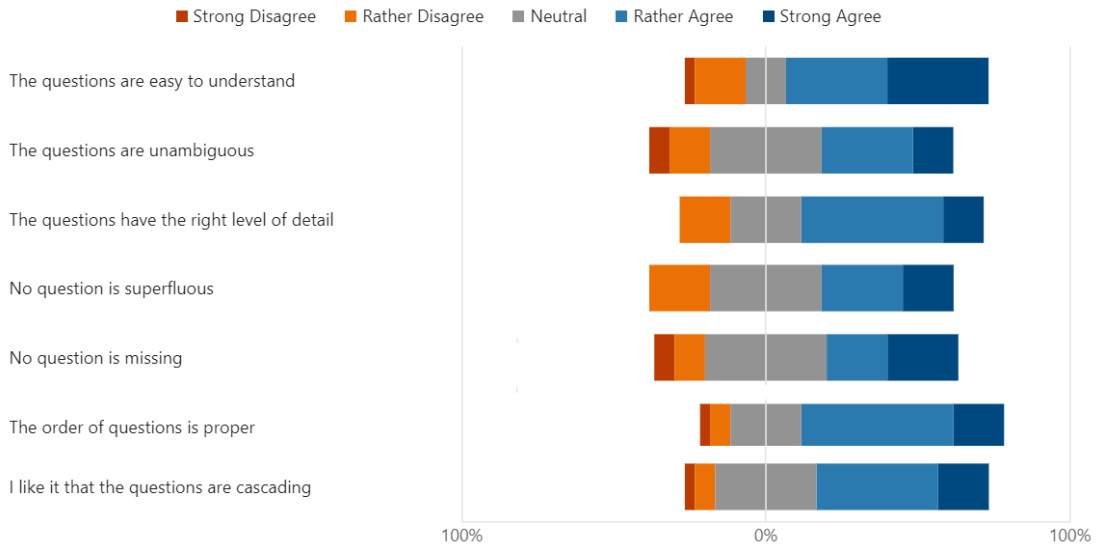

Fig. 1. Evaluation of question form, content validity, and questionnaire structure.

The responses regarding the envisaged areas of use are shown in Figure 2.

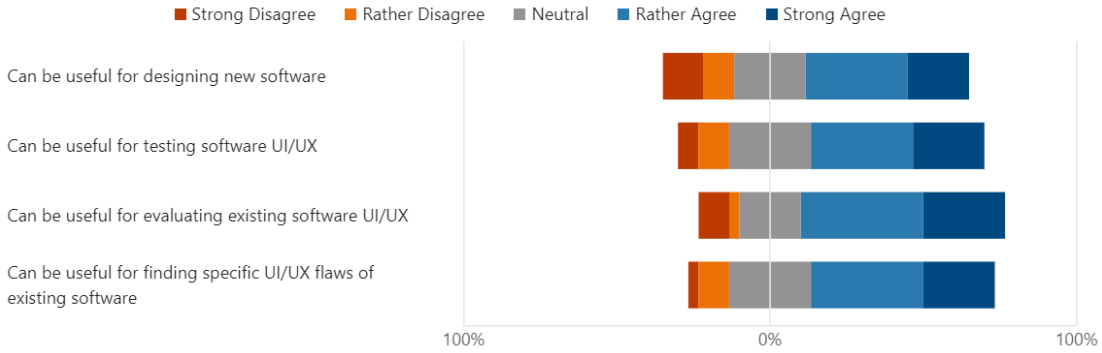

Fig. 2. Evaluation of envisaged areas of use for CQFAU.

More respondents see CQFAU usable in all four listed areas than think otherwise: the prevalence of positive responses over the negative ones was the highest for "can be useful for evaluating existing software UI/UX" (67% vs. 13%) and the lowest for "can be useful for designing new software" (53% vs. 23%). We can thus conclude that the respondents consider CQFAU as a capable tool for its intended purposes.

5. Conclusion

In this paper, we have introduced CQFAU: a Cascading Questionnaire for Feature-Oriented Assessment of Usability. Compared to the existing usability assessment instruments (see section 2), it is designed to assess a specific feature (or their set) of the evaluated software system. It is also distinct in using cascading questions and the binary scale. As for now, only preliminary evaluation of the instrument has been performed. It yielded positive results, which show promise for using the proposed questionnaire. Nonetheless, its limited scope and scale, as well as the selection of respondents, call for further evaluation of CQFAU. Our future work includes the repetition of the same

evaluation survey yet with a larger and more representative group of respondents, having a larger share of subject matter experts, as well as using more complicated functions for exemplary usability assessment. We would like also to check the test-retest reliability of the instrument as well as check its criterion validity by comparing its results to those obtained with other usability assessment questionnaires on the example of software systems featuring one primary function (such as Google search).

References

1. Bailey, J.E., Pearson, S.W.: Development of a tool for measuring and analyzing computer user satisfaction. Management science. 29 (5), 530–545 (1983)
2. Lin, H.X., Choong, Y.-Y., Salvendy, G.: A proposed index of usability: a method for comparing the relative usability of different software systems. Behaviour & information technology. 16 (4–5), 267–277 (1997)
3. Yang, C.-Y.: Website Designer as an Evaluator: A Formative Evaluation Method for Website Interface Development. In: Jacko, J.A. (ed.) Human-Computer Interaction. New Trends. pp. 372–381. Springer, Berlin, Heidelberg (2009)
4. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction. 7 (1), 57–78 (1995)
5. Luck, I.: Why you should use cascading questions in your surveys, https://customergauge.com/blog/cascading-questions-surveys-net-promoter, (2023)
6. Poepjes, R., Lane, M.: An Information Security Awareness Capability Model (ISACM). Presented at the 10th Australian Information Security Management Conference, Perth, Western Australia (2012)
7. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly. 319–340 (1989)
8. Nielsen, J.: Usability engineering. Morgan Kaufmann, San Francisco, CA, USA (1994)
9. Brooke, J.: SUS – A quick and dirty usability scale. Usability evaluation in industry. 189 (194), 4–7 (1996)
10. Chin, J.P., Diehl, V.A., Norman, L.K.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the SIGCHI conference on Human factors in computing systems – CHI ’88. pp. 213–218. ACM Press, Washington, DC, USA (1988)
11. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: HCI and Usability for Education and Work: USAB 2008. pp. 63–76. Springer, Cham, Switzerland (2008)
12. Lund, A.M.: Measuring usability with the USE questionnaire. Usability Interface. 8 (2), 3–6 (2001)
13. Kirakowski, J., Corbett, M.: SUMI: the Software Usability Measurement Inventory. Brit J Educational Tech. 24 (3), 210–212 (1993)
14. Finstad, K.: The usability metric for user experience. Interacting with computers. 22 (5), 323–327 (2010)
15. Borsci, S., Federici, S., Bacci, S., Gnaldi, M., Bartolucci, F.: Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience. International Journal of Human–Computer Interaction. 31 (8), 484–495 (2015)
16. Dolnicar, S., Grün, B., Leisch, F.: Quick, Simple and Reliable: Forced Binary Survey Questions. International Journal of Market Research. 53 (2), 231–252 (2011)
17. Fowler, F.J.: Improving survey questions: design and evaluation. Sage Publications, Thousand Oaks (1995)
18. Kitchenham, B., Pfleeger, S.L.: Principles of survey research part 4: questionnaire evaluation. ACM SIGSOFT Software Engineering Notes. 27 (3), 20–23 (2002)