Investigation of alternative interface designs for long-list questions–The case of a computer-assisted survey in Germany

Jessica M. E. Herzing

LINES/FORS, University of Lausanne, Switzerland, and University of Mannheim, Germany

jessica.herzing@unil.ch

Abstract

This study aims to address the questionnaire design challenges in cases wherein questions involve a large number of response options. Traditionally, these long-list questions are asked in open-ended or closed-ended formats. However, alternative interface design options are emerging in computer-assisted surveys that combine both interface designs. To investigate trade-offs of these alternative designs, a split-ballot experiment was conducted with a) a long list of radio buttons, b) a search tree (nested list of response options), and c) a combo box (combination of a text box and a drop-down box). Based on the question on the highest educational qualification attained from the Innovation Sample of the German Socio-Economic Panel, we investigated the interface design that facilitates respondents optimally and enhances the measurement quality. The findings indicate that combo boxes reduce the response burden and increase measurement details, whereas search trees and long lists reduce post-coding efforts.

Keywords: instrument design; questionnaire design; look-up database; education; measurement

Introduction

Questions with a large number of response options, such as questions on occupation, prescription drugs, country of residence, brand names, names of product groups, fields of study, or educational qualifications are not easily answered by respondents and cannot be easily designed by survey practitioners (Couper & Zhang, 2016; Keusch, 2014; Schierholz, Gensicke, Tschersich, & Kreuter, 2018; Stern, 2008; Tijdens, 2014, 2015). To account for a large number of response options, questions on the referred topics listed above are typically asked either in an open-ended format (e.g. a standard text field) or in a closed-ended format (e.g. in the form of long lists of response options). Conversely, open-ended question formats are challenging for survey organisations because of the codability of the collected data; respondents may provide answers that are invalid or difficult to convert into a codable format (Conrad, Couper, & Sakshaug, 2016). Open-ended question formats impose lesser burden on respondents (Couper & Zhang, 2016; Stern, 2008).

In contrast to open-ended question formats, closed-ended question formats provide respondents with additional information on the question context and increase the codability of responses (Fowler, 1995). However, closed-ended question formats may not present all potential response options (Fowler, 1995; Schneider, 2008), as the number of response options is limited to the page/screen space (Couper, Tourangeau, Conrad, & Crawford, 2004). In addition, closed-ended questions formats will likely introduce response order effects (Galesic, Tourangeau, Couper, & Conrad, 2008; Krosnick & Alwin, 1987) and confuse respondents with unknown response options (Couper, et al., 2004; Lenzner, Kaczmirek, & Lenzner, 2010). Thus, open-ended and closed-ended interface designs for long-list questions challenge respondents and survey researchers in different ways.

To address some of the challenges related to the established interface designs of long-list questions, survey practitioners developed alternative interface designs in computer-assisted surveys, such as combo boxes (combination of a text box and a drop-down box) and search trees (nested list of response options; Couper & Zhang, 2016; Funke & Reips, 2007; Tijdens, 2014, 2015). Although these alternative interface designs for long-list questions may support respondents in their efforts to answer questions and improve data quality, we believe that research on the use of these interface designs has received little attention and has been inconclusive. Thus, it seems prudent to evaluate alternative interface designs for long-list questions in computer-assisted surveys from the respondent and researcher perspectives.

To evaluate alternative interface designs for long-list questions, this study investigates whether presenting a long list of response options (list of radio buttons), a search tree, or a combo box facilitate respondents in answering long-list questions. Furthermore, this study examines whether or not there are differences in data quality between a long list of response options, a search tree, or a combo box. Finally, this study investigates and assesses the extent to which differences exist across the experimental conditions influenced by respondent characteristics. Therefore, this study makes a significant contribution to social research methodology because: a) it is the first study to compare the aforementioned interface designs with each other; b) it contributes to the improvement of data quality; c) it evaluates the burden of respondents when answering long-list questions; and d) it scrutinises the current approach of measuring educational qualifications in computer-assisted surveys.

Alternative interface designs for long-list questions

The possibility to deploy plug-in-based programming languages (e.g. Java or hypertext markup language (HTML) scripts) in survey software enabled various visual interface designs for long-list questions. For example, auto-completion text boxes (completes the entered text automatically) and auto-suggest text boxes (response suggestions are shown below the text field; Funke & Reips, 2007). More recently, search trees (nested list of response options; Tijdens, 2014, 2015), semantic matching tools (identifies semantic similarities in two structures, e.g. between a response and an entry in an underlying database, as reported in Tijdens, 2014, 2015), and combo boxes (a combination of a text field and a drop-down box, also referred to as a combination box or database look-up, as reported in Couper & Zhang, 2016) were developed. The focus of this study is on search trees and combo boxes.

Search trees are nested lists of response options that present different levels of subgroups of response options (Figure 1, Tijdens, 2014, 2015). At a first glance, respondents see a single table with approximately 20 subgroups or generic response categories (single-level). These generic response categories are ordered hierarchically or alphabetically. After the respondent has selected a generic response category, more specific and detailed response options are displayed in a second list (second level with up to 20 response categories). Consequently, a second-level search tree can provide approximately 400 response categories, thus allowing greater granularity and detail of measurement than long lists of response options in established closed-ended question formats. However, search trees do not allow respondents to give a text input on the same survey page. Consequently, search trees work well when a finite list of response options is available.

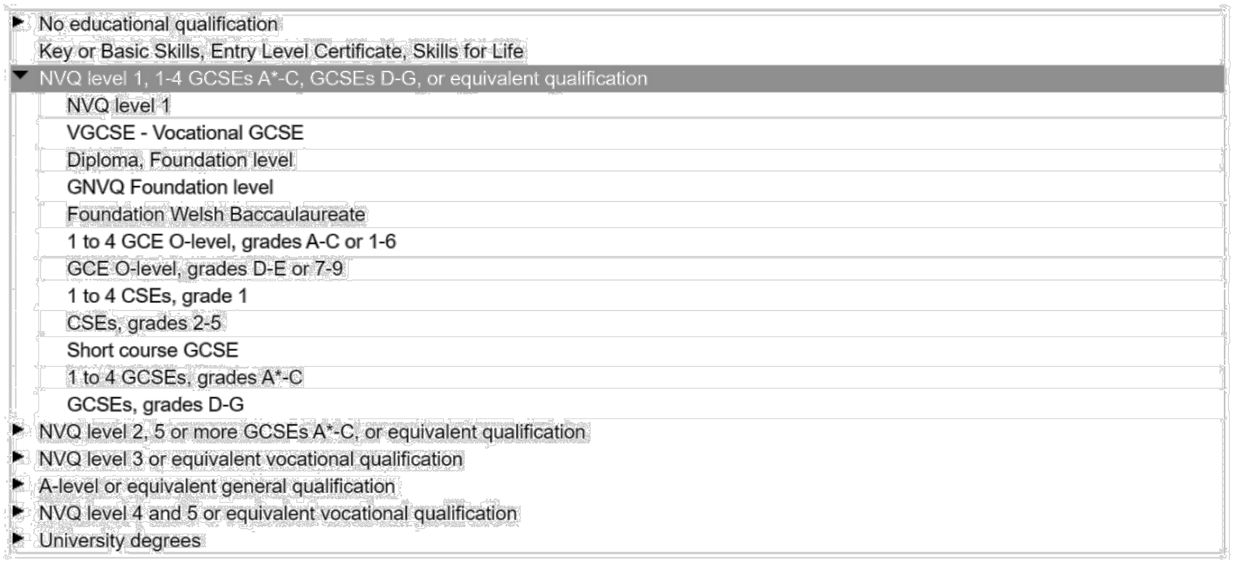

Figure 1. Alternative interface design - Search tree exemplary for levels of education in United Kingdom (England, Wales, and Northern Ireland).

In general, combo boxes constitute the front end of a database query (and allow several thousand entries) which is incorporated in the survey procedure (Figure 2). Initially, combo boxes do not differ from standard text fields, and do not give respondents any guidance regarding the answer format. However, at the moment the respondent begins to type, various response suggestions are offered below the text field (type suggestions). Each additional letter reduces the number of suggestions[1]. Respondents can type a response, select a response from the list of response suggestions, or use a mixture of both approaches. Therefore, combo boxes allow greater granularity and measurement detail than long lists and search trees.

[1] The implemented search algorithm uses text string matching without a fuzzy search. However, the search algorithm ignores special characters, the number of space characters, as well as upper and lower case letters. The suggestions are presented in alphabetical and hierarchical orders.

Figure 2. Alternative interface design - Combo box exemplary for levels of education in United Kingdom (England, Wales, and Northern Ireland).

The concern with search trees relates to the fact that respondents need to navigate through generic response categories to find their preferred response option (for example on grouping educational qualifications please refer to Redline, Tourangeau, Couper, Conrad, & Ye, 2009). However, based on the assumption that respondents understand the generic response categories, respondents have to read fewer response options when they use search trees compared to long lists. In the case of combo boxes, respondents can answer by entering a text response or by choosing a response from a filtered list of response suggestions. Thus, search trees should enable respondents to answer more readily than long lists do, but not as effortless as combo boxes do.

For survey practitioners, it is vital to achieve high-data quality. In this regard, it is essential for survey practitioners to identify measurement differences between interface designs as these may introduce measurement errors (Groves et al., 2011, ch. 8.6). Combo boxes offer more flexibility in answering questions compared to search trees or long lists. Hence, it is more likely that combo boxes produce measurements with greater detail. However, this comes along with greater measurement differences between responses given in combo boxes versus responses made in established interface designs. Consequently, combo boxes may yield answers that are less consistent than those from search trees when they are compared to established interface designs.

Depending on the requirements of the underlying research question, it might be necessary to provide the same interface design multiple times. However, the willingness to answer additional questions with the specific interface design may be influenced by the response difficulty (e.g. Krosnick & Presser, 2009, p. 274). Presuming that combo boxes provide additional assistance, respondents should be more willing to respond to questions with combo boxes multiple times compared with search trees or long lists.

Responses in combo boxes can be automatically coded unless they do not match the entries of the underlying database. In that case, the answers need to be post-coded. Thus, the coverage of potential response options in the underlying database is crucial for the automatic coding in the combo boxes. In contrast to combo boxes, long lists and search trees offer a closed-ended response format. Hence, all responses can be coded automatically unless an 'alternative' category with text input is provided. Therefore, implementing search trees or long lists may reduce post-coding efforts compared with combo boxes.

Example on the measurement of educational qualifications

Our study exemplifies the issue of long-list questions based on the use of the interface design of the question on educational qualification, as the measurement of this commonly used socio-economic variable is a challenging survey research question. The established question format for the highest educational qualification is asked with a long list of radio buttons (for examples Hoffmeyer–Zlotnik, 2016; Redline et al., 2009). However, there is an increase in the complexity in the measure of the educational qualification in Germany and Europe due to the increasing differentiation of educational systems that makes it challenging to present complex educational systems (Schneider, 2008, p. 311). This increase in complexity results in difficulties in recording and coding the answers in classified educational schemes (presented for the German case in table A1 with the International Standard Classification of Education [ISCED]; Schneider, 2013; UNESCO-UIS, 2014).

Given the challenge of designing long-list questions for the measure of educational qualifications, this study evaluates three alternative interface designs[2]. In the following, we strive to maximise data quality for each interface design based on the use of the maximum number of plausible educational levels. Because respondents are not restricted in their responses in the case of combo boxes, the underlying database contains all official and outdated educational levels, as well as synonyms for educational levels. In the case of a search tree, respondents are restricted to the presented response categories. Therefore, synonyms are of minor concern in search trees as one can expect that respondents can map their educational qualifications according to the provided options. Thus, only official and outdated educational qualifications are presented in a hierarchical structure based on the sub-grouping of the responses (Schneider & Ortmanns, 2019). When designing long lists, one should avoid scrolling. Hence, the list of educational qualifications was restricted to the most common official and outdated educational levels as these could be displayed in a proper way on 14´´ or 15´´ laptops (laptop size is commonly used by the survey organization). These varying numbers of response categories across different interface designs also reflect different levels of granularities for the measures of educational qualifications (outlined in table A1), it reflects the different needs of researchers, and makes use of the strengths of each design. Consequently, this

[2] The sample size did not allow more experimental conditions (for information on power analysis see Döring & Bortz, 1995; Lachin, 1981).

article compares three different interface designs regarding the question of educational qualifications[3] in Germany:

(1) A long list of radio buttons with 29 response options from which respondents choose an appropriate response (Figure A1). The 26 educational qualifications were hierarchically ordered and categories for ‘no qualification’, ‘other qualification’, and ‘refuse to answer’, were offered. In contrast to the established interface design which contains a list with 10 very generic response options (Figure A2), this long list allows more granularity.

(2) A two-level search tree with 38 response options (Figure 1, and for the German version, see Figure A3). The single-level search tree contained six very generic response categories for educational levels, a category for ‘no qualification’, and an ‘other’ button. The two-level search tree contained six to eight specific response categories. If no selection was made in the search tree, or the ‘other’ option was not chosen, then respondents could answer in a standard text field on a new survey page. Compared with the established interface design and the long list, the search tree allows a higher level of measurement detail.

(3) A combo box which covered 417 response options (for illustration, see Figure 2, and for the German version, see Figure A4) also allowed a standard text input or no answer. The response options of the combo box cover generic and specific terms for educational qualifications, outdated educational qualifications, and synonyms that are not covered by

[3] German response options are presented because translating educational qualifications is prone to errors (Schneider, Joye, & Wolf, 2016). The combo box and search tree are based on Windows Presentation Foundation (WPF) technology and were recalled by the Kantar survey software nipo developed by TNS infratest, Germany. For further information on the tools see www.surveycodings.org/education.

the other interface designs. The combo box allows the highest measurement detail in comparison with the established interface design, the long list, and the search tree.

Study design

The random split-ballot experiment was conducted in accordance with the German Socio-Economic Panel Innovative Sample (SOEP-IS) which was initiated in 2011 (referred to as 'wave 1'). This panel interviews panel members on a yearly basis (Richter & Schupp, 2012, 2015). In 2014 (referred to as 'wave 2'), the combo box, the search tree, and the long list, were implemented in a split-ballot experiment in accordance with the SOEP-IS (Bohlender & Glemser, 2016, pp. 14–16). In total, 5,141 respondents participated in the SOEP-IS. However, only respondents who participated in the panel since 2011 (Sample I, potentially 1,278 respondents) were eligible for the experiment on interface designs, as these had answered the established educational question in wave 1 (for an illustration, see Figure A2).

Furthermore, only respondents who obtained their highest qualification in Germany and who did not change their educational level in the period between wave 1 and wave 2 were included in the sample (these respondents were excluded from participating in the survey). Moreover, respondents who had extreme response times were also excluded (the 1st and the 99th percentiles). Owing to these restrictions, the sample ended up with 1,039 respondents, while 349 respondents were randomly assigned to the long list, 339 respondents were randomly assigned to the combo box, and 351 respondents were randomly assigned to the search tree (for a composition of the experimental groups, see table A2).

Typically, SOEP-IS respondents were interviewed in person (face-to-face interviews, CAPI). However, for this study, interviewers were asked to turn the computer around so that respondents could answer the question with the alternative interface design on their own (CASI).

The question wording for all three experimental conditions was, 'What is your highest educational qualification?' Respondents assigned to the long list design were asked to choose their highest educational qualification from the list. The respondents assigned to the combobox or the search tree were instructed to select the best matching results for the highest educational qualification they attained.

Operationalisation

Multiple indicators can measure whether interface designs enable respondents to respond more readily. Hereinafter, *response times* are used as a proxy for the complexity of the interface design, and hence, as a proxy for the response burden (Olson & Parkhurst, 2013, p. 45; Malhotra, 2008; Turner, Sturgis, & Martin, 2014; Yan & Tourangeau, 2008). While there were no response times available for the established interface design (wave 1), response times for each of the three experimental conditions were captured by the client-side[4] (wave 2). Owing to the skewed distribution of the response times (skewness = 14.5, kurtosis = 224.5), the outliers were eliminated following the exclusion of the fastest and slowest 1% based on a graphical display and a sensitivity analysis (21 observations were faster than 3.7 s and slower than 378 s). Furthermore, the natural logarithm was applied to normalise the values (Ratcliff, 1993).

Another indicator used for the complexity of interface designs are *response edits* (Healey, 2007; Heerwegh & Loosveldt, 2002) because edits occur when respondents either misread or misunderstood the question or accidentally clicked on the wrong response option. Response editing examined whether response editing occurred before respondents gave their final answer. However, it

[4] Client-side paradata are collected at the level of the respondent's computer (Heerwegh, 2003).

was not possible to capture entry changes for the experimental condition of the long list of radio buttons. Therefore, this analysis focuses on respondents who answered with help of the combo box or search tree.

Measurement differences were operationalised based on the *consistency* between the answer given in the established interface design in wave 1[5] and the answers, which were given in the three alternative interface designs in wave 2. Consistency between answers was measured by transforming the responses from the experiments into the educational coding used by the SOEP-IS in wave 1 (six-point scale, see SOEP, 2014, p. 54, and table A1).

The willingness to answer multiple questions with the same interface design was measured by the *number of educational qualifications*[6]. Starting with a question on the highest educational qualification, respondents were asked whether they had any other educational qualifications. If the answer was 'yes', they were expected to name the educational qualification (this question was repeated up to six times) with the use of the same interface design as that used in the question on the highest educational qualification.

Code-ability was operationalised based on whether a) the responses of the combo box and the search tree were automatically coded or b) a human post-survey coding was used. When a response was not automatically coded, we differentiated whether the human coder was able to code the response or not. The answers were not considered automatically codable if the 'other' category was chosen in the search tree or long list.

---

[5] The responder's educational level was harmonized with their answers in the panel waves 2012 and 2013 to avoid mismatches between the interface designs owing to changes in the respondent's personal educational history.

[6] This was based on a screening question that pertained to multiple educational qualifications.

Finally, this experiment examined whether the logarithms of the response times, editing responses, and consistency between waves, differed according to the respondent characteristics. Socio-demographic characteristics (age in years and level of educational qualification as used in the SOEP) were included in the multivariate models given that Yan and Tourangeau (2008) showed that these variables were associated with questions that were considered demanding for cognition. Furthermore, we examined the interactions of age and educational level with the alternative interface designs to assess whether the facilitation of respondents and the data quality varied across the alternative interface designs based on respondent characteristics[7].

Results

First, the *response times* were compared across the three experimental conditions (table A3). The long list required the maximum amount of time before an answer was provided, with an average response time of 50 s and a median of 43 s, whereas the combo box was associated with the minimum response time, with an average response time of 41 s and a median of 32 s. The response times of the search tree lies in between the response times of the other experimental conditions (table A3).

Overall, there is a significant difference in the response times between the three alternative interface designs ($F_{2;1,039} = 4.6, p = 0.01$). While the difference in the mean and median response times between the long list and the search tree is not significant, the difference in the mean and

[7] However, no multivariate models for the data quality indicators of code-ability and number of educational qualifications are estimated, as these indicators have low-cell frequencies for many respondent characteristics and low-variations in some experimental conditions.

median response times between the long list and the combo box is significant (table A4). However, there is no significant difference between the combo box and the search tree (table A4).

Model 1 (table 1, linear regression) reports the impact of respondent characteristics on a logarithmic response time scale (in seconds)[8]. Medium vocational and higher vocational qualifications are positively associated with logarithmic response times. Furthermore, the use of a combo box is significantly faster than the use of a long list or a search tree ($F_{2;135} = 3.09, p = 0.05$).

The interaction of response times with age and the use of the combo box are positive and significant at the 99% level (table 1, model 1). Model 1 indicates that the interaction effects of medium vocational, higher vocational, and higher educational qualifications, are negatively associated with the use of combo boxes compared to the use of long lists. Furthermore, difference tests reveal that respondents with medium vocational, higher vocational, and higher educational qualifications, respond significantly faster with the use of the combo box compared with respondents with the same educational levels based on the use of a search tree ($F_{2;135} = 13.54, p = 0.001$; $F_{2;135} = 13{,}46, p = 0.001$; $F_{2;135} = 9.04, p = 0.001$).

Second, this study investigated whether there were differences in the *editing responses* between the interface designs of the search tree and the combo box. Table 2 indicates that 41% of the respondents edited their response during the use of the combo box, and 51% of the respondents edited their responses during the use of the search tree ($\chi^2 = 7.35$, $p = 0.007$).

[8] The estimates are also robust when estimating fixed-effect models to account for respondents who are clustered in interviewers.

Table 1. Linear regression of respondent characteristics and alternative interface designs on logarithmic response time scales (model 1), and logistic regressions of respondent characteristics and alternative interface designs on response edits (model 2) and answer consistency (model 3).

	Model 1		Model 2		Model 3	
	Response times		Response edits		Answer consistency	
Independent variables	$\hat{\beta}$ coef.	Std. err.	$\hat{\beta}$ coef.	Std. err.	$\hat{\beta}$ coef.	Std. err.
Age	-0.01	0.00	-0.01**	0.01	0.01	0.01
Being female	-0.07	0.05	0.15	0.16	0.28	0.17
Ref. General elementary qualification						
Medium vocational qualification	0.37**	0.17	-1.78***	0.33	1.69***	0.45
Vocational qualification and/or Abitur	0.14	0.34	-1.39**	0.56	-0.37	0.62
Higher vocational qualification	0.47**	0.24	-1.12*	0.53	0.24	0.63
Higher educational qualification	0.26	0.19	0.04	0.43	1.75**	0.62
Ref. Long list						
Search tree	-0.10	0.25			1.31	0.74
Combo box	-0.63**	0.28	-1.64**	0.69	0.58	0.68
Ref. Age long list*						
Age* search tree	0.01	0.00			-0.01	0.01
Age*combo box	0.01**	0.00	0.01	0.01	-0.02	0.01
*Ref. Educational qualification*long list*						
Medium vocational*search tree	0.40	0.22			-1.14	0.61
Medium vocational*combo box	-0.46**	0.20	1.16***	0.34	0.13	0.60
Vocational and Abitur*search tree	0.47	0.39			-1.15	0.85
Vocational and Abitur*combo box	-0.26	0.40	0.98	0.73	0.41	0.89
Higher vocational*search tree	0.39	0.32			-0.64	0.75
Higher vocational*combo box	-0.76***	0.28	0.76	0.75	-0.42	0.77
Higher education*search tree	0.21	0.24			-0.85	0.75
Higher education*combo box	-0.56***	0.21	-0.94	0.55	0.11	0.68
Number of observations	1,037		689		1,037	

NOTES - $\hat{\beta}$ coef. = beta coefficients, Std. err. = standard errors, Ref. = Reference category. Std. err. adjusted for interviewer clusters. Model 2 includes respondents to the search tree and combo box only (reference category is the search tree interface). In models 1 and 3, two cases were omitted because there were not enough cases on the value of ‘inadequate’ regarding the educational level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2. Descriptive statistics of four indicator variables according to the experimental condition (in percentage).

Indicator variables	Long list	Search tree	Combo box	Test of difference
Response editing	n.a.	51	41	**
Consistency	84	82	80	n.s.
Number of educational qualifications [a]				
1 qualification	88	80	81	
2 qualifications	10	15	15	†.
≥ 3 qualifications	2	5	4	
Codability				
Automatically codable	100	95	52	
Post-codable	0	2	38	***
Not codable [b]	0	3	10	
Number of observations	349	351	339	

NOTES - n.a. = not applicable; n.s. = not significant. [a] Only valid educational qualifications were considered for the human post-coding. [b] Not codable is defined as an ambiguous or insufficient response (e.g. an occupation was named). † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1 and model 2 present the results of logistic regression[9] on the editing responses under the control of respondent characteristics. Respondents with medium vocational qualifications, vocational qualifications and/or Abitur holders, and higher vocational qualifications did not edit their responses to the same extent as respondents with other educational qualifications did. Furthermore, use of the combo box resulted in fewer response edits than the use of the search tree, which is statistically significant at the 99% level. Nevertheless, the results indicate that respondents with medium vocational qualifications who used the combo box changed their responses more often compared with respondents with the same educational qualifications who used the search tree.

[9] In line with the recommendations of Best and Wolf (2015, p. 157), only the direction and statistical significance of the beta coefficients is interpreted in the cases of the logistic regression models 2 and 3.

To investigate the differences of the various interface designs, table 2 presents the percentage of measurement *consistency* between the established interface design used in wave 1 and the three alternative interface designs used in wave 2. All three alternative interface designs achieved a measurement consistency which was > 80%. Hence, there was no difference in the measurements between the three experimental conditions (table 2).

To investigate whether the measurement consistency differs according to the respondent characteristics, a logistic regression with socio-demographic covariates was estimated (table 1, model 3). Respondents with medium vocational or higher educational qualifications provided more consistent answers than respondents with general elementary qualifications. As indicated in the descriptive analysis, the alternative interface designs were not associated with the measurement consistency (table 1, model 3).

The *numbers of educational qualifications* are listed in table 2; 88% of the respondents who received the long list of response options indicated that they had one educational qualification, whereas in the other two interface designs, approximately 80% mentioned one educational qualification. Out of the respondents who got a long list of response options, 10% indicated that they had a second educational qualification. By contrast, 15% of respondents indicated that they had a second educational qualification when responding with the combo box or the search tree. Consequently, respondents indicated a similar number of educational qualifications when they used the combo box or search tree, and they indicated fewer educational qualifications when they used the long list (table 2, $\chi^2 = 8.66$, $p = 0.070$).

In addition, table 2 presents the *code-ability* of the highest educational qualification mentioned (in percentage). When the long list was used, 100% of the responses were codable, as all respondents found their highest educational qualification in the list of response options. In the case of the search tree, 95% of the responses were coded automatically. However, 5% of the respondents could not identify their educational qualifications in the search tree and answered a follow-up question with a standard text field. Out of this 5% of responses, 2% of the responses were codable by human coders (seven cases), whereas 3% of the responses were uncodable by human coders (10 cases named an occupation). In the case of the combo box, 52% of the responses were automatically coded. In 38% of the responses which were entered in the combo box and needed manual coding, human coders identified a code[10]. In 10% of the cases, no coding of the entered response was possible as the response was ambiguous or insufficient (e.g. named a lower educational qualification or an occupation). In total, there was a significant variation between the three alternative interface designs regarding the code-ability of responses in favour of the long list and the search tree (table 2, $\chi^2 = 741.83$, $p = 0.001$).

Discussion

Several indicators were used to examine whether a long list of radio buttons, a search tree, or a combo box, best facilitated the responses to long-list questions in computer-assisted surveys. Based on the example of the highest educational qualification attained, response times were significantly lower when the combo box was used compared with the use of the search tree or the long list. As the combo boxes are similar to standard text fields, this result is in line with the findings of Stern (2008) and Couper and Zhang (2016) who showed that standard text fields and combo boxes are less burdensome for respondents compared to long lists or drop-down boxes. Additionally, the results indicated that as the age increased, the advantage of the combo boxes concerning response times became smaller. Hence, combo boxes appear to present more difficulties for older

[10] These human coded responses were included in the database for later use.

respondents. Furthermore, respondents with medium vocational, higher vocational, and higher educational qualifications, are faster when the combo box is used compared with respondents with the same educational qualifications who used the long list. Thus, combo boxes can be answered faster. However, this effect is moderated by the respondent's age and educational level.

Besides respondents with medium vocational qualifications, the split-ballot experiment revealed that there were significantly fewer response edits when the combo box was used compared with the use of the search tree. These findings indicate that most respondents have to expend fewer efforts in finding their intended responses in the case of combo boxes.

The comparison of measurement consistencies shows no variation across the three experimental conditions. Consequently, the given example provides no evidence that the implementation of either the long lists, search trees, or combo boxes, result in significant measurement differences compared with the previous measures assumed with established interface designs.

In addition, the study shows that respondents answering with the combo box and the search tree mention more educational qualifications than respondents in the long list design. Consequently, for researchers who are presenting an interface design multiple times, the combo box and the search tree seem to have clear advantages over long lists.

The amount of post-survey coding shows that researchers have to expend more efforts in post-survey coding when they use the combo box compared to the long list or the search tree. The low percentage of codable responses when answering with the combo box can be explained a) either by the fact that many entries did not exist in the underlying database, or b) because of typographical errors (Schneider, Briceno-Rosas, Herzing, & Ortmanns, 2016). Thus, as long as there is no fuzzy match and the underlying database does not cover the majority of response categories, long lists and search trees have an advantage over combo boxes concerning the post-survey coding effort.

Conclusions

Alternative interface designs for long-list questions have already been implemented in computer-assisted surveys; however, little is known about the optimal interface design to help respondents provide answers to enhance the quality of measurements (Couper & Zhang, 2016, and Tijdens, 2015). This study achieved an improvement in the interface designs of long-list questions based on comparisons of the combo box, search tree, and long list.

This study suggests that combo boxes and search trees are alternatives for the establishment of interface designs of long-list questions. However, these designs are associated with a trade-off between accuracy and costs. Based on our indicators for response burden, we can assume that combo boxes seem to provide maximum help to respondents answering long-list questions. Although combo boxes cover the largest number of response categories, they increase post-survey coding efforts, thus increasing costs. Conversely search trees can cover fewer response categories than combo boxes (but more than long lists), but they do not cause considerable increases in the amount of coding efforts compared with long lists. Therefore, neither the search tree nor the combo boxes are the magic solutions for the design challenge associated with long-list questions. Accordingly, researchers have to choose one or the other interface design based on their research intentions (also Couper & Zhang, 2016; Tijdens, 2015). In this regard, one promising example is the implementation of combo boxes and search trees for measuring educational qualifications in cross-cultural, migrant, or refugee surveys (Schneider, Briceno-Rosas, Ortmanns, & Herzing, 2018).

This study had certain limitations concerning the interface designs. The experimental set-up is driven by the argument that maximises quality in each interface design, and that uses the strengths of each design to produce comparable data rather than identical data (designs are not interchangeable), which results in different number of response categories. While one cannot rule out that the results of this study may be confounded by the number of response options presented, it can also be argued that this difference is in favour of the long-list design, as this interface design is associated with the lowest number of answer options. Consequently, one may expect even greater advantages in the case of the combo box if the number of categories is the same for all experimental conditions.

In addition, one may argue that the search algorithm and the database of the combo box was not ideal and will result in more post-survey codings. Future studies will have to show whether the optimisation of search algorithms and the database can improve the database request, and hence increase the amount of automatically coded responses for combo boxes.

Despite these limitations, the results are encouraging for those who want to implement combo boxes in their survey procedures. Likewise, if these alternative interface designs worked in the case of education, they should be suitable for different types of questions with many response options, such as political party candidates. This evaluation of alternative interface designs contributes to the research on the interface design for long-list questions and shows how survey practitioners can improve the measurement of the key socio-demographic variable of educational qualification.

Data availability

The data used in the analyses of this article are freely available as part of the Scientific Use Files (SUFs) of the DIW Berlin (SOEP-IS) survey data. They can be requested from the SOEP Data Archive at https://www.diw.de/en/diw_02.c.222829.en/access.html. The data sets used are catalogued under the following DOI numbers: 10.5684/soep.is.2011, 10.5684/soep.is.2012, 10.5684/soep.is.2013, and 10.5684/soep.is.2014.

Funding

This work was supported by the Leibniz-Association through the Leibniz competition under grant number SAW-2013-GESIS-5 249.

Acknowledgments

The author gratefully acknowledge support from the SOEP Innovation Sample at the DIW, Berlin and the Collaborative Research Center (SFB) 884 "Political Economy of Reforms" at the University of Mannheim, Germany. I especially thank Silke Schneider, Verena Ortmanns, Beatrice Rammstedt, and David Richter for their support in making this experiment possible. I am grateful for comments on a previous version from Annelies G. Blom, Frauke Kreuter, Edith D. de Leeuw, Florian Keusch, and Barbara Felderer.

References

Best, H., & Wolf, C. (2015). Logistic regression. In H. Best & C. Wolf (Eds.), *The SAGE Handbook of Regression Analysis and Causal Inference* (Chap. 8, pp. 153–171). Los Alamitos, CA: Sage.

Bohlender, A., & Glemser, A. (2016). *SOEP-IS 2014 – Methodenbericht zum Befragungsjahr 2014 des SOEP-Innovationssamples*. SOEP Survey Papers 339: Series B. Berlin: DIW/SOEP

Conrad, F. G., Couper, M. P., & Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes. *Journal of Official Statistics*, 32, 75–92. doi:10.1515/jos-2016-0003

Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get: Response options for web surveys. *Social Science Computer Review*, 22, 111–127. doi:10.1177/0894439303256555

Couper, M. P., & Zhang, C. (2016). Helping respondents provide good answers in web surveys. *Survey Research Methods*, 10, 49–64. doi:10.18148/srm/2016.v10i1.6273

Döring, N., & Bortz, J. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin, DE: Springer.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Applied Social Research Methods Series. Thousand Oaks, London, New Dehli: Sage Publications.

Funke, F., & Reips, U.-D. (2007). *Dynamic form: Online Surveys 2.0*. Paper presented at the General Online Research Conference (GOR 2007), Leipzig, Germany.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, 72, 892–913. doi:10.1093/poq/nfn059

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey Methodology* (2nd). Wiley Series in Survey Methodology. Hoboken, NJ: John Wiley & Sons.

Healey, B. (2007). Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in web surveys. *Social Science Computer Review*, 25, 111–128. doi:10.1177/0894439306293888

Heerwegh, D. (2003). Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey. *Social Science Computer Review*, 21, 360–373. doi:10.1177/0894439303253985

Heerwegh, D., & Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review*, 20, 471–484. doi:10.1177/089443902237323

Hoffmeyer-Zlotnik, J. H. P. (2016). Standardisation and harmonisation of socio-demographic variables. *GESIS Survey Guidelines*. Mannheim, DE: GESIS – Leibniz Institute for the Social Sciences. doi:10.15465/gesis-sg_en_012

Keusch, F. (2014). The influence of answer box format on response behavior on list-style open-ended questions. *Journal of Survey Statistics and Methodology*, 2, 305–322. doi:10.1093/jssam/smu007

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201–219. doi:10.1086/269029

Krosnick, J. A., & Presser, S. (2009). Question and questionnaire design. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (Chap. 9, pp. 263–315). San Diego, CA: Elsevier.

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2, 93–113.

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24, 1003–1020. doi:10.1002/acp.1602

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72, 914–934. doi:10.1093/poq/nfn050

Olson, K., & Parkhurst, B. (2013). Collecting paradata for measurement error evaluations. In Kreuter (Ed.), *Improving surveys with paradata: Analytic uses of process information* (Chap. 3, Vol. 581, pp. 43–72). Wiley Series in Survey Methodology. Hoboken, NJ: John Wiley & Sons.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532.

Redline, C. D., Tourangeau, R., Couper, M. P., Conrad, F. G., & Ye, C. (2009). The effects of grouping response options in factual questions with many options. In *JPSM Research Paper*. Retrieved from http://www.fcsm.gov/09papers/Redline_IX-B.pdf, Washington, DC: Annual Conference of the Federal Committee on Statistical Methodology.

Richter, D., & Schupp, J. (2012). *SOEP Innovation Sample (SOEP-IS) – Description, structure and documentation*. SOEP Papers on Multidisciplinary Panel Data Research. Berlin, DE: Deutsches Institut für Wirtschaftsforschung, DIW.

Richter, D., & Schupp, J. (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch*, 135, 389–399. doi:10.3790/schm.135.3.389

Schierholz, M., Gensicke, M., Tschersich, N., & Kreuter, F. (2018). Occupation coding during the interview. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181, 379–407. doi: 10.1111/rssa.12297

Schneider, S. L., Briceno-Rosas, R., Herzing, J. M. E., & Ortmanns, V. (2016). Overcoming the shortcomings of long list showcards: Measuring education with an adaptive database lookup. In *9th International Conference on Social Science Methodology*. RC33 Conference, Leicester, UK.

Schneider, S. L., & Ortmanns, V. (2019). *Database of educational attainment, with explanatory note* (Deliverable 8.8 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221). Retrieved from: www.seriss.eu/resources/deliverables

Schneider, S. L. (2008). Suggestions for the cross-national measurement of educational attainment: Refining the ISCED-97 and improving data collection and coding procedures. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries* (Chap. 17, pp. 311–330). Mannheim, DE: MZES.

Schneider, S. L. (2013). The International Standard Classification of Education 2011. In E. Birkelund (Ed.), *Class and Stratification Analysis* (Vol. 30, pp. 365–379). Comparative Social Research. Bradford, UK: Emerald Group Publishing Limited.

Schneider, S. L., Briceno-Rosas, R., Ortmanns, V., & Herzing, J. M. E. (2018). Measuring migrants' educational attainment: The CAMCES tool in the IAB-SOEP Migration Sample. In D. Behr (Ed.), *Surveying the migrant population: Consideration of linguistic and cultural issues*. Cologne, DE: GESIS Schriftenreihe. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-58550-6

Schneider, S. L., Joye, D., & Wolf, C. (2016). When translation is not enough: Background variables in comparative surveys. In C. Wolf, D. Joye, T. W. Smith, & Y.-C. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (Chap. 20, pp. 288–307). London, UK: Sage.

SOEP Group. (2014). SOEP 2013 – *Documentation of person-related status and generated variables in PGEN for SOEP v30*. SOEP Survey Paper 250 (Series D). Berlin, DE: DIW/SOEP.

SOEP-IS Group. (2017). SOEP-IS 2011 – Questionnaire for the SOEP Innovation Sample. SOEP Survey Papers 456: Series A – Survey Instruments (Erhebungsinstrumente). Berlin, DE: DIW Berlin/SOEP.

SOEP-IS Group. (2018). SOEP-IS 2014 – Fragebogen für die SOEP-Innovations-Stichprobe (Update soep.is.2016.1). SOEP Survey Papers 519: Series A – Survey Instruments (Erhebungsinstrumente). Berlin, DE: DIW Berlin/SOEP.

Stern, M. J. (2008). The use of client-side paradata in analyzing the effects of visual layout on changing responses in web surveys. *Field Methods*, 20, 377–398. doi: 10.1177/1525822X08320421

Tijdens, K. (2014). Dropout rates and response times of an occupation search tree in a web survey, *Journal of Official Statistics* 30, 23-43. doi:10.2478/jos-2014-0002

Tijdens, K. (2015). Self-identification of occupation in web surveys: Requirements for search trees and look-up tables. *Survey Methods: Insights from the Field*. Retrieved from https://surveyinsights.org/?p=6967. doi:10.13094/SMIF-2015-00008

Turner, G., Sturgis, P., & Martin, D. (2014). Can response latencies be used to detect survey satisficing on cognitively demanding questions? *Journal of Survey Statistics and Methodology*, 3, 89–108. doi:10.2196/publichealth.5240

UNESCO-UIS. (2014). ISCED 2011 International Standard Classification of Education. Montreal: UNESCO-UIS. Retrieved from http://uis.unesco.org/en/isced-mappings.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22, 51–68. doi: 10.1002/acp.1331

Appendix

Table A1. Transfer of the ISCED codings used for the different interface designs.

ISCED categories		Detailed alternative ISCED 2011 generated by the alternative interface designs		SOEP coding	
Code	Label	Code	Label	Code	Label
0-1	Less than lower secondary education	0	Less than primary education	1	Inadequately
		100	Primary education		
2	Lower secondary education	243	General lower secondary completed, without direct access to upper secondary education	2	General elementary
		244	General lower secondary completed, with direct access to upper secondary education		
		253	Vocational lower secondary completed, without direct access to upper secondary education		
		254	Vocational lower secondary completed, with direct access to upper secondary education	3	Medium vocational
3 general	General upper secondary education	343	General upper secondary completed, without direct access to tertiary education		
		344	General upper secondary completed, with direct access to tertiary education		
3 vocational	Vocational upper secondary education	353	Vocational upper secondary completed, without direct access to tertiary education		
		354	Vocational upper secondary completed, with direct access to tertiary education		
4	Post-secondary non-tertiary education	443	General post-secondary non-tertiary education completed, without direct access to tertiary education	4	Vocational and Abitur
		444	General post-secondary non-tertiary education completed, with direct access to tertiary education		
		453	Vocational post-secondary non-tertiary education completed, without direct access to tertiary education		
		454	Vocational post-secondary non-tertiary education completed, with direct access to tertiary education		
5	Short-cycle tertiary education	560	Short-cycle tertiary education	5	Higher vocational
6	Bachelor’s level	660	Bachelor’s or equivalent level	6	Higher educational
7	Master’s level	760	Master’s or equivalent level		
8	Doctoral level	860	Doctoral or equivalent level		

NOTE. – ISCED coding is based on Schneider, Briceno-Rosas, Ortmanns, and Herzing (2018, table 5.5), while the SOEP coding is based on the SOEP Group (2014, p. 55) documentation.

Table A2. Sample composition according to experimental group.

	All	Long list	Search tree	Combo box
Mean age in years	51	51	52	51
Being female (in %)	50.6	49.3	47.9	54.9
Educational level[a] (in %)				
Inadequately	0.2	0.3	0.0	0.3
General elementary	13.3	10.9	14.5	14.5
Medium vocational	56.5	59.9	56.1	53.4
Vocational and Abitur	6.0	4.3	6.3	7.4
Higher vocational	5.3	5.7	4.8	5.3
Higher educational	18.8	18.9	18.2	19.2
Overall	1,039	349	351	339

NOTE. –.[a] Educational level is based on SOEP coding according to the SOEP Group (2014, p. 55) documentation. Difference tests did not indicate a significant difference across experimental groups.

Table A3. Response times in seconds for different experimental conditions (interface designs).

Experimental conditions	Mean	SD	Median	Min.	Max.	25th percentile	75th percentile	n
Long list	50	38	43	4	244	17	75	349
Search tree	45	41	35	4	319	22	54	351
Combo box	41	38	32	4	377	20	52	339
Overall	45	39	34	4	377	20	61	1,039

NOTES: - SD = standard deviation, Min. = Minimum, Max. = Maximum, n = number of observations. Twenty-one observations were excluded because their response times were in the 1st and the 99th percentiles.

Table A4. Difference tests for response times between experimental conditions (interface designs).

Experimental conditions	Tests of difference for means[a]	Tests of difference for medians[b]
Long list vs. search tree	$t = -1.56, p = 0.36$	$\chi^2 = 1.65, d.f. = 1, p = 0.12$
Long list vs. combo box	$t = -3.03, p = 0.01$	$\chi^2 = 4.89, d.f. = 1, p = 0.03$
Search tree vs. combo box	$t = 1.49, p = 0.41$	$\chi^2 = 3.07, d.f. = 1, p = 0.08$

NOTES - Twenty-one observations were excluded because their response times were in the 1st or the 99th percentiles. *d.f.* = degrees of freedom, χ^2 = chi-square test, *t* = t-value, *p* = p-value.

[a] Difference test is based on a pairwise comparison of means with Bonferroni's correction.

[b] Difference test is based on a median χ^2 test.

Figures

- ○ No vocational training
- ○ Degree for basic vocational training year
- ○ Degree for full-time vocational school, commercial college (basic professional skills)
- ○ Professional in-plant training period with certificate, but no apprenticeship
- ○ Partial skilled-worker / craftsperson degree
- ○ Job-oriented degree from full-time vocational school
- ○ Yearlong school of health care
- ○ 2 to 3 years long school of heath care
- ○ Commercial or agricultural apprenticeship, skilled-worker / craftsperson degree
- ○ Commercial / mercantile or other apprenticeship
- ○ Vocational school degree for state-approved educator
- ○ Master craftsperson, technician or equivalent vocational school degree
- ○ Training for civil servants in lower or middle grade
- ○ Bachelor, college of advanced vocational studies / dual university
- ○ Master, college of advanced vocational studies / dual university
- ○ Diploma, college of advanced vocational studies / dual university
- ○ Bachelor, administrational / advanced technical college
- ○ Master, administrational / advanced technical college
- ○ Diploma, administrational / advanced technical college (GDR: Engineer and vocational school degree)
- ○ Bachelor, university or other college/ academy (e.g. for art or music)
- ○ Master, university or other college/ academy (e.g. for art or music)
- ○ Diploma, university or other college/ academy (e.g. for art or music)
- ○ Master ("Magister"), university
- ○ State examination, university or other college/ academy (e.g. for art or music)
- ○ Postgraduate studies ("Aufbaustudium"), university
- ○ Doctorate / Ph.D
- ○ Post-doctoral dissertation ("Habilitation")

- ○ Other degree
- ○ No answer

Figure A1. Established interface design - Long list with 29 response options for levels of education in Germany (SOEP-IS Group, 2018).

☐ Apprenticeship in crafts / trades / agriculture

☐ Commercial or other apprenticeship

☐ Full-time vocational school / commercial college

☐ Health sector school

☐ Vocational school such as a school for master craftspeople or technical college

☐ Technical college / school of engineering

☐ University without a doctoral degree

☐ University with a doctoral degree

☐ Other vocational training or university degree

☐ No answer

Figure A2. Established interface design - Long list with 10 items for levels of education in Germany (SOEP-IS Group, 2017).

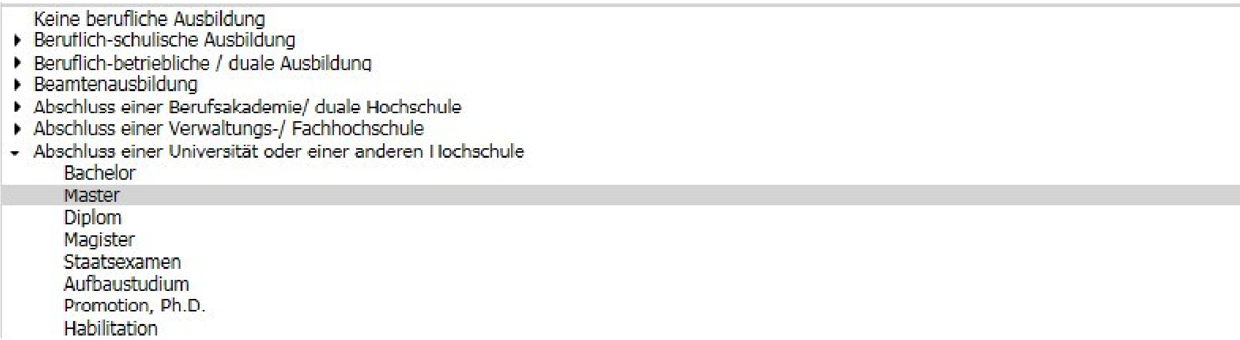

Figure A3. Alternative interface design - Search tree with 38 response options for level of education for Germany (slight design differences compared to figure 1 are due to later design adjustments for the international version).

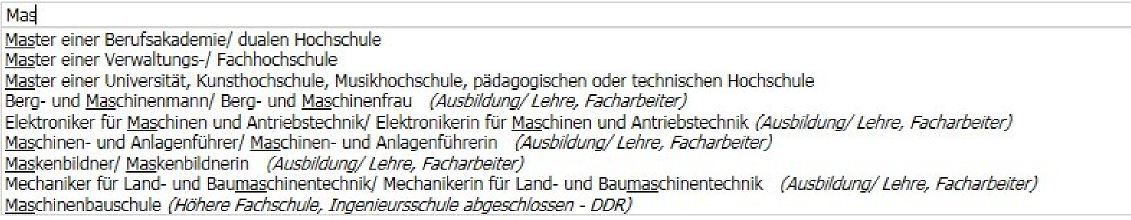

Figure A4. Alternative interface design - Combo box with 417 response suggestions for level of education for Germany (exemplified after three characters were entered; slight design differences compared to figure 2 are due to later design adjustments for the international version).