Designing Usable Web Forms – Empirical Evaluation of Web Form Improvement Guidelines

Mirjam Seckler[1], Silvia Heinz[1], Javier A. Bargas-Avila[2], Klaus Opwis[1], Alexandre N. Tuch[1 3]

[1]Department of Psychology
University of Basel, CH
{forename.surname}@unibas.ch

[2]Google / YouTube User
Experience Research, Zurich, CH
javier.bargas@me.com

[3]Dept. of Computer Science
University of Copenhagen, DK
a.tuch@unibas.ch

ABSTRACT

This study reports a controlled eye tracking experiment (N = 65) that shows the combined effectiveness of 20 guidelines to improve interactive online forms when applied to forms found on real company websites. Results indicate that improved web forms lead to faster completion times, fewer form submission trials, and fewer eye movements. Data from subjective questionnaires and interviews further show increased user satisfaction. Overall, our findings highlight the importance for web designers to improve their web forms using UX guidelines.

Author Keywords

Web Forms; Form Guidelines; Form Evaluation; Internet; World Wide Web; Form Interaction

ACM Classification Keywords

H.3.4 Systems and Software: Performance evaluation;
H.5.2 User Interfaces: Evaluation/methodology;
H.5.2 User Interfaces: Interaction styles

INTRODUCTION

Technological development of the Internet has changed its appearance and functionality drastically in the last 15 years. Powerful and flexible technologies have added varying levels of interactivity to the World Wide Web. Despite this evolution, web forms – which offer rather limited and unilateral ways of interaction [14] – remain one of the core interaction elements between users and website owners [29]. These forms are used for registration, subscription services, customer feedback, checkout, to initiate transactions between users and companies, or as data input forms to search or share information [31]. Web forms stand between users and website owners and can therefore be regarded as gatekeepers. Due to this gatekeeper role, any kind of problems and obstacles that users experience during

form filling can lead to increased drop-out rates and data loss. Accordingly, website developers should pay special attention to improving their forms and making them as usable as possible.

In recent years, an increasing number of publications have looked at a broad range of aspects surrounding web form interaction to help developers improve their forms. These studies shed light on selected aspects of web form interaction, but rarely research the form filling process using holistic approaches. Therefore, various authors have gathered together the different sources of knowledge in this field and compiled them as checklists [17] or guidelines [7, 18, 21]. Bargas-Avila and colleagues, for instance, present 20 rules that aim at improving form content, layout, input types, error handling and submission [7]. Currently there is no empirical study that applies these guidelines in a holistic approach to web forms and shows whether there are effects on efficiency, effectiveness and user satisfaction.

It is this gap that we aim to close with the present study. The main research goal is to conduct an empirical experiment to understand whether improving web forms using current guidelines leads to a significant improvement of total user experience. For this we selected a sample of existing web forms from popular news websites, and improved them according to the 20 guidelines presented in Bargas-Avila et al. [7]. In a controlled lab experiment we let participants use original and improved forms, while we measured efficiency, effectiveness and user satisfaction.

This work contributes to the field of HCI in three ways:
(1) The findings of this paper are empirically tested guidelines that can be used by practitioners.
(2) Thanks to the applied multi-method approach, we were able to better understand the impact of the individual guidelines on different aspects of user experience.
(3) Finally, our study shows that there is a difference between how experts estimate the relevance of the individual guidelines for user experience and how these guidelines actually affect the users' experience.

RELATED WORK

An online form contains different elements that provide form filling options to users: for instance text fields, radio-buttons, drop-down menus or checkboxes. Online forms are

used when user input is required (e.g. registration forms, message boards, login dialogues).

The usability of such forms can vary vastly. Small variations in form design can lead to an increase or decrease of interaction speed, errors and/or user satisfaction. It was shown, for instance, that the placement of error messages impacts efficiency, effectiveness and satisfaction. Locations near the erroneous input field lead to better performance than error messages at the top and the bottom of the form – placements that have been shown to be the most wide spread in the Internet [29].

Due to the importance of form usability, there is a growing body of research and guidelines published on how to make online forms more usable. These include topics such as error message improvement [2, 5, 29], error prevention [6, 26], improvement of various form interaction elements [3, 4, 10, 11], improvement for different devices [27], or accessibility improvement [23]. Some publications present empirical data, whereas others are based on best practices of experts in the fields of Human-Computer Interaction and User Experience [18, 19, 31].

There are extensive reviews on form guidelines research such as publications from Nielsen [24], Jarrett and Gaffney [19], and Wroblewsky [31]. One review that focuses particularly on guidelines that are based on published empirical research is provided by Bargas-Avila et al. [7]. Based on their review, the authors derive a set of 20 practical guidelines that can be used to develop usable web forms or improve the usability of existing web forms (see

---

Web Form Design Guidelines

Form content
  1. Let people provide answers in a format that they are familiar with from common situations and keep questions in an intuitive sequence.
  2. If the answer is unambiguous, allow answers in any format.
  3. Keep the form as short and simple as possible and do not ask for unnecessary input.
  4. (a) If possible and reasonable, separate required from optional fields and (b) use color and asterisks to mark required fields.

Form layout
  5. To enable people to fill in a form as quickly as possible, place the labels above the corresponding input fields.
  6. Do not separate a form into more than one column and only ask one question per row.
  7. Match the size of the input fields to the expected length of the answer.

Input types
  8. Use checkboxes, radio buttons or drop-down menus to restrict the number of options and for entries that can easily be mistyped. Also use them if it is not clear to users in advance what kind of answer is expected from them.
  9. Use checkboxes instead of list boxes for multiple selection items.
  10. For up to four options, use radio buttons; when more than four options are required, use a drop-down menu to save screen real estate.
  11. Order options in an intuitive sequence (e.g., weekdays in the sequence Monday, Tuesday, etc.). If no meaningful sequence is possible, order them alphabetically.
  12. (a) For date entries use a drop-down menu when it is crucial to avoid format errors. Use only one input field and place (b) the format requirements with symbols (MM, YYYY) left or inside the text box to achieve faster completion time.

Error handling
  13. If answers are required in a specific format, state this in advance, communicating the imposed rule (format specification) without an additional example.
  14. Error messages should be polite and explain to the user in familiar language that a mistake has occurred. Eventually the error message should apologize for the mistake and it should clearly describe what the mistake is and how it can be corrected.
  15. After an error occurred, never clear the already completed fields.
  16. Always show error messages after the form has been filled and sent. Show them all together embedded in the form.
  17. Error messages must be noticeable at a glance, using color, icons and text to highlight the problem area and must be written in a familiar language, explaining what the error is and how it can be corrected.

Form submission
  18. Disable the submit button as soon as it has been clicked to avoid multiple submissions.
  19. After the form has been sent, show a confirmation site, which expresses thanks for the submission and states what will happen next. Send a similar confirmation by e-mail.
  20. Do not provide reset buttons, as they can be clicked by accident. If used anyway, make them visually distinctive from submit buttons and place them left-aligned with the cancel button on the right of the submit button.

Table 1. 20 guidelines for usable web form design (from Bargas-Avila et al. [7]).

Table 1). The overall application of these guidelines is meant to improve the form's usability, shorten completion times, prevent errors, and enhance overall user satisfaction [7]. To the authors' best knowledge, there has been no empirical evidence that the usage of these guidelines accomplishes the established claims. Therefore a carefully designed experiment was conducted to answer this question.

METHOD

Study Design
In order to investigate as to how forms can be improved by the application of the guidelines compiled by Bargas-Avila et al. [7], we conducted an eye tracking lab study, where participants had to fill in either original or improved versions of three online forms taken from real company websites (between-subject design). Usability was measured by means of objective data such as task completion time, type of errors, effectiveness of corrections as well as eye tracking data (number of fixations, total fixation duration and total time of saccades), but also by subjective ratings on satisfaction, usability, cognitive load and by short interviews about quality of experience.

Participants
Participants were recruited from an internal database, containing people interested in attending studies. In total 65 participants (42 female) took part in the study. Thirty-two were assigned to the original form and 33 to the improved form condition (see below). The mean age of the participants was 27.5 years (SD = 9.7; range = 18-67) and all indicated to be experienced Internet users (M = 5.4, SD = 0.9 with 1 = "no experience"; 7 = "expert"). Participants received about 20$ or course credits as compensation.

Independent sample t-tests showed no significant differences between the two experimental groups regarding age, level of education, computer knowledge, web knowledge, online shopping knowledge and Internet usage. A chi-square test indicated that there are also no significant differences regarding gender distribution.

Selection and Improvement of Web Forms
By screening www.ranking.com for high traffic websites we ensured getting realistic and commonly used web forms to demonstrate that the 20 guidelines work not only for an average website with a form or even for poorly designed forms but also for frequently used ones. We focused on top ranked German-language newspapers and magazines that provide an online registration form (N = 23). We chose high traffic news websites because they often include web forms with the most common input fields (login, password and postal address) and are of decent overall length. Subsequently, we evaluated these forms with the 20 design guidelines provided by Bargas-Avila et al. [7]. Moreover,

Nr.	Guideline	Expert Rating M (range)	Violated by*
15	Never clear the already completed fields.	5.00 (5-5)	-
11	Order options in an intuitive sequence.	4.71 (3-5)	Spiegel (1)
19	Provide a confirmation site.	4.64 (4-5)	-
14	Texting of error messages: (…)	4.57 (3-5)	Suedd (2)
16	Show all error messages after sending the form.	4.29 (3-5)	Spiegel (2), Suedd (2)
20	Do not provide reset buttons.	4.14 (1-5)	NZZ (2)
13	State a specific format in advance.	4.14 (3-5)	Spiegel (1), NZZ (2), Suedd (1)
18	Disable the submit button as soon as it has been clicked.	4.07 (2-5)	Spiegel (2), NZZ (2), Suedd (2)
4a	Separate required from optional fields.	4.07 (2-5)	NZZ (2)
9	Use checkboxes instead of list boxes (…)	3.86 (2-5)	-
8	Use checkboxes, radio buttons or drop-down (…)	3.86 (2-5)	-
3	Do not ask for unnecessary input.	3.86 (1-5)	Spiegel (1), Suedd (1)
1	Let people provide answers in a familiar format.	3.79 (2-5)	-
12a	Date entries (…)	3.57 (2-5)	Suedd (1)
17	Show error messages in red at the right side.	3.57 (2-5)	Spiegel (2), NZZ (2), Suedd (2)
2	If the answer is unambiguous (…)	3.50 (2-5)	-
6	(…) only ask for one input per column.	3.36 (1-5)	Spiegel (2), Suedd (2)
7	Match the size of the input fields (…)	3.29 (2-5)	NZZ (2), Suedd (2)
12b	(…) the year field shoud be twice as long (…)	2.79 (1-5)	Suedd (2)
5	(…) place the lables above the input field	2.71 (1-5)	NZZ (2)
10	Use of radio buttons and drop-down menu: (…)	2.36 (1-4)	Spiegel (2), NZZ (2), Suedd (2)
4b	Use color to mark required fields.	2.21 (1-4)	Spiegel (2), NZZ (2), Suedd (2)

*Note: (1) partial violated, (2) fully violated

Table 2. Expert ratings and guideline violations for each form.

we screened the literature to update this guideline set. As result, we refined guideline 17 [29].

Two raters independently rated for each form whether a guideline was fully, partially or not violated (Cohen's kappa = 0.70). Additionally, 14 HCI experts rated independently each of the 20 guidelines on how serious the consequences of a violation would be for potential users (from 1 = not serious to 5 = serious; Cronbach's α = .90). See Table 2 for these expert ratings.

Based on these two ratings we ranked the forms from good to bad and selected three of different quality: One of rather
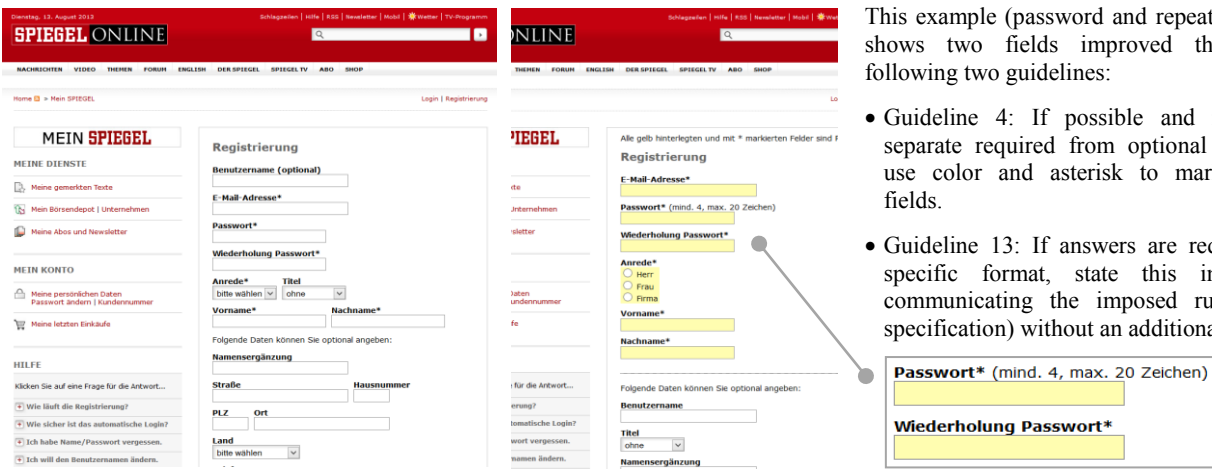
Figure 1. Copy of the original Spiegel™ form (left), improved form (middle), improvement example (right)

This example (password and repeat password) shows two fields improved through the following two guidelines:

- Guideline 4: If possible and reasonable, separate required from optional fields and use color and asterisk to mark required fields.

- Guideline 13: If answers are required in a specific format, state this in advance communicating the imposed rule (format specification) without an additional example.

good quality (Spiegel.de; ranked #11), one of medium quality (nzz.ch; #13) and one of rather poor quality (sueddeutsche.de; #18). Nonetheless, the pool of websites in our ranking is based on top traffic websites – we expect that our three web forms represent rather high quality examples. In total, the NZZ and the Spiegel form violated 9 guidelines each, while the Sueddeutsche form violated 12. See Table 2 for guideline violations for each form.

We refrained from selecting any form from the top third (rank 1 to 8), since these forms had only minor violations and hence showed little to no potential for improvement. By means of reverse engineering of the structure, function and operation, we built a copy of the original form and an improved version according to the 20 guidelines (see Figure 1 for an example). We refrained from applying guideline 3 ("Keep the form as short and simple as possible and do not ask for unnecessary input") in this study, as this would have required in-depth knowledge of the companies' business strategies and goals.
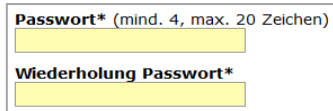
Measurements
Usability was assessed by means of user performance and subjective ratings. User performance included: time efficiency (task completion time, number of fixations, total fixation duration and total time of saccades) and effectiveness of corrections (number of trials to submit a form, error types). Furthermore, we used the KLM Form Analyzer Tool [20] to compare the different form versions. Eye tracking data were collected with a SMI RED eye tracker using Experiment Center 3.2.17 software, sampling rate = 60 Hz, data analysis using BeGaze 3.2.28.

We used the following subjective ratings: The NASA Task Load Index (TLX) for mental workload [15], the System Usability Scale (SUS) [8] and After Scenario Questionnaire (ASQ) [22] for perceived usability in general, and the Form Usability Scale (FUS) [1] for perceived form usability. Moreover, we conducted a post-test interview consisting of two questions: (1) "What did you like about the form?" and (2) "What did you perceive as annoying about the form?".

As the FUS is not a published questionnaire yet, this is a short introduction. The FUS is a validated questionnaire for measuring the usability of online forms [1]. It consists of 9 items each to be rated on a Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree). The total FUS score is obtained by computing the mean of all items. Items: (1) I perceived the length of the form to be appropriate. (2) I was able to fill in the form quickly. (3) I perceived the order of the questions in the form as logical. (4) Mandatory fields were clearly visible in the form. (5) I always knew which information was expected of me. (6) I knew at every input which rules I had to stick to (e.g. possible answer length, password requirements). (7) In the event of a problem, I was instructed by an error message how to solve the problem. (8) The purpose and use of the form was clear. (9) In general I am satisfied with the form.

Procedure
At the beginning, participants had to fill in a practice trial form. The quality of this form was medium (rank #14; Computerbase.de). Afterwards, participants were randomly assigned to one of the experimental conditions (original vs. improved). Participants were then sent to a landing page with general information about the selected newspapers and a link to the registration form. They were told to follow that link and to register. After successful completion of the form, participants rated the form with a set of questionnaires. This procedure was repeated for each online form. At the end participants were interviewed on how they experienced the interaction with the forms. The study investigator asked first for positive (what was pleasing) experiences and the participants could answer for as long as they wanted. Then they were asked for negative experiences (what was annoying).

RESULTS

For all statistical tests an alpha level of .05 was used. Moreover, all data were checked to ensure that they met the requirements for the statistical tests. All time metrics had to be log-transformed to achieve normal distribution.

User Performance

Number of form submission

As expected, users performed better with the improved version of the forms. In all three forms they needed fewer trials to successfully submit the form: Suddeutsche (χ^2 = 11.20, $p < .001$), NZZ (χ^2 = 12.93, $p < .001$), and Spiegel (χ^2 = 3.29, $p = .035$). See Table 3 for corresponding data.

Form	Trials	Original	Improved
Sueddeutsche	1	10	24
	≥ 2	22	9
NZZ	1	9	24
	≥ 2	23	9
Spiegel	1	22	28
	≥ 2	11	4

Table 3. Number of trials until form was successfully submitted.

Initial errors

Descriptive data showed that errors due to missing format rules specifications were frequent for the NZZ form (see Table 4). Chi-square tests showed that this error type was significantly more prevalent for the original condition than all other error types for NZZ ($\chi2$ = 7.17, $p = .007$). For the two other forms, no significant differences between the different error types and conditions were found.

Error types	Original	Improved
Missing specification	17	2
Field left blank	1	1
Captcha wrong	0	2
Mistyping	1	4
Error combination	4	0

Table 4. Initial errors for the NZZ form.

Consecutive errors

Significant differences for errors made after the form has been validated once (consecutive errors, see Bargas-Avila et al. [5]) were found for the two conditions of Sueddeutsche, $p = .033$ (Fisher's exact test). Descriptive data showed that in the original condition participants often ignored the error messages and resubmitted the form without corrections (see Table 5). No significant differences between error types were found for the two other forms.

Error types	Original	Improved
No corrections	14	0
No input	0	1

Table 5. Consecutive errors for the Sueddeutsche form.

Task completion time

As a consequence of the number of submissions, improved versions of all forms also performed better regarding task completion time than their original counterpart (see Table 6). An independent sample t-test showed significant differences for NZZ (t(63) = 4.39, $p < .001$, Cohen's d = 1.00) and for Sueddeutsche (t(63)= 3.91, $p < .001$, Cohen's d = .93). No significant effect was found for Spiegel (t(63)= 1.23, $p < .111$, Cohen's d = .38).

Form	Condition	N	M (SD)	Time improvement
Suedd.	original	32	113 (36)	
	improved	33	85 (25)	- 25%
NZZ	original	32	105 (46)	
	improved	33	70 (20)	- 33%
Spiegel	original	32	104 (66)	
	improved	33	85 (30)	- 18%

Note: Reported values are not log-transformed; statistical tests are based on log-transformed data.

Table 6. Average task completion time in seconds.

To further compare task completion times of the two form conditions, we checked the two forms with the Keystroke Level Model (KLM) [9]. We used the KLM Form Analyzer Tool from Karousos et al. [20] with the default settings except for running the analysis with the option "average typist". For all improved forms the KLM predicted time was lower than for the original forms (see Table 7). Nonetheless, participants in our study needed more time than predicted by the KLM analyzer.

Form	KLM predicted time (sec)		Improvement
	original	improved	
Suedd.	68	52	-23%
NZZ	53	49	-8%
Spiegel	91	84	-7%

Table 7. KLM form analyzer predicted time.

Eye Tracking

The eye tracking data were analyzed using non-parametric Mann-Whitney U tests, as data were not normally distributed. The data shown in Table 8 support results found with the user performance data. Participants assigned to the improved form condition were able to fill in the form more efficiently and needed significantly fewer fixations for the first view time (load until first submission) for Sueddeutsche and NZZ, but not for the Spiegel form: Sueddeutsche (Z = 2.57, $p < .005$, r = .322), NZZ (Z = 4.10, $p < .001$, r = .525), Spiegel (Z = 1.50, $p = .067$, r = .192). The total amount of time participants spent fixating a form before the first submission was shorter in the improved condition, indicating that they needed less time to process the information on screen. Total fixation duration was significantly shorter for Sueddeutsche (Z = 1.71, $p = .044$, r = .214) and NZZ (Z = 3.29, $p < .001$, r = .421). No significance difference could be shown for Spiegel (Z = 0.59, $p = .277$, r = .076).

Form	Number of fixations M (SD)	Fixation duration in sec M (SD)	Saccades total time in sec M (SD)
Suedd. orig. (N=31)	157 (54)	62 (23)	7 (6)
Suedd. improv. (N=33)	126 (41)	53 (18)	4 (3)
NZZ orig. (N=30)	155 (70)	62 (28)	9 (9)
NZZ improv. (N=31)	96 (37)	41 (15)	4 (3)
Spiegel orig. (N=30)	146 (70)	58 (34)	6 (4)
Spiegel improv. (N=31)	121 (43)	50 (20)	5 (4)

Table 8. Eye tracking measures for the original and the improved condition by form.

Analyzing the total time of saccades shows that participants in the original form of the Sueddeutsche ($Z = 2.20$, $p = .014$, $r = .275$) and the NZZ form ($Z = 3.88$, $p < .001$, $r = .497$) spent more time searching for information. For the Spiegel form no significant differences could be shown ($Z = 1.18$, $p = .119$, $r = .151$). Figures 2 and 3 visualize scan paths of participants in the original and the improved condition (duration 38 seconds). The participants filling in the improved form show a much straightforward scan path without unnecessary fixations whereas the side-by-side layout with left-aligned labels of the original form provoked longer saccades and more fixations for participants to orient themselves.
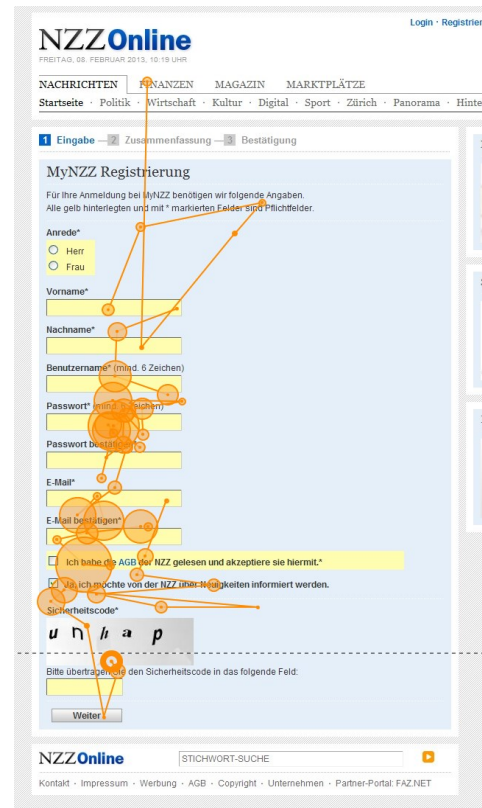

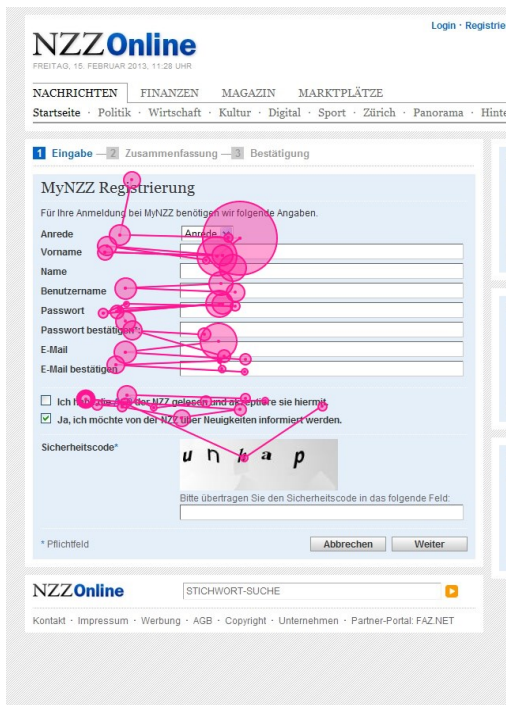**Figure 3. Sample extract of a scanpath in the improved version of the NZZ[TM] form.**

Figure 2. Sample extract of a scanpath in the original version of the NZZ[TM] form.

Subjective Ratings
As not all data follow normal distribution, we applied the non-parametric Mann-Whitney U test to investigate the differences between the improved and the original versions of the forms. Overall, the improved forms received better ratings than their original counter parts. Participants perceived the improved versions as more usable (ASQ, $Z = 2.29$, $p = .011$; FUS, $Z = 2.71$, $p < .001$; SUS, $Z = 2.89$, $p < .001$), as less demanding (NASA-TLX, $Z = 1.85$, $p = .032$) and were more satisfied with them (i.e., FUS item 9), $Z = 1.99$, $p = .024$). However, when analyzing the three different forms separately, differences emerge. As shown in Table 9, only the NZZ form received significantly better ratings on all scales. The Sueddeutsche form, in contrast, only shows higher ASQ ratings. For the Spiegel form none of the comparisons turn out significant. Nevertheless, one should notice that all comparisons between the original and improved versions of the forms show a tendency towards the expected direction.

Effects on single items of the FUS
The original versions of the three forms have different usability issues. Therefore we analyzed the forms separately on single item level of the FUS, which is a questionnaire designed to measure form usability. Figure 4 shows that applying the guidelines on the Sueddeutsche form leads to improvements regarding the user's ability to fill in the form

Scale	Form	Original (n=32)		Improved (n=32)		Z	r^{1}	Improvement
		M	SD	M	SD			
ASQ	Suedd.	5.03	1.24	5.71	1.18	2.48	**.31**	10%
	NZZ	5.40	1.46	6.35	0.70	3.00	**.38**	14%
	Spiegel	5.79	1.56	5.93	1.03	0.60	.07	2%
FUS	Suedd.	4.60	0.87	4.83	0.62	0.77	**.10**	6%
	NZZ	4.75	0.81	5.49	0.44	3.84	**.48**	14%
	Spiegel	5.17	0.73	5.32	0.70	0.94	.12	5%
SUS	Suedd.	3.86	0.78	4.13	0.50	0.88	.11	5%
	NZZ	4.14	0.70	4.71	0.35	3.80	**.47**	11%
	Spiegel	4.17	0.74	4.36	0.71	1.39	.17	4%
NasaTLX*	Suedd.	22.11	15.12	17.11	12.74	1.61	.20	-5%
	NZZ	18.98	14.40	12.29	8.29	2.21	**.28**	-7%
	Spiegel	18.49	15.56	16.25	13.67	0.40	.05	-2%
Satisfaction (last FUS item)	Suedd.	4.50	1.11	4.56	1.05	0.12	.01	6%
	NZZ	4.72	1.37	5.47	0.88	2.57	**.32**	16%
	Spiegel	4.84	1.11	5.06	1.13	0.98	.12	8%

Note. *Lower values show lower workload. Values in bold are significant at the .05 level (one-tailed test), [1]effect size *r* Mann-Whitney U test ($r \geq .10$ = small, $r \geq .30$ = medium, $r \geq .50$ = large [12]).

Table 9. Descriptive statistics for questionnaire scales.

quickly ($r = .23$) and the user's perception of the helpfulness of error messages ($r = .56$). The NZZ form shows improvements on five items: "I was able to fill in the form quickly" ($r = .38$), "Mandatory fields were clearly visible in the form" ($r = .34$), "I always knew which information was expected" ($r = .46$), "I knew at every input which rules I had to stick to" ($r = .64$), and "In the event of a problem I was instructed by an error message how to solve the problem" ($r = .41$). Finally, the improved version of the Spiegel form shows higher ratings only on the item "I knew at every input which rules I had to stick to" ($r = .49$).

Effects on single items of the NASA-TLX
As the NASA-TLX measures workload in a rather broad sense, it might be that its overall score is not able to capture the subtle differences in design between the original and improved versions. Therefore we conducted an analysis on single item level of the NASA-TLX. Results show that the improved version of both, the Sueddeutsche and the NZZ form, is perceived as being significantly less frustrating ($r = .23$, resp. $r = .37$) and users feel more successful in performing the task with it ($r = .34$, resp. $r = .36$). There are no effects on workload with the Spiegel form.
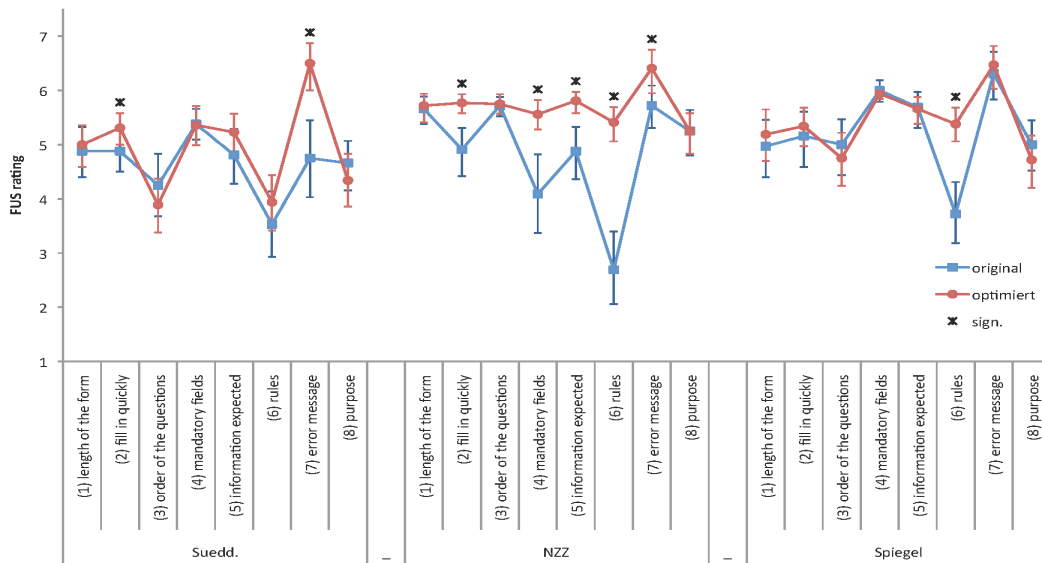

Figure 4. Single item analysis of all FUS questions for original and improved versions.

Interview data

Most frequently mentioned issues
All interview data were analyzed by grouping similar issues for positive and negative comments

- Example of a positive comment: *"I think the form is clear, I immediately knew what I had to fill in and where. And I got an error message telling me how to do it right."*

- Example of a negative comment: *"It was annoying not to know the rules for the username and password first but only learn about them in the second step."*

We further made subgroups for each form and version. In a first step, we counted the number of issues per group showing that the most mentioned issues over all original forms were missing format specifications, insufficient identification of required and optional fields and that there were too many fields overall. Positive comments regarding the original forms were about easy and fast filling, clear identification of required and optional fields, and well-structured and clearly arranged forms. The most frequently reported negative aspects over all improved forms were: unappealing design of the whole site, too many fields, and the cumbersome Captcha fields. The positive comments concerned easy and fast filling in, clear identification of required and optional fields, and the logical sequence of the fields. See Table 10 for details.

Differences between the two versions in issues mentioned
As the most mentioned issues differ between the original and original versions, we analyzed the comments by means of chi-square tests. Participants assigned to the original form condition mentioned significantly more often missing format specifications ($\chi2 = 7.74$, $p = .003$) and insufficient identification of required and optional fields ($\chi2 = 4.93$, $p = .013$) than participants assigned to the improved form versions. Detailed analysis considering the three different forms separately shows that these results are mainly due to the differences between the two versions of the NZZ form (missing format specifications: $\chi2 = 13.54$, $p < .001$ and insufficient identification of required and optional fields: *Fisher's p = .002*).

Unexpectedly, participants assigned to the improved forms mentioned significantly more often not liking the design of the whole site (as the forms were totally on the left and on the right were advertisements), $\chi2 = 7.74$, $p = .005$ instead of expressing negative comments about the forms themselves. Detailed analysis considering the three different forms separately shows that these results are due to differences between the two versions of the Sueddeutsche, $\chi2 = 5.85$, $p = .016$. No significant differences were found for the other most frequently mentioned issues.

DISCUSSION
This study showed that by applying the 20 web form improvement guidelines, all three web forms showed improvements in regard to user performance and subjective ratings. Eye tracking data revealed furthermore that the original forms needed more fixations, longer total fixation duration and longer total saccade duration than the improved forms.

Our findings highlight the importance for web designers to apply web form guidelines. A closer look at the form submission trials shows that there is great potential for increasing the number of successful first-trial submissions by applying the guidelines. Thereby website owners can minimize the risk that users leave their site as a consequence of unsuccessful form submissions. Especially guideline 13 (addressing missing format specifications) and guideline 17 (addressing the location and design of error messages) had a remarkable effect on submission trials. This finding is in line with previous research on form guidelines [4, 29].

Furthermore, data for task completion times show an improvement between 18% and 33%. These values are even better than predicted by the Keystroke Level Model Analyzer Tool from Karousos et al. [20] that predicts improvements between 7% and 23%. Eye tracking data also indicate that participants could fill in the improved forms more efficiently as they needed fewer fixations and saccades [13, 16]. This indicates that participants needed

	Original			Improved		
Positive comments	Suedd.	NZZ	Spiegel	Suedd.	NZZ	Spiegel
easy and fast filling in	14	17	10	12	16	12
well-structured and clearly arranged	3	7	7	5	7	7
clear identification of required and optional fields	5	1	13	3	9	14
logical sequence of the fields	1	5	5	6	10	4
Negative comments						
missing format specifications	5	15	2	4	2	2
insufficient identification of required and optional fields	1	10	2	2	0	1
too many fields	6	1	6	4	1	6
design of the whole site	3	0	5	11	5	6
Captcha	4	8	0	4	5	0

Table 10. Number of positive and negative comments for original and improved versions.

less time looking for specific information during form filling in the improved versions and further supports the performance data. This result is comparable to findings of former usability studies on forms [25].

Subjective ratings showed improvement of up to 16%. Items with a relation to guideline 17 (error messages, see [2, 5, 29]) and guideline 13 (format specification, [4]) showed frequent significant improvements. Finally, interview comments showed that the two conditions differed also regarding subjective feedback. While participants assigned to the original form condition mentioned significantly more often missing format specifications and insufficient identification of required and optional fields, participants assigned to the improved form condition more often criticize the layout of the whole site and not issues about the form itself. Therefore, it can be concluded that from the users' point of view, guideline 13, (addressing missing format specifications) and guideline 4 (highlighting the importance of clear identification of required and optional fields), are the most important. These findings support results of former usability studies on form guidelines [4, 26, 30].

Furthermore, our study shows that the ratings of experts and users differ remarkably. While participants assigned to the original form condition mentioned most often missing format specifications and insufficient identification of required and optional fields, experts rated these two aspects as only moderately important (as seventh and ninth out of 20, respectively). Furthermore, although Spiegel and Sueddeutsche violate two of the five most important expert-rated guidelines (see Table 2), these two forms often performed better than the NZZ form.

To sum up, the effort to improve web forms is relatively small compared to the impact on usability, as shown by our study results.

LIMITATIONS AND FURTHER RESEARCH
There are two important limitations regarding this study. First, the study took place in a lab and therefore controlled aspects that may arise when people fill in forms in real world situations. Distracting context factors were reduced to a minimum and participants concentrated on filling in forms and did not work in parallel on other tasks. Furthermore, the study focuses on newspaper online registration forms. Further research is needed to explore whether the findings from this study can be replicated with other type of forms (e.g. longer forms with more than one page or other use cases such as web shops, social networks or e-gov forms). Moreover, it would be interesting to study the implications outside the lab and perform extended A/B testings. Additionally, from an economic standpoint it would be important to know how the guidelines influence not only usability aspects, but also conversion rates. Another emerging topic that will be relevant for the future will be guidelines tailored for mobile applications.

CONCLUSION
This study demonstrates how form improvement guidelines can help improve the usability of web forms. In contrast to former research that focused on the evaluation of single aspects, the present study uses a holistic approach. In a controlled lab experiment we were able to show the combined effectiveness of 20 guidelines on real web forms. The forms used were taken from real websites and therefore reveal that web forms are often implemented in suboptimal ways that lead to lower transaction speed and customer satisfaction. In the worst case, users may not be able to complete the transaction at all. Our results show that even forms on high traffic websites can benefit from an improvement. Furthermore, we showed the advantages of a multi-method approach to evaluate guidelines. We hope this paper animates other researchers to empirically validate existing or new guidelines.

ACKNOWLEDGMENTS

REFERENCES
1 Aeberhard, A. (2011). FUS - Form Usability Scale. Development of a Usability Measuring Tool for Online Forms. Unpublished master's thesis. University of Basel, Switzerland.

2 Al-Saleh, M., Al-Wabil, A., Al-Attas, E., Al-Abdulkarim, A., Chaurasia, M., Alfaifi, R. (2012). Inline immediate feedback in arabic web forms: An eye tracking study of transactional tasks. In *Proc. IIT 2012*, 333-338.

3 Bargas-Avila, J. A., Brenzikofer, O., Tuch, A. N., Roth, S. P., & Opwis, K. (2011a). Working towards usable forms on the World Wide Web: Optimizing date entry input fields. *Advances in Human Computer Interaction*, Article ID 202701.

4 Bargas-Avila, J. A., Brenzikofer, O., Tuch, A. N., Roth, S. P., & Opwis, K. (2011b). Working towards usable forms on the World Wide Web: Optimizing multiple selection interface elements. *Advances in Human Computer Interaction,* Article ID 347171.

5   Bargas-Avila, J.A., Oberholzer, G., Schmutz, P., de Vito, M. & Opwis, K. (2007). Usable Error Message Presentation in the World Wide Web: Don't Show Errors Right Away. *Interacting with Computers*, 19, 330-341.

6   Bargas-Avila, J. A., Orsini, S., Piosczyk, H., Urwyler, D., & Opwis, K. (2010a). Enhancing online forms: Use format specifications for fields with format restrictions to help respondents, *Interacting with Computers,* 23(1), 33-39.

7   Bargas-Avila, J. A, Brenzikofer, O., Roth, S., Tuch, A. N., Orsini, S., & Opwis, K. (2010b). Simple but Crucial User Interfaces in the World Wide Web: Introducing 20 Guidelines for Usable Web Form Design. In: R. Matrai (Ed.), *User Interfaces*, 1-10. InTech.

8   Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189-194). London: Taylor & Francis.

9   Card, S. K., Moran, T. P., & Newell, A. (Eds.) (1983). *The psychology of human computer interaction*. Routledge.

10   Christian, L., Dillman, D., & Smyth, J. (2007). Helping respondents get it right the first time: the influence of words, symbols, and graphics in web surveys. *Public Opinion Quarterly*, 71(1), 113 - 125.

11   Couper, M., Tourangeau, R., Conrad, F., & Crawford, S. (2004). What they see is what we get: response options for web surveys. *Social Science Computer Review,* 22(1), 111 – 127.

12   Field, A. (2009). Non-parametric tests. In Discovering statistics using SPSS (third edition., pp. 539-583). London: SAGE.

13   Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer.

14   Harms, J. (2013). Research Goals for Evolving the 'Form' User Interface Metaphor towards more Interactivity. *Lecture Notes in Computer Science*, 7946, 819-822.

15   Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. P. A. Hancock & N. Meshkati (Eds.). *Human mental workload* (pp. 139-183). Amsterdam: Elsevier Science.

16   Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

17   Idrus, Z., Razak, N. H. A., Talib, N. H. A., & Tajuddin, T. (2010). Using Three Layer Model (TLM) in web form design: WeFDeC checklist development. *Computer Engineering and Applications (ICCEA)*, 385-389.

18   James, J., Beaumont, A., Stephens, J., Ullman, C. (2002). Usable Forms for the Web. Glasshaus, Krefeld.

19   Jarrett, C. & Ganey, G. (2008). Forms that work: Designing web forms for usability. Morgan Kaufmann.

20   Karousos, N., Katsanos, C., Tselios, N., & Xenos, M. (2013). Effortless tool-based evaluation of web form filling tasks using keystroke level model and fitts law. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 1851-1856. ACM.

21   Koyani, S. (2006). Research-based web design & usability guidelines. US General Services Administration, Washington.

22   Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. *SIGCHI Bulletin*, 23(1), 78-81.

23   Money, A. G., Fernando, S., Elliman, T., & Lines, L. (2010). A trial protocol for evaluating assistive online forms for older adults. *Proc. ECIS*, Paper 90.

24   Nielsen, J. (2005). Sixty Guidelines From 1986 Revisited. <http://www.nngroup.com/articles/sixty-guidelines-from-1986-revisited/>

25   Nielsen, J., & Pernice, K. (2010). *Eyetracking web usability*. New Riders.

26   Pauwels, S. L., Hübscher, C., Leuthold, S., Bargas-Avila, J. A. & Opwis, K. (2009). Error prevention in online forms: Use color instead of asterisks to mark required fields. *Interacting with Computers,* 21(4), 257-262.

27   Rukzio, E., Hamard, J., Noda, C., & De Luca, A. (2006). Visualization of Uncertainty in Context Aware Mobile Applications. In *Proc. MobileHCI'06*, 247-250.

28   Seckler, M., Heinz, S., Bargas-Avila, J.A., Opwis, K., & Tuch, A.N. (2013). Empirical evaluation of 20 web form optimization guidelines. In *Proc. CHI '13*, 1893-1898.

29   Seckler, M., Tuch, A. N., Opwis, K., & Bargas-Avila, J. A. (2012). User-friendly Locations of Error Messages in Web Forms: Put them on the right side of the erroneous input field. *Interacting with Computers*, 24(3), 107-118.

30   Tullis, T., Pons, A., 1997. Designating required vs. optional input fields. In: Conference on Human Factors in Computing Systems. ACM New York, NY, USA, pp. 259–26

31   Wroblewski, L. (2008). *Web Form Design: Filling in the Blanks*. Rosenfeld Media.