Impact of a Split into Single Items on the Response Rate of the User Experience Questionnaire Short (UEQ-S)

Marco Schaa[1][a], Jessica Kollmorgen[2][b], Martin Schrepp[3][c] and Jörg Thomaschewski[2][d]
[1]*Allpax GmbH & Co. KG, Papenburg, Germany*
[2]*University of Applied Sciences Emden/Leer, Emden, Germany*
[3]*SAP SE, Walldorf, Germany*

Keywords: Increase Response Rates, User Experience, UX Evaluation, UX Measurement, UX Survey, UX Questionnaire, Single Item Version.

Abstract: Standardized questionnaires are an efficient and reliable method to measure the user experience of a product, service or system. However, response rates of such surveys are often quite low. The length of a survey has an impact on the willingness to respond. We investigate in this paper if it is useful to split a questionnaire into single items to reduce the time needed for participation. In a real web shop on the confirmation page of their order, customers are either asked to answer all the 8 items of the UEQ-S or just one randomly selected item. Results showed that the presentation of single items increased the response rate. The increase was statistically significant, but from a practical point of view not big enough to justify this method. The measured scores for the single items were statistically different for 2 of the 8 items. Thus, some context effects of neighboring items seem to impact the scores in the full UEQ-S version.

1 INTRODUCTION

A continuous user experience (UX) evaluation of products is important to stay competitive in the market. User requirements go beyond pure functionality, i.e. usability, and subjective aspects such as aesthetics or trust should be taken into account depending on the product category (Kollmorgen et al., 2024).

To evaluate the UX of a product, it is necessary to use both qualitative and quantitative methods. As an example of qualitative tests, there are inspection methods which help to generate best practices and ideas for improvement, such as the cognitive walkthrough (Lewis and Wharton, 1997) or heuristic evaluation (Nielsen and Molich, 1990). These methods help to uncover usability problems and generate detailed ideas for improvements. However, especially usability tests are quite expensive and thus limited to small target groups.

These qualitative methods help to gain a deeper understanding of the user and possibly identify unexpected problems. In addition, quantitative evaluation methods are necessary to identify trends and enable a comparison of different products or product versions.

A popular quantitative method to collect user feedback is surveys. They require little effort and can be used to collect feedback from larger user groups. If a survey contains a standardized UX questionnaire, for example the UEQ (Laugwitz et al., 2008), the SUS (Brooke, 1996), or the UMUX (Finstad, 2010), then it produces quantitative results that can be used to compare products or product versions concerning their UX quality. This is useful if the goal is to measure if the UX quality changes over time, for example to check if a redesign of a web shop has a positive impact on its perceived UX.

Surveys can be distributed over different channels, for example, email campaigns, feedback buttons inside a product, or links in posts on social media. Implementing the survey directly in the product has the advantage that is requires little effort for a UX researcher to organize the data collection and ensures a continuous stream of data. In addition, users give feedback while they are using the product, i.e., it is to be expected that the answers are close to the experi-

[a] https://orcid.org/0009-0009-7552-8321
[b] https://orcid.org/0000-0003-0649-3750
[c] https://orcid.org/0000-0001-7855-2524
[d] https://orcid.org/0000-0001-6364-5808

Schaa, M., Kollmorgen, J., Schrepp, M. and Thomaschewski, J.
Impact of a Split into Single Items on the Response Rate of the User Experience Questionnaire Short (UEQ-S).
DOI: 10.5220/0013046400003825

In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024)*, pages 363-371
ISBN: 978-989-758-718-4; ISSN: 2184-3252

enced interaction.

In order to utilize these advantages and achieve continuous product improvements, people are constantly asked to provide feedback over surveys for different products in their limited time of the day. This naturally creates a survey fatigue among the participants and response rates are often quite low (Whelan, 2008).

Therefore, it is practically highly relevant to present a survey in a way that attracts as many users as possible. Some incentives to give feedback can increase the response rate. Nevertheless, here it is likely that the overall data quality decreases, since many participants just take part in the survey to get the incentive and minimize the time they spend to think about their answers. This may result in a low data quality.

Another way to influence response rates is the usage of very short surveys. However, a short survey creates less detailed insights compared with a full UX questionnaire, like the UEQ or SUS. In many situations, especially in e-commerce, there is a high number of potential respondents. Thus, a high response rate is desirable for gaining information on the perceived UX as long as not too much information is lost due to the brevity of the questionnaire. This leads to our following two research questions:

RQ1: Do the item and scale scores collected over single item version differ from the corresponding scores collected with the full UEQ-S version?

We try to estimate the results of the full questionnaire (UEQ-S) from the responses to single items. Thus, it is crucial that the method is able to provide accurate results. Items are always interpreted in a context. This context is mainly defined by the evaluated product, but neighboring items can also influence the interpretation. Therefore, it is important to investigate if the isolated version of single items has an effect on the results.

RQ2: How does a split into single items impact the response rate?

After determining whether the different survey results behave similarly, it is necessary to consider whether the response rate is also influenced by the different display versions. We assume that the response rate increases for the single item version - but how big is the increase and is it big enough to be of practical relevance?

In terms of product quality assurance, it is still the case that the functional requirements of the user experience are primarily taken into account (ISO, 2024).

In this article, we therefore aim to support quality assurance by taking into account the capacities of the participants and at the same time measuring a complete picture of the user experience. We analyze the idea of splitting a longer UX questionnaire into single items to increase the response rate using a real product in the e-commerce sector. Instead of presenting the complete questionnaire, each participant is presented only one randomly selected item. In this way, a high number of answers to single items is collected and these data are used to examine the impacts and differences between the full and single item UEQ-S version, with regard to the comparability of the values and the response rate.

The article structures as follows: Section 2 presents related literature on short and single item questionnaires, as well as the UEQ-S in specific. Section 3 then explains the study conducted with the corresponding evaluation method and study design, the results of which are analyzed in Section 4. Section 5 then discusses and answers the research questions before presenting a summary with limitations and outlook in Section 6.

2 BACKGROUND

In the following, literature is presented which serves as the basis for the study carried out. This refers to single item and short questionnaires.

2.1 Related Work

Time is precious and that is even more true in the field of business and quality assurance of products. Asking the customer to fill in a long form without any incentives will likely result in low response rates. They may also break off the survey or the answers are incomplete caused by participants fatigue. These biases can be reduced by the use of single item surveys, which only present a single item to the user what can be done in a few seconds. However, the possibility of using single item surveys for the measuring of a construct and also archive a high or moderate context validity and reliability depends on the construct itself.

Fuchs et al. (Fuchs and Diamantopoulos, 2009) found that the use of single item measurement can be appropriate under certain conditions. The authors created guidelines for relevant criteria to evaluate single item measurements for use in business administration.

Cuvillier et al. (Cuvillier et al., 2021) came to a similar conclusion using the example of financial institution websites. They investigated functional and non-functional user requirements using the Net Promoter Score (NPS) and the Customer Satisfaction Score (CSAT). They investigated the influence of the context on the hedonic and pragmatic experiences of

the users, whereby it was shown that hedonic features correlated slightly better with the single-item scales.

There are also systematic studies on this topic for other fields, such as Organizational Psychology. Matthews et al. (Matthews et al., 2022) developed corresponding single item questionnaires for 91 questionnaires that have proven effective in this area. The authors analyzed them for criteria and content validity, reliability and usability and also investigated if there is a negative relation between the construct breadth (near vs. broad) and the validity and reliability. Their approach was to develop new, more comprehensive items instead of just splitting the existing surveys into their single item pieces. As a result, they showed that the use of single item measurements does not automatically mean the sacrifice of validity and reliability but has to be considered due to the advantages of short surveys. With this study, the authors provide a ready-to-use catalog of these constructs as single item measures for usage in organizational psychology. They also recommend to develop further single item measures of other constructs they did not include.

Single item measuring in the field of user experience was already investigated for the System Usability Scale (SUS) (Brooke, 1996) in relation to automated driving. Himmels et al. (Himmels et al., 2021) developed a single item version of the SUS, the SIUX scale, and compared the single item version to the multi item measures SUS and UMUX in terms of stability, reliability and validity. Their results supports the hypotheses that the SIUX is even more sensitive to differences in event related user experience than the UMUX and more sensitive to cumulative user experience than the SUS.

The related work shows that single item surveys are not in general a bad research approach, but have to be individually examined for each construct. However, to the best of our knowledge, we have not identified any study that examines the impact of splitting into individual items on response rates. This is a necessary study to support quality assurance while taking into account the challenge of obtaining user feedback. In this article, this analysis will be done for the UEQ-S, the short version of the User Experience Questionnaire (Laugwitz et al., 2008).

2.2 Fundamentals

Questionnaires are a quantitative method for recording the perceived UX of products in general, as they make it easy to gather information. For these scenarios, the UEQ is a standardized questionnaire that allows to measure UX (Laugwitz et al., 2008) concerning several task-related (pragmatic) and non-task-related (hedonic) UX aspects by 26 items. The item format is a semantic differential with a 7-point answer scale (scored from -3 to +3). An example of an item is:

inefficient o o o o o o o efficient

Filling out a complete UEQ can be done in approximately 3 to 5 minutes. Nevertheless, this is still too much effort for a participant in several typical application scenarios, e.g., following an order in a web shop. Therefore, a short version, called UEQ-S, was developed which includes just eight items of the UEQ, corresponding to the two high-level dimensions *Pragmatic Quality* and *Hedonic Quality* (Schrepp et al., 2017). The eight UEQ-S items (English version, for translations see www.ueq-online.org) are:

- obstructive / supportive (UEQ1)
- complicated / easy (UEQ2)
- inefficient / efficient (UEQ3)
- confusing / clear (UEQ4)
- boring / exciting (UEQ5)
- not interesting / interesting (UEQ6)
- conventional / inventive (UEQ7)
- usual / leading edge (UEQ8)

The first four items form the scale *Pragmatic Quality*, the last four items the scale *Hedonic Quality*. The labels in brackets behind an item are used in section 4 to refer to the items.

3 STUDY DESIGN

The following section describes the study design with the web shop used, the implementation of the full UEQ-S and the single item version, the data collection and the evaluation method to answer the research questions.

3.1 Investigated Web Shop

The data was collected from a German online shop. The company distributes various product categories (e.g., vacuum technology, cleaning technology, infrared heaters, gastronomy supplies, laboratory equipment), primarily targeting commercial customers (90%), but also serving private customers (10%). Private customers predominantly focus on gastronomy supplies, specifically household vacuum

sealers, low-temperature cooking devices, and infrared heaters. The online store had around 13,000 items listed at the time of the study.

Due to the relative high number of visitors on the online shop (an average of 117 purchases per day), a representative amount of data could be generated. In addition, the web shop follows data protection guidelines so that the anonymity of the participants could be guaranteed.

3.2 Implementation of the Feedback Mechanism and Data Collection

The UX measurement using the UEQ-S was conducted at the online shop at the end of the checkout process, specifically after a successful purchase and payment, on the checkout success page of the online shop to ensure minimal disruption to the customer's purchasing journey and to avoid potential cart abandonment (see Figures 3 and 4 in the Appendix). During the data collection period from February 2, 2024, to June 2, 2024 (4 months), the questionnaire was presented to a total of 14,469 customers, in average 119 customers per day.

As the utilized shop system is proprietary and closed-source (developed in ASP.NET by 4Sellers, Rain am Lech), the direct integration of a survey module was not implemented due to cost considerations. Instead, a PHP script was developed on a separate Apache server and embedded into the shop's template via an iFrame.

This PHP script generated a random number between 1 and 12 with each request and accordingly displayed either a) a single item or b) the entire UEQ-S. The full UEQ-S was displayed four times more frequently than each individual item (4 of the 12 random numbers pointed to a full UEQ-S and the others to a single item) to ensure an adequate number of complete UEQ-S datasets for comparison purposes. Upon clicking the "Submit Feedback" button, the value(s) of the item(s), along with timestamps of the display and submission were stored in a CSV file on the server.

In addition, each view was recorded in a separate CSV file, including the random ID, timestamp of the view, and customer GUID, to determine the response rate and analyze which questions from the UEQ-S were more or less frequently answered by customers.

Overall, each item (UEQ1-UEQ8) was answered differently often in the full version as well as in the single item version of the UEQ-S. This is due to the fact that there were also participants who only completed part of the questionnaire in the full version and then left the web shop. The number of responses per item for both versions are shown accordingly in Table 1. This resulted in a total of 1,387 data records, of which 413 data records were recorded with the full version and 964 data records with the single item version.

Table 1: Number of answers per item in the full and single item UEQ-S version.

Item	N (full)	N (single)
UEQ1	373	114
UEQ2	407	165
UEQ3	377	149
UEQ4	402	141
UEQ5	338	102
UEQ6	344	110
UEQ7	343	102
UEQ8	342	90

4 RESULTS

The data collected in the web shop was statistically analyzed using Excel. Analyses on statistical key figures including mean values, confidences, reliability and consistency as well as the response rate are presented in this section.

4.1 Difference Between Full and Single Item UEQ-S Scores

The first research question **RQ1: Do the item and scale scores collected over single item version differ from the corresponding scores collected with the full UEQ-S version?** asks if the scale scores from the single item version differ from the results measured with the full UEQ-S version, which is analyzed in the following using statistical key figures.

4.1.1 Mean and Confidence

First, it is examined whether there are differences in the ratings of the web shop between the two versions (single vs. full). Figure 1 shows the scale scores and confidence intervals for the 8 UEQ-S items in both versions.

The detailed scores, standard deviations and confidence intervals are shown in Table 2 for the full version and in Table 3 for the single item version.

To subsequently assess the significance of the individual differences, two-sample t-tests ($p<0.05$) were performed for the items UEQ1-UEQ8 of the two versions. For the items UEQ5 ($t=5.29$, $df=421$, $p<0.0000002$) and UEQ6 ($t=5.96$, $df=429$,

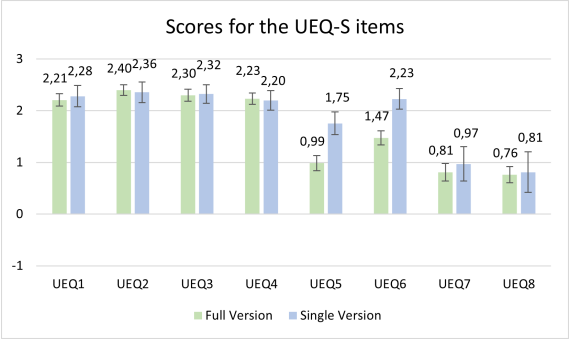

Figure 1: Mean item scores and confidence intervals for the investigated web shop measured with the full and single item UEQ-S version (number of answers per item is shown in Table 1).

Table 2: Means, standard deviations and confidence intervals for the item scores of the **full version** (number of answers per item is shown in Table 1).

Item	Mean	St.Dev.	Confidence Int.
UEQ1	2.21	1.18	2.09 - 2.33
UEQ2	2.40	1.09	2.29 - 2.50
UEQ3	2.30	1.15	2.18 - 2.42
UEQ4	2.23	1.14	2.12 - 2.34
UEQ5	0.99	1.36	0.84 - 1.13
UEQ6	1.47	1.28	1.33 - 1.61
UEQ7	0.81	1.59	0.64 - 0.98
UEQ8	0.76	1.49	0.60 - 0.92

$p<0.000000005$), the score for the single item version is significantly higher than for the full UEQ-S (two-sample t-test assuming equal variances). In addition, the strength of the significant influence was checked using the effect size (Cohen's d). The result was an effect size of d=0.59 for UEQ5 and d=0.61 for UEQ6, so that a medium effect (>0.5, <0.8) was determined. For the other items, the differences determined in the same way are not statistically significant and no effect could be shown.

Furthermore, the mean scale scores for the *Pragmatic Quality* (PQ, items UEQ1-UEQ4), *Hedonic Quality* (HQ, items UEQ5-UEQ8) and the overall score (all items) are displayed in Figure 2 and in detail shown in Table 4. Significant differences with a strong effect were found for both the HQ dimension (t=5.54, df=1769, p=0.000000034; Cohen's d=1.80) and the overall value (t=4.89, df=3888, p=0.000001; Cohen's d=3.17).

4.1.2 Consistency and Reliability

First, the consistency of the dimensions *Pragmatic Quality* and *Hedonic Quality* is examined. For this purpose, the correlation of the items UEQ1-UEQ4 (see Table 5) and UEQ5-UEQ8 (see Table 6) is examined to ensure that they measure the same construct in both the full and single item version.

Table 3: Means, standard deviations and confidence intervals for the item scores of the **single item version** (number of answers per item is shown in Table 1).

Item	Mean	St.Dev.	Confidence Int.
UEQ1	2.28	1.17	2.08 - 2.49
UEQ2	2.36	1.31	2.16 - 2.56
UEQ3	2.32	1.10	2.14 - 2.50
UEQ4	2.12	1.15	2.01 - 2.39
UEQ5	1.76	1.14	1.53 - 1.98
UEQ6	2.23	1.06	2.03 - 2.43
UEQ7	0.97	1.70	0.64 - 1.30
UEQ8	0.81	1.90	0.42 - 1.20

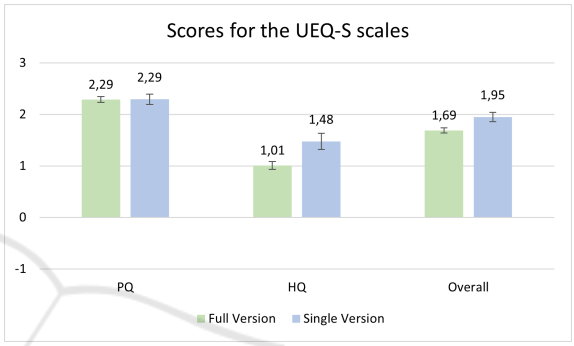

Figure 2: Mean scale scores for the investigated web shop measured with the full and single item UEQ-S version (number of answers per item is shown in Table 1).

It becomes clear that there is a high positive linear correlation (>0.7) for both versions and that the items of the dimension *Pragmatic Quality* have a high consistency (>0.9).

There is also a high positive linear correlation (>0.7) for the items that are assigned to *Hedonic Quality* in both versions, so that a high consistency of the dimension can also be seen here.

Furthermore, the item-by-item correlation is analyzed in the following in order to additionally determine the reliability, meaning whether the individually surveyed items of the single version exhibit similar behavior to the items connected to the full version. This serves to verify the assertion that a change in the rating of an item in the full version also means a change of the rating of the corresponding single displayed item. If, for example, item UEQ2 (complicated/easy) is rated better in the full version after future changes in the web shop, it can be assumed that item UEQ2 will also be rated better in the single item version and vice versa. With a high correlation (>0.7), the items in both versions would therefore exhibit similar behaviour.

As can be seen in Table 7, the items UEQ1-UEQ4

Table 4: Means, standard deviations and confidence intervals for the scale scores (PQ, HQ and overall) of both versions (number of answers per item is shown in Table 1).

Scale (Full)	Mean	St.Dev.	Confidence Int.
PQ	2.31	1.04	2.21 - 2.41
HQ	1.05	1.33	1.92 - 1.18
Overall	1.80	1.07	1.70 - 1.91
Scale (Single)	Mean	St.Dev.	Confidence Int.
PQ	2.29	1.18	2.11 - 2.47
HQ	1.48	1.58	1.23 - 1.72
Overall	1.95	1.42	1.73 - 2.17

Table 5: Correlation coefficient of the items UEQ1-UEQ4 for the *Pragmatic Quality* of the full and single item version (number of answers per item is shown in Table 1).

Full	UEQ1	UEQ2	UEQ3	UEQ4
UEQ1	1			
UEQ2	0.75	1		
UEQ3	0.79	0.87	1	
UEQ4	0.76	0.76	0.77	1
Single	UEQ1	UEQ2	UEQ3	UEQ4
UEQ1	1			
UEQ2	0.94	1		
UEQ3	0.96	0.94	1	
UEQ4	0.91	0.93	0.94	1

of the full and single item version, which belong to the *Pragmatic Quality* dimension, all show a high positive linear correlation to each other, which confirms the reliability of the behaviour of the four pragmatic items.

In contrast, the items UEQ4-UEQ8, which are assigned to the dimension of *Hedonic Quality*, only show a medium positive linear correlation.

4.2 Response Rate

To answer **RQ2: How does a split into single items impact the response rate?**, the response rates (number of participants who submitted an answer divided by the number of participants in the corresponding version) for the single item version and the full UEQ-S version are compared in the following.

Overall, we had 14,469 participants that reached the checkout page during the period of the study. For 4,809 participants, the full UEQ-S was presented and for 9,660, it was only a single item. The overall response rate for the full UEQ-S version was 9% and the response rate for the single item version was 10%. There is a statistically significant dependency (Chi-Square=7.21, df=1, $p<0.0072$) between the display condition (single item versus full UEQ-S) and the willingness to respond. The single item version increased the response rate (two-sample T-test assuming unequal variances, t=2.75, df=10198, p=0.006).

Table 6: Correlation coefficient of the items UEQ5-UEQ8 for the *Hedonic Quality* of the single and full version (number of answers per item is shown in Table 1).

Full	UEQ1	UEQ2	UEQ3	UEQ4
UEQ1	1			
UEQ2	0.79	1		
UEQ3	0.75	0.69	1	
UEQ4	0.74	0.66	0.83	1
Single	UEQ1	UEQ2	UEQ3	UEQ4
UEQ1	1			
UEQ2	0.88	1		
UEQ3	0.94	0.88	1	
UEQ4	0.93	0.89	0.96	1

Table 7: Correlation coefficient of the individual items of the single item version compared to the items of the full version (number of answers per item is shown in Table 1).

Item	Correlation
UEQ1	0.64
UEQ2	0.77
UEQ3	0.78
UEQ4	0.76
UEQ5	0.43
UEQ6	0.55
UEQ7	0.39
UEQ8	0.45

Furthermore, the strength of the significant influence was checked using the effect size (Cohen's d). The result was an effect size of d=1.98, so that a strong effect (>0.8) was proven.

Furthermore, the response rate in the single item version differs between the 8 items (see Table 8).

Table 8: Mean overall and item-related (UEQ1-8, with UEQ1-4 for assessing the pragmatic and UEQ5-8 the hedonic quality) response rates for the full and single item UEQ-S version (number of answers per item is shown in Table 1).

Response Rate	Full	Single
Overall	9%	10%
UEQ1	8%	9%
UEQ2	8%	14%
UEQ3	8%	11%
UEQ4	8%	12%
UEQ5	7%	9%
UEQ6	7%	9%
UEQ7	7%	8%
UEQ8	7%	7%

The pragmatic items (items UEQ1-UEQ4) showed higher response rates than hedonic items (UEQ5-UEQ8). The mean response rate in the single item version for pragmatic items was 12% and for

hedonic items 8%. There is a statistically significant dependency between the UEQ-S subscale (PQ or HQ) and the willingness to answer the item (Chi-Square 29.06, df=1, p<0,00000007). Pragmatic items have a statistical significantly higher response rate than hedonic items (two-sample T-test assuming unequal variances, t=-55.24, df=550, p<1.3e-226). A strong effect (d=1.99) was also calculated here.

Concerning the time required to create an answer (time between showing the order confirmation screen and clicking the submit button), there is also a difference between the two versions. The median time required to answer was 21 seconds for the single item version and 43 seconds for the full UEQ-S version.

5 DISCUSSION

The results are discussed in the following in order to answer the research questions and provide recommendations for action.

5.1 Differences in the Full and Single Item Version

Statistical key figures were calculated in connection with the first research question **RQ1: Do the item and scale scores collected over single item version differ from the corresponding scores collected with the full UEQ-S version?**

It was found that the mean values of the full and single item version do not differ with regard to the four items UEQ1-UEQ4, which belong to the dimension *Pragmatic Quality* (see Section 4.1.1). Thus, the users were satisfied with the pragmatic aspects of their shopping experience in the web store, as the mean values of >2 (scale from -3 to +3) were obtained in both versions.

However, the situation is different for the hedonic items. Here, significant differences with a strong effect were found for the two items UEQ5 (boring/exciting) and UEQ6 (not interesting/interesting), as the mean score for the single item version was higher than for the full version. As a result of these significant differences identified for the two hedonic items, the mean rating of the dimension *Hedonic Quality* and the overall rating of the full and single item version also differed significantly from each other. On the one hand, this can be explained by a context effect, since the answers to these interpretation-free items are selected more carefully in relation to other items, i.e., the answers to the hedonic quality are selected more carefully in connection with the other 6 items. Since the pragmatic items were rated very positively, the hedonic items may be intentionally rated lower in this context in order to illustrate a greater gap in the evaluation. Furthermore, the lack of explanation of the items in the web shop query can lead to a greater scope for subjective interpretation than, for example, for item UEQ7 (usual/leading edge). It is therefore possible that a more detailed description of the single items in the query would have led to different results, like it was done by Matthews et al. (Matthews et al., 2022) (see Section 2.1) by developing new, more descriptive items in their single-item-measurements.

The statistical key figures on similar mean values are additionally analyzed by the consistency and reliability of the version comparison. This was necessary to analyse the sensitivity of the individual constructs (see Section 2.1). The identified high positive linear correlation proves that the four items of the dimensions *Pragmatic Quality* (UEQ1-UEQ4) as well as the items of *Hedonic Quality* (UEQ5-UEQ8) show a high correlation with each other in both versions and thus a high consistency of the dimensions is given. Hence, both the single and full versions measure the same dimensions even with the differently displayed items.

Furthermore, also a high positive correlation for the individual items UEQ1-UEQ4 of the single and full version was shown. Thus, it can be assumed that the individual items of the *Pragmatic Quality* are interpreted similarly by the participants, and that the items measure the same construct. In contrast, the items UEQ5-UEQ8 only show a medium correlation. However, it was shown that these medium correlations are again out of all 8 items only significant for the two hedonic items UEQ5 (boring/exciting) and UEQ6 (not interesting/interesting), which means that there is a connection between the other 6 items. In summary, a high reliability could be demonstrated for the dimension *Pragmatic Quality* and a medium reliability for the *Hedonic Quality*, but this is only relevant for two items.

This shows differences between the context studied and related work. Cullivier et al. (Cuvillier et al., 2021) examined single item formats in relation to financial institution websites. The authors found a slightly increased influence on the hedonic items, which is the opposite in our case in the e-commerce area and in relation to an online store.

Our research showed that both the full and single item version measured very similar scores, and that the consistency of the dimensions remains intact when splitting the UEQ-S into single items. In the future, however, it would be advisable to add more precise descriptions to the semantic differentials in order to avoid possible differences due to room for interpre-

tation.

5.2 Influence on the Response Rate

In connection with the second research question **RQ2: How does a split into single items impact the response rate?**, we also investigated whether there were differences in the response rates for the full and single item version. It was found that the response rate for the single item version was higher for each item and increased by an average of 1% overall.

There were also differences in the increase in response rates for the two dimensions. The response rates for the pragmatic items rose from an average of 8% to 12%. This effect can be explained by the type of web shop surveyed, which supplies goods that customers need for their work, which corresponds to the target group of commercial customers (90%). In this case, shopping is not a fun task, but above all a pragmatic need. Therefore, the hedonic items do not match the customers' expectations as well as the pragmatic ones, which is also reflected in the lower increase in response rates from 7% to 8%.

These results confirmed the assumption that the display of single items provides a significant increase in the response rate compared to the full UEQ-S with a strong effect. Thus, depending on the use case, the use of single items can be recommended, especially with regard to the pragmatic items UEQ1-UEQ4. However, although there is an increase, it is very small. This would certainly be worthwhile for longer questionnaires. However, with a questionnaire as short as the UEQ-S, the split into single items is not recommended, as the gain in information content by using the only slightly longer questionnaire (on average 43 seconds for the Full UEQ-S instead of 21 seconds for a single item) is greater than the time saved.

We were thus able to close the research gap and examined the influence of the split into single items. researchers and practitioners can use this approach as a guide and the study can be transferred to other contexts, for example to examine other product categories or end devices.

6 CONCLUSION AND LIMITATIONS

This article examines whether an increase in the response rate can be achieved by splitting a UX questionnaire (UEQ-S) into individual items.

As it was shown, it is possible to show randomized original single items from the UEQ-S to the customers to measure user experience in web shops without a big negative impact on reliability and consistency. Overall, the measures in this single item display did not differentiate much from the full UEQ-S measurements. However, regarding the comparison between *Pragmatic* and *Hedonic Quality*, there are significant differences between the single item measuring and the full UEQ-S, which can be explained by context effects. In connection with well-rated pragmatic items, the neighboring hedonic items may be rated lower in order to create a differentiation than with individually displayed items.

Furthermore, answering a single item takes about half the time of the full eight-item survey, so maybe the customer is forced to think longer about the solo item, e.g., due to the missing context of the other items.

As the full and single item versions were therefore generally comparable, it was possible to analyse the response rate. An increase in the response rate of 1% was identified for the single item version. Since the UEQ-S is already compact with only eight items in length, the reduction of the information content to the single item version in ordinary web shops is therefore not recommendable. Instead, the willingness of answering a small amount of items is almost as high as answering a single question, but will provide more differentiation on information to the research topic.

However, there might be scenarios where using the randomized single item UEQ-S can be helpful. Since a single item needs little space, this version can be used in very limited screen sizes or even voice assistant systems. Answering only one question instead of eight may make a difference in accepting this kind of customer survey in these cases. This could be investigated in future studies.

As limitation to this study can be seen that the measurement took place after placing an order on the checkout success page in the web shop. Therefore, only customers that were able to navigate within the shop, finding the right product and placing an order in this shop (impact on results in *Pragmatic Quality*) and had the trust into the shop to spend their money were selected to participate in this study, what could explain the overall good results. Furthermore, the analyzed web shop has a distinct customer group as target audience. Due to the mainly B2B customers, there could be differences between the rating by B2B and B2C customers in relation to the UX qualities a web shop must provide. Thus, the results of this study can't be generalized for web shops of all business models.

Like Matthews et al. (Matthews et al., 2022) did for constructs in organizational psychology, a catalog of consistency/reliability proven single item measures

in the field of user experience could be useful. The UEQ+ would be a good base for this future work due to its variable kit architecture, adopting to the most use cases in measuring user experience.

REFERENCES

Brooke, J. (1996). Sus: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*, volume 189, pages 189–194. Taylor & Francis.

Cuvillier, M., Léger, P.-M., and Sénécal, S. (2021). Quantity over quality: Do single-item scales reflect what users truly experienced? In *Computers in Human Behavior Reports*. Elsevier.

Finstad, K. (2010). The usability metric for user experience. In *Interacting with Computers*, volume 22, pages 323–327. Oxford University Press Oxford, UK.

Fuchs, C. and Diamantopoulos, A. (2009). Using single-item measures for construct measurement. pages 195–210. Schaeffer Poeschel Verlag.

Himmels, C., Jarosch, O., Omozik, K., and Buchner, A. (2021). Measuring user experience in automated driving. In *AutomotiveUI '21*, pages 09–14, Leeds, United Kingdom. Association for Computing Machinery.

ISO (2024). Iso 25002:2024(en) systems and software engineering - systems and software quality requirements and evaluation (square) - quality model overview and usage. International Organization for Standardization.

Kollmorgen, J., Hinderks, A., and Thomaschewski, J. (2024). Selecting the appropriate user experience questionnaire and guidance for interpretation: the ueq family. In *International Journal of Interactive Multimedia and Artificial Intelligence*, pages 1–14.

Laugwitz, B., Held, T., and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Holzinger, A., editor, *HCI and Usability for Education and Work*, pages 63–76, Berlin, Heidelberg. Springer Berlin Heidelberg.

Lewis, C. and Wharton, C. (1997). Cognitive walkthroughs. In *Handbook of human-computer interaction*, pages 717–732. Elsevier.

Matthews, R. A., Pineault, L., and Hong, Y. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. In *Journal of Business and Psychology*, pages 639–673. Springer Science+Business Media.

Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 249–256. Association for Computing Machinery.

Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (ueq-s). In *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.* UNIR.

Whelan, T. J. (2008). Antecedents of anonymity perceptions in web-based surveys.

APPENDIX

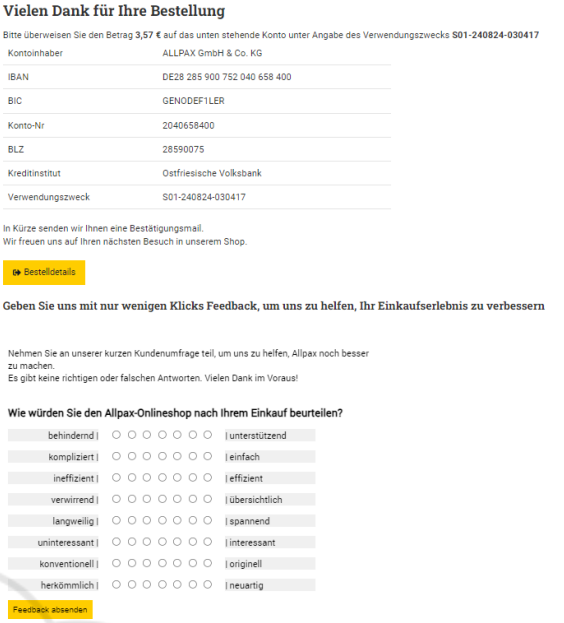

Figure 3: Order confirmation screen in the full UEQS version.

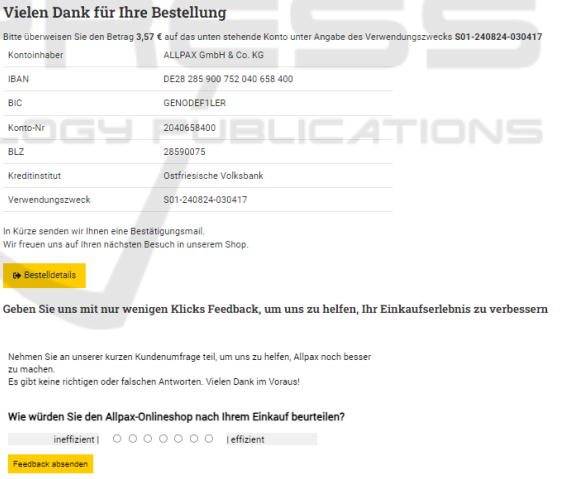

Figure 4: Order confirmation screen in the single item version.