

How to create short form of UEQ+ based questionnaires?

Martin Schrepp, Heike Sandkühler, Jörg Thomaschewski

The UEQ+ Framework

The UEQ+ is not a UX questionnaire in the classical sense.

It is a collection of (currently) 20 UX scales.

These scales can be combined by UX researchers to create a UX questionnaire that fits perfectly to their research questions.

| Available Scales | | |
|--|--|---|
| Currently the following scales are available. This list may grow in the future. The scales are not independent! Please check the handbook before you create your first questionnaire with the UEQ+. | | |
| Attractiveness Overall impression of the product. Do users like or dislike it? | Perspicuity Is it easy to get familiar with the product and to learn how to use it? | Efficiency Can users solve their tasks without unnecessary effort? Does it react fast? |
| Dependability Does the user feel in control of the interaction? Is it secure and predictable? | Stimulation Is it exciting and motivating to use the product? Is it fun to use? | Novelty Is the design of the product creative? Does it catch the interest of users? |
| Aesthetics Does the product look beautiful and appealing? | Adaptability Can the product be adapted to personal preferences or personal working styles? | Usefulness Does using the product bring advantages? |
| Intuitive use Can the product be used immediately without any training or help? | Value Does the product design look professional and of high quality? | Trustworthiness of Content Is the information provided by the product of good quality and reliable? |
| Quality of Content Is the information provided by the product actual and well-prepared? | Trust Are the users data in safe hands and not misused to harm him or her? | Haptics Feelings which result from touching the product. |
| Acoustics Impact of sounds or operating noise of the product to the user experience. | Clarity Impression towards order, structure and visual complexity of a graphical user interface. | Response behavior Does a voice assistant behave respectful, polite, and trustworthy? |
| Response quality Does the responses of a voice assistant cover the user's information needs? | Comprehensibility Does a voice assistant correctly understand the users instructions and questions using natural language? | |

Benefits of a modular approach

- Which UX aspects are of interest for a product evaluation depends heavily on the special situation.
- Type and main use cases of a product determine which UX aspects are important for overall satisfaction of the users.
- The manufacturer may have some additional needs (marketing requirements) to measure certain UX aspects which are not important from the user perspective.
- The modular approach allows to set-up UX questionnaires that fit perfectly to the questions that should be answered!

Downside: It requires a deeper understanding of the product and UX qualities in general to set up a UEQ+ based study. The UEQ+ is not for a one-shot evaluation, but for situations in which a UX questionnaire should be used for long-term UX monitoring.

UEQ+ scale format

...

In my opinion, handling and using the product are

not understandable ● ● ● ● ● ● ● understandable

difficult to learn ● ● ● ● ● ● ● easy to learn

complicated ● ● ● ● ● ● ● easy

confusing ● ● ● ● ● ● ● clear

I consider the product property described by these terms as

Completely irrelevant ● ● ● ● ● ● ● Very important

In my opinion, handling and working with the product are

not interesting ● ● ● ● ● ● ● interesting

boring ● ● ● ● ● ● ● exiting

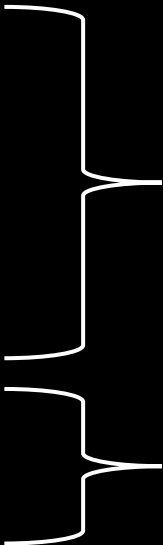
inferior ● ● ● ● ● ● ● valueable

demotivating ● ● ● ● ● ● ● motivating

I consider the product property described by these terms as

Completely irrelevant ● ● ● ● ● ● ● Very important

...



Introductory sentence and
4 Items describing the semantic
meaning of the scale

Rating of the importance of the
scale

Importance rating is used to compute a
KPI as weighted sum over all scales.

Short forms

In certain typical research situations it is required that a UX questionnaire can be filled in a very short time.

Otherwise, participants will simply drop-out.

Examples:

- Questionnaires triggered automatically when a user logs out of a product.
- Experimental studies that require a participant to rate several products or product variants in one session.
- Situations, in which the same participant is asked to provide feedback to the same product repeatedly (for example, every three month).
- If the UX questionnaire is part of a larger product experience questionnaire.

How to shorten UX questionnaires?

The standard method to create short forms of UX questionnaires is to select simply a subset of items.

As a result typically not all scales are measured by the short form, but a smaller number of high-level scales or just an overall score.

Examples:

- VISAWI: 4 scales concerning visual aesthetics in full form, only an overall score in short form.
- UEQ: 6 scales in full form, only two meta-scales (pragmatic/hedonic quality) and an overall value in short form.

Due to the modular approach in the UEQ+ this will not work!

Basic idea for short forms of UEQ+ questionnaires?

There are two potential mechanisms to make UEQ+ based questionnaires shorter:

In my opinion, handling and using the product are

not understandable ● ● ● ● ● ● understandable

difficult to learn ● ● ● ● ● ● easy to learn

~~complicated ● ● ● ● ● ● easy~~

~~confusing ● ● ● ● ● ● clear~~

~~I consider the product property described by these terms as~~

~~Completely irrelevant ● ● ● ● ● ● Very important~~

Drop some items and hope that the remaining items still predict the scale value good enough

Drop the importance question and reuse values from previous studies for KPI calculation

Of course it is possible to drop scales, but since we assume that only those scales are chosen that are required to answer the research questions behind an UEQ+ study, this is in practice not very useful.

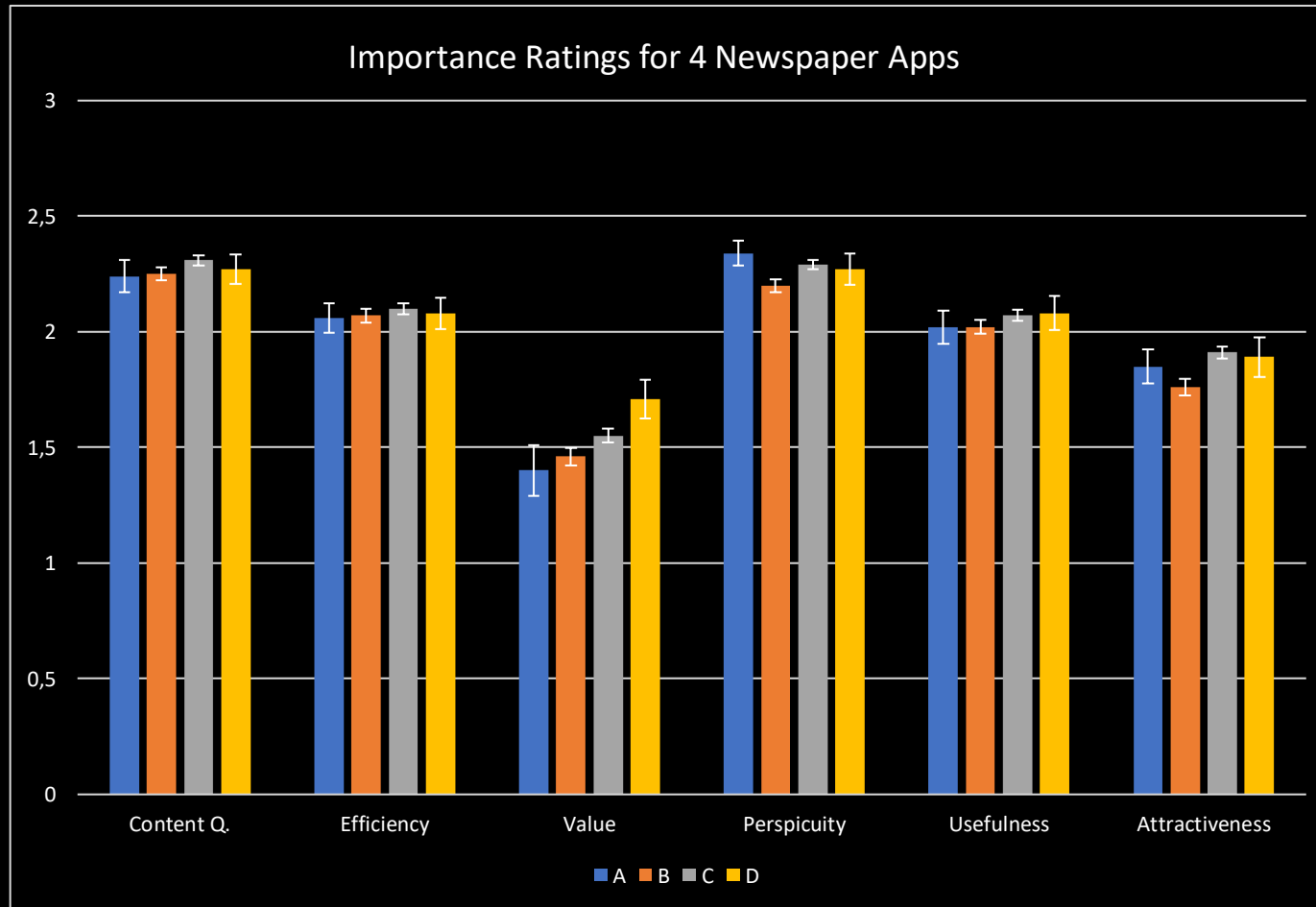
Can we reuse importance ratings?

The rating of a product on a scale can change heavily between two measurements if the product gets better or worse concerning the underlying quality.

But the importance of a scale for the overall UX quality should be quite stable over time (if the main use cases of the product are not changed).

Since the required switch in perspective between rating the 4 items and the importance costs time, it will shorten the time required to fill the questionnaire massively if we can drop the importance rating and reuse values from former studies.

Can we reuse importance ratings?



- Ratings quite similar.
- Importance rating only relevant for KPI calculation.
- If use the importance rating from an arbitrary product to calculate the KPI of the others there would be only minor changes!

A: 1.66 vs. 1.65

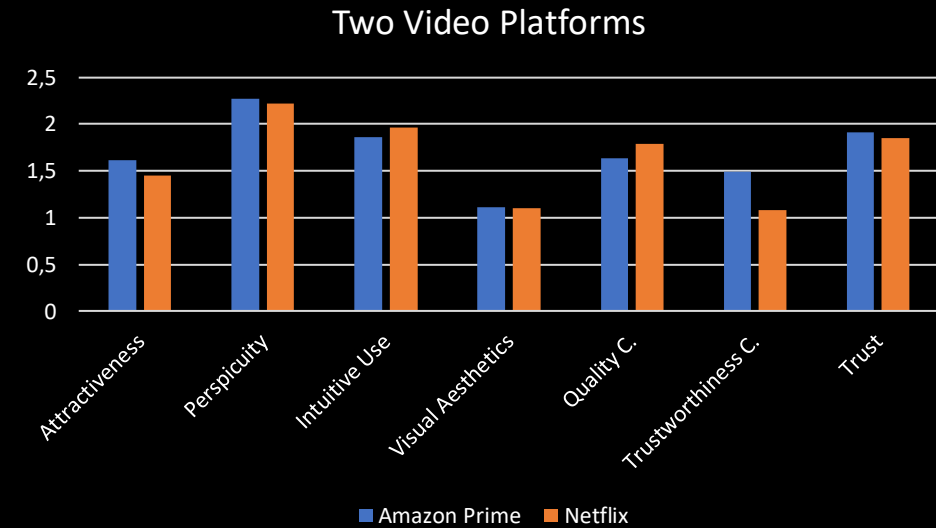
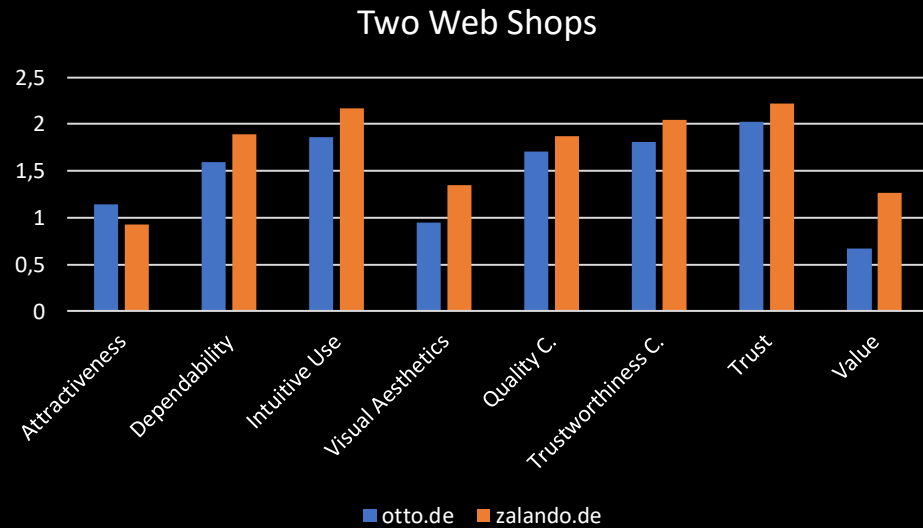
B: 1.58 vs. 1.57

C: 1.82 vs. 1.81

white = real, blue = based on D

We can reuse important ratings from previous studies or nearly identical products!

Can we reuse importance ratings?



It is theoretically plausible that the importance of an UX aspect does mainly depend on the product type and not on the concrete product.

Thus, it may even be possible to reuse importance ratings from products in the same category.

KPI (calculate KPI with importance ratings obtained from a similar product):

- Otto.de 1.27 (1.31 if ratings from Zalando are used), Zalando 1.70 (1.75 if ratings from Otto are used)
- Netflix 1.73 (1.72 if ratings from Amazon are used), Amazon 1.35 (1.33 if ratings from Netflix are used)

Can we estimate the scale means with fewer items?

| Scale | Mean | Combinations of 3 items | | | | Combinations of 2 items | | | | | |
|----------------|------|-------------------------|-------|-------|-------|-------------------------|------|------|------|------|------|
| | | 1,2,3 | 1,2,4 | 1,3,4 | 2,3,4 | 1,2 | 1,3 | 1,4 | 2,3 | 2,4 | 3,4 |
| Content Q. | 5.72 | 5.64 | 5.77 | 5.73 | 5.71 | 5.69 | 5.63 | 5.83 | 5.60 | 5.80 | 5.74 |
| Efficiency | 5.20 | 5.14 | 5.21 | 5.17 | 5.26 | 5.13 | 5.08 | 5.18 | 5.21 | 5.32 | 5.26 |
| Value | 5.48 | 5.63 | 5.47 | 5.39 | 5.44 | 5.69 | 5.56 | 5.33 | 5.63 | 5.40 | 5.27 |
| Perspiciuity | 5.71 | 5.74 | 5.70 | 5.66 | 5.73 | 5.74 | 5.68 | 5.62 | 5.79 | 5.73 | 5.67 |
| Usefulness | 5.75 | 5.77 | 5.79 | 5.75 | 5.71 | 5.82 | 5.77 | 5.80 | 5.70 | 5.74 | 5.68 |
| Attractiveness | 5.53 | 5.54 | 5.51 | 5.51 | 5.56 | 5.52 | 5.51 | 5.48 | 5.58 | 5.55 | 5.54 |

Results for a newspaper app, 1125 participants.

Of course reducing the number of items reduces reliability for small item sets.

But if the target group is big it is possible to reduce the items per scale and still get a good estimation!

A first practical application

| UEQ+ Scale | A (ePaper, 4 items) | B (ePaper, 2 items) | C (Portal, 4 items) | D (Portal, 2 items) |
|----------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Content Qual. | 1.68 (1.11) | 1.56 (1.27) | 0.86 (1.54) | 0.89 (1.32) |
| Efficiency | 1.40 (1.28) | 1.20 (1.46) | 0.30 (1.69) | -0.10 (1.56) |
| Usefulness | 1.90 (1.12) | 1.98 (1.16) | - | - |
| Value | 1.33 (1.26) | 1.47 (1.22) | - | - |
| Perspiciuity | 1.85 (1.20) | 1.78 (1.38) | - | - |
| Adaptability | - | - | -0.36 (1.65) | 0,14 (1.38) |
| Aesthetics | - | - | 0.18 (1.65) | 0.21 (1.53) |
| Intuitive Use | - | - | 0.57 (1.62) | 0.43 (1.66) |
| KPI | 1,66 (0.91) | 1.60 (1.13) | 0.37 (1.28) | 0.33 (1.22) |

Impact of leaving two items out is quite limited!

Given the accuracy required for practical applications this seem justified.

Suggestions to create short UEQ+ versions

The analysis of the available data allows the following suggestions:

- It is possible to reuse the importance ratings from previous studies or from studies concerning similar products.
- The impact of such a reuse of importance ratings on the KPI calculation is limited.
- It is possible to reduce the number of items per scale. However, it may depend on the product which items to drop.
- If you use the UEQ+ for a permanent evaluation of a product, then start with the full scales for the first evaluation, reuse the importance ratings and reduce the number of items for the further evaluations.

Of course further analysis with other data sets are required to get deeper insights.