Coding Issues of Open-Ended Questions in a Cross-Cultural Context

Evi Scholz, Brita Dorer, and Cornelia Zuell

ABSTRACT
Although cross-cultural surveys increasingly use open-ended questions to obtain detailed information on respondents’ attitudes, the issue of coding quality is rarely addressed. These questions are always challenging but even more so in multilingual, cross-cultural research contexts as the different survey languages make response coding more difficult and costly. In this paper, we examine coding issues of open-ended questions and the impact of translation on coding results by comparing codings of translated responses (two-step approach with translation and coding) with codings of the same responses in the original languages (one-step approach using bilingual coders). We draw on data from the project CICOM, specifically respondents’ answers in English and Spanish to open-ended questions about the meaning of left and right. Our goal is to determine whether the coding approach makes a difference to data quality and to identify error sources in the process. Positive news is that both coding approaches resulted in good quality data. We identify several error sources related, first, to respondents’ short answers; second, to the translation process; and third, to the coding process. The response context and the cultural background of translators and coders appear to be important.

Introduction
Coding textual data is generally a demanding task. A high-quality categorization scheme with detailed coding rules and exhaustive and mutually exclusive categories is required. Coders must have the necessary cognitive abilities—that is, they must be capable of understanding the coding rules and applying them consistently (Krippendorff 2004:127). Moreover, they should also have a level of familiarity with the topic under consideration and should undergo intensive coder training to ensure a high quality of the coded data (Krippendorff 2004:128–130).

Coding textual data is especially demanding in multilingual, cross-cultural studies due to the different languages in which they are conducted (Bristle et al. 2019:10). This
additional layer of different languages constitutes an extra challenge because the members of the project teams do not usually have a command of all of these languages. The syntax and semantics of the data languages are even more important, as coders must have the same understanding of the meaning of specific categories across languages. A special type of textual data is found in surveys—namely, answers to open-ended survey questions. Open-ended questions in cross-cultural surveys are particularly challenging because the surveys are fielded in different countries and cultures, usually in different languages. All these issues relate to the quality of coding, and finally to the quality of the resulting coded data for quantitative analyses.

To determine whether the coding approach makes a difference to data quality, and to identify sources of coding error, the present paper examines the quality of the coding of answers to the open-ended probe questions about the meaning of left and right in a cross-cultural survey fielded in English and Spanish. The next section provides some background information on the challenges of and approaches to coding answers to open-ended questions in cross-cultural surveys and on translation as a source of error. This is followed by sections presenting the research question, the data used and methods applied, and the analytical approach. Our project findings are summarized in the Results section and discussed in the Interpretation and Conclusions section, which also provides some basic recommendations. The paper ends with an outlook on how future research could further examine the coding issues of open-ended questions in cross-cultural survey projects.

Background

Coding challenges in cross-cultural projects

Cross-cultural research projects report several coding challenges. Riffe et al. (2019) attribute coding differences to problems such as poorly worded definitions of categories in the scheme and concepts or coding rules that are not understandable in all cultures and languages and cannot therefore be operationalized or applied in all study countries. One problem is that category definitions might be understood differently across cultures. While the aspect of understanding category definitions is important in any coding process, it is even more relevant in cross-cultural projects. If coders in different countries have different understandings of a scheme’s categories, they will either assign the answers to different categories, or they will assign answers to the same category although they mean different things across countries.

Peter and Lauf (2002) conducted two case studies to investigate the coding reliability in cross-national studies focusing on television broadcast news of BBC. Texts were coded by international students of the University of Amsterdam applying a categorization scheme developed in English. Peter and Lauf found that even well trained and guided coders may build in a particular understanding of certain categories which might produce a problem in cross-cultural comparative content analyses that is not or less existing in non-comparative analyses (Peter and Lauf 2002:816). Satisfactory coding quality could thus not be taken for granted. However, the authors identified certain coder characteristics as additional important factors. The authors stressed that "it seems important to sharpen our awareness of the pitfalls of cross-cultural testing and the
potential necessity of a coder-trainer test in addition to the conventional intercoder test,” (Peter and Lauf 2002:827) and they suggested an alternative, multi-step procedure for the cross-national assessment of coding reliability.

Open-ended questions in cross-cultural surveys

Open-ended questions serve many purposes in the empirical social sciences. In surveys, they are used as instruments to assess respondents’ attitudes or knowledge in more detail than is possible with closed questions; they serve as a technical means to avoid excessively long lists of response options; and they are used if the range of possible answers is unknown (Singer and Couper 2017; Zuell 2016). In cognitive pretests, open-ended questions are used to determine why a certain response option was chosen, or whether and how well the meaning of a question was understood (Hadler et al. 2018). In addition, they serve to test for differences in the understanding of questions in a cross-cultural perspective (Zuell and Scholz 2019).

Answering open-ended questions is generally more demanding for respondents than answering closed questions, as they must formulate responses in their own words. Hence, answers are often short and incomplete—for example, a list of keywords or grammatically incomplete sentences. On the other hand, some respondents tell stories or provide answers that are not always related to the question asked.

Although the coding of responses to open-ended questions requires extra efforts, they are or have been used by many high-quality cross-cultural studies. However, these studies have provided little or no information on the coding process or coding quality. Yet, this information would be very important because the coded answers are used in many analyses, and a high quality of coding is a prerequisite for high-quality analyses.

To give some examples of issues related to using open-ended questions in multilingual contexts and to the coding approaches used: Many surveys collect information about respondents’ occupations by asking open-ended questions about their current jobs and the work done in those jobs. In multilingual, cross-cultural surveys, these answers are typically coded using the International Standard Classification of Occupations (ISCO-08; International Labour Office (ILO) 2012). This categorization scheme is available in English, and the study countries must code the answers given in the country-specific languages following the (English-language) instructions provided by the ILO. Mostly no information is available on the coding process and the resulting data quality.

The Political Action II study conducted in 1980 in the Netherlands, the United States, and Germany included an open-ended question on the meaning of the political terms left and right. In each country, all answers were coded by the same human coders using a categorization scheme developed in English. No details of the composition of the coder teams are available, and because of the lack of information about the coding process, the quality of the coding cannot be assessed.

The European Social Survey (ESS) asks some open-ended questions on demographic variables (country of birth, languages spoken, occupation, etc.). The answers are coded decentrally in the individual ESS member countries following standard categorization schemes. Information about coding quality is not available.
Approaches to coding answers to open-ended questions in cross-cultural surveys

How can the various problems related to coding answers to open-ended questions in cross-cultural surveys be overcome? Essentially, three different approaches to coding these answers could be applied:

1. The categorization scheme could be developed in all survey languages, and the original answers could then be coded decentrally by native speakers of these languages. In this approach, the development of a multi-language categorization scheme requires great efforts and the coders could be monolingual native speakers.

2. The categorization scheme could be developed in the main language of the project (for instance, in English), and the original answers to the open-ended questions could then be centrally coded by native speakers of the original language provided these coders are bilinguals—that is, persons “who use two or more languages (or dialects) in their everyday lives” (Grosjean 2015:573). The advantage of applying a one-step approach using bilingual coders is that they understand the texts in one of their two strong languages and code in the other language without a notable risk of losing information due to language switching.

3. The categorization scheme could be developed in the main language of the project. All original answers of the respondents of the different countries could be translated into the main project language by expert translators and then coded by one central team of coders. The advantage of this two-step approach is that the coding process is centrally organized, and all coders receive the same training.

Translation as a source of error

Linguistic systems differ from one another, and translation does not usually mean that only one correct translation can be produced. In many cases, there is room for interpretation due to unclear or missing context, which might give the translated text a specific direction and might change the meaning slightly. Subjectivity is an integral component of translation and cannot be eliminated (Behr 2015). Thus, each translation might be a source of bias or error resulting from text interpretation by the translator. The translation of a categorization scheme into several languages might produce systematic biases or errors when the translated versions are used for coding. Translations of answers to open-ended questions might also produce biases or errors in the coding of one or more terms. In the process of coding answers to open-ended questions in multilingual, cross-cultural surveys, translation might thus result in deviating coding and analysis results. Due to the role of translation in the whole process of answer coding, it is thus essential to test for its impact on data quality. To test for intercoder reliability, Behr (2015) implemented double coding for translated texts. After coding English-language translations of responses to open-ended questions in a cross-national survey, Thrasher et al. (2011) had the codings reviewed by those who collected the original data in the non-English-speaking study countries and
discrepancies were discussed and resolved by the country teams and the central study coordination.

**Research question**

The three approaches to coding answers to open-ended questions in cross-cultural survey research outlined above all have pros and cons, irrespective of procedural logistics or costs. Hardly any studies investigate the quality (reliability and validity) of the coding results across countries (inter-country reliability).

Our focus in this paper is on the reliability aspect of coding quality. Using answers to open-ended questions about the meaning of *left* and *right* from a cross-cultural survey fielded in English and Spanish, we investigate the impact of translation on coding results by comparing the codings of the translated answers to the codings of the same answers in the original languages. The central elements of our study design are summarized in [Figure 1].

| Approach         | Main Project Language | Original Answers | Translation of Original Answers | Categorization Scheme | Coders                                      |
|------------------|-----------------------|------------------|---------------------------------|-----------------------|---------------------------------------------|
| One-step approach | German                | English Spanish  | No                              | Central scheme in German language | Two bilingual coders: one English/German; one Spanish/German |
| Two-step approach | German                | English Spanish  | Into main project language (German) by one expert translator for each language | One monolingual coder: native speaker of main project language (German) | One monolingual coder: native speaker of main project language (German) |

**Figure 1.** Coding approaches compared in our case study.
The results of our study are relevant for cross-cultural comparative research because decisions increasingly have to be made when organizing the coding of multilingual answers. In such contexts, the right balance must be found between spending resources on translation and coding, on the one hand, and the quality of the resulting codings, on the other.

Our research question is whether differences in coding outcomes can be detected between (a) the one-step approach, where answers to open-ended questions are coded in the original languages by bilingual coders according to a German-language categorization scheme (with different coders for different original languages), and (b) the two-step approach, where answers to open-ended questions are first translated from the original languages into German, and then coded by only one German coder according to the same central German-language categorization scheme.

Several sources of bias and error might occur in the process between collecting the multilingual qualitative data (i.e., answers to open-ended questions) and conducting quantitative analyses of such cross-cultural data. These might result from the sometimes unclear content of the respondents’ answers, from the translations, from coding, or from a combination of these error sources. Translators interpret texts while translating, and human coders interpret texts while coding.

Data and methods

Data

The data for this study were drawn from the CICOM project which was carried out by GESIS – Leibniz Institute for the Social Sciences in 2011. CICOM was a research project with the aim to improve indicator validity in cross-cultural surveys. Within the framework of the project, an experimental web survey was conducted in Canada, Denmark, Germany, Hungary, Spain, and the United States in 2011 in the countries’ main national languages (in Canada in English only). Respondents were drawn from centrally organized, non-probability online access panels. Participants in the CICOM 2011 survey were adult citizens of the respective countries, with an upper age cutoff of 65 years. The target sample size per country was 500. To balance the samples for some basic variables, quotas for sex, age, and education were applied to select respondents (Behr et al. 2014). The visual layout and the technical implementation of the web survey were identical across all countries. The only element that differed was the language. In contrast to many commercial web surveys, item response was not forced. Rather, soft checks and item probes were programmed.

Open-ended questions on the meaning of left and right

To test the performance of the two above-mentioned approaches, we used the open-ended questions on the meaning of left and right, which were asked directly after the self-placement left–right scale. The left–right dimension is a central element of political science research and plays an important role in all sorts of analyses and studies. The use of the left–right concept to measure ideology is much discussed among political scientists. It is frequently used in social science surveys, and a detailed categorization
scheme applicable to cross-cultural data is available (Zuell and Scholz 2012). The question text was based on that used in the face-to-face German General Social Survey 2008 (ALLBUS) and adapted to the purposes of the CICOM web survey.

Translation of questions and responses

Both the left–right scale and the open-ended questions were translated from German into the respective survey languages by two professional translators per language. The translations were then reconciled by a third professional translator, reviewed by translation and survey experts in the CICOM team, and where necessary further discussed with the reconciler. The complete question wording of the left–right self-placement and of the open-ended questions about the meaning of left and right, as used in Canada, Spain, and the United States are documented in Figure 2.

Answers to the open-ended questions were translated from all survey languages into German. In our case study, we focused on two of the five CICOM languages: English and Spanish. The reason for this choice was that these are the languages understood by the authors of this paper. For English, the translation was produced by a specially trained translation student who was an intern at GESIS at the time of the project. For

| Country    | Question Wording |
|------------|------------------|
| Canada and USA | Many people use the terms “left” and “right” to designate different political positions.  
We have a scale here that runs from left to right. Where would you position your own political views on this scale?  
We have just asked you to place yourself on a scale from left to right.  
Now please tell us what ideas you associate with the term “left”.  
And what ideas do you associate with the term “right”? |
| Spain      | Mucha gente usa los conceptos “izquierda” y “derecha” para designar posiciones políticas diferentes.  
Aquí tenemos una escala que va de izquierda a derecha. Si piensa en sus opiniones políticas, ¿dónde las situaría en esta escala?  
Le acabamos de pedir su clasificación en la escala de izquierda a derecha.  
Por favor, indique qué asocia con el concepto “izquierda”.  
¿Y qué asocia con el concepto “derecha”? |
| Germany    | Viele Leute verwenden die Begriffe „links“ und „rechts“, wenn es darum geht, unterschiedliche politische Einstellungen zu kennzeichnen. Wir haben hier eine Skala, die von links nach rechts verläuft. Wenn Sie an Ihre eigenen politischen Ansichten denken, wo würden Sie diese Ansichten auf dieser Skala einstufen?  
Eben haben wir Sie zu Ihrer Einordnung auf einer Skala von links nach rechts gefragt.  
Bitte geben Sie nun an, was Sie mit dem Begriff „links“ verbinden.  
Und was verbinden Sie mit dem Begriff „rechts“? |

Figure 2. Question wording of the left–right questions in English, Spanish, and German (texts only).
Spanish, the professional translator was a native speaker of German who had lived for around 10 years in Spain (Behr 2015). The translation process was centrally organized by the CICOM project team. To ensure that the translations would be as close to the source text as possible, the translators were instructed to produce a documentary translation—that is, to avoid cultural adaptations, to forgo optimizing wording, and to preserve vagueness and inconsistencies. Word-by-word translation was not desired and typing errors in the source language were to be ignored. The translators were also instructed to add comments in all cases where mere translation did not seem sufficient to reflect the content expressed in the source languages. As the translators’ comments were not taken into account in the coding process, they will not be discussed any further in this paper. Because the translation instructions affected the content of the translated answer texts and their coding, they might thus influence our results.

**Coding of answers to open-ended questions**

In our study, the answers to the open-ended questions about the meaning of left and right often consisted of several arguments reflecting respondents’ ideas. This means that one answer may have been coded into different categories. The maximum number of codes assigned to one answer was 12. For example, one respondent’s answer to the open-ended question on the meaning of left reads: “social justice, equality of opportunity, social democratic party.” It was counted as three arguments, and thus coded into three categories (“justice,” “solidarity,” and “SPD”).

To gain a better understanding of the impact of translation on coding results, we applied two approaches. In the two-step approach with translation and coding, the answers were first translated into German (within the original CICOM project). In a second step, the translated answers were then coded by a German native speaker using a detailed categorization scheme in German language (Zuell and Scholz 2012). This scheme was based on a scheme developed by Fuchs and Klingemann (1989, 1990) and on an updated and amended version of that scheme developed by Bauer-Kaase (2001). The scheme was completely revised, updated, and extended; validated by data from several German surveys; and adjusted to the needs of the cross-cultural CICOM project. The final categorization scheme comprises more than 250 categories reflecting associations with left and right (for the main dimensions, see Figure 3).

Applying the one-step approach without translation, two bilingual coders (one of English and one of Spanish mother tongue; both had an extremely good command of the second language, German) were trained in using our complex categorization scheme. The bilinguals were social science students who had been living in Germany for several years at the time of coding. The coder of the Spanish answers was from Colombia; the coder of the Canadian and U.S. answers was from the United States. According to Grosjean’s definition (2015), both coders were bilinguals.

Our coders thus had experience in the social sciences and were qualified to understand the topic of left and right. Moreover, as bilinguals, they were able to code texts written in their mother tongue (English or Spanish) while applying the German-language categorization scheme in their second language (German). They coded all original answers to the two open-ended questions from their respective mother tongues. Hence,
the answers to the open-ended questions on left–right orientation administered in Spain were coded by a Spanish/German bilingual, and the answers to the open-ended questions administered in the United States and Canada were coded by an English/German bilingual.

Intercoder disagreement is a common phenomenon when two or more human coders code texts (He and Schonlau 2020). In order to improve coding quality, the present authors, as expert coders, corrected the avoidable coding errors of all coders in both the translated and the original texts before testing for intercoder agreement. We thus identified a gold standard for codings (from all selected language versions), resulting in a database free of evident coding errors. Evident errors result, for example, from typing errors or errors where the digits of codes are transposed (e.g., “2010” instead of “2100”). Other coding errors, for instance, codings based on culture-specific misinterpretation of texts, were not corrected in the data.

Finally, we compared the codings of the original answers in English and Spanish with those of the translated answers. We assumed that all differences between the two sets of codings in the dataset resulted from (a) unclear content of respondents’ answers, (b) translation errors and issues, or (c) coding issues. To identify whether the source of these differences was linked to (b), translation, we worked closely with local experts (see Dorer 2021).

Analytical approach

Testing the quality of the coding process by having the same sample of texts coded by two or more different coders and calculating a reliability coefficient are standard practice for monolingual projects (see Freelon 2010). To our knowledge, coding reliabilities have not been tested in multilingual cross-cultural projects so far. The sample of texts

| Dimension              | Categories (examples)               |
|------------------------|------------------------------------|
| Ideologies             | Socialism; Capitalism              |
| General social values  | Equality; Solidarity               |
| Specific social values | Democracy; Welfare state           |
| Social change          | Human; Terror                      |
| Social groups          | Upper class; Workers               |
| Political actors       | Individual parties; Politicians    |
| Issues                 | Social security; Unemployment      |
| Affective evaluation   | Competent; Honest                  |
| Non-codable answers    | Don’t know; Nonsense               |

Figure 3. Main dimensions of the categorization scheme.
for this kind of quality testing should contain texts from all countries included in the study. However, coders are usually unable to code these texts because they do not have the appropriate language skills. Thus, in such projects, reliability is typically country-specific, and no measure is calculated at the cross-country level.

We tested for the quality of codings by calculating the reliability between the codings of the original texts and the codings of the translated texts. We then identified deviations between the two and checked for systematic sources of these deviations. Our goal was to find out which coding approach is less error-prone.

The strength of reliability gives an indication of whether the text base for coding—either in the original or in the translated form—affects the coding results. Following Krippendorff (2004:241), who proposed relying only on variables with reliabilities (Cohen’s kappa) above .80, we assumed that, with a high reliability of .80, it was of minor importance whether the original or the translated version of the text was coded. On the other hand, once the reliability is below .80, the decision whether to translate the answers before coding or to code them directly from the original language has a strong influence both on the codings themselves and on the analyses based on the resulting data. Studying the deviations between the codings of the original texts and the translated texts helps to determine whether they are systematic or randomly distributed.

Results

To gain initial insights into the number of differences and problems, we first calculated the reliability between the codings of the answers to the two open-ended questions on the meaning of left and right applying the one-step and the two-step approaches. Following Holsti (1969:140), we calculated coding reliability as the ratio of the coding agreement to the total number of coding decisions. The number of coding decisions covered both codings in the original language and codings of translated texts. Coding agreement was based on pairs of codings and would be 100% if the number of codings and their values were identical.

One positive result of our study was that coding agreement relating to the two open-ended questions was very high in all three countries (see Table 1). It was highest for Canada and lowest for the United States. Although disagreement was low in all three

Table 1. Coding agreement between one-step and two-step approach.

|        | Canada | Spain | USA | All Languages |
|--------|--------|-------|-----|---------------|
| **Left** |        |       |     |               |
| N of codings\(^1\) | 1440   | 1858  | 1646 | 4944          |
| N of agreements\(^2\) | 697    | 893   | 767  | 2357          |
| N of disagreements\(^3\) | 46     | 72    | 112  | 230           |
| Coding agreement | 97%    | 96%   | 93%  | 95%           |
| **Right** |       |       |     |               |
| N of codings\(^1\) | 1318   | 1839  | 1585 | 4742          |
| N of agreements\(^2\) | 647    | 875   | 756  | 2278          |
| N of disagreements\(^3\) | 24     | 89    | 73   | 186           |
| Coding agreement | 98%    | 95%   | 95%  | 96%           |

\(^1\)N of codings means the sum of all codings assigned by the respective two coders. It includes the N of agreements (which are based on pairs multiplied by 2) plus the N of disagreements.

\(^2\)N of agreements means the number of pairs of codings (between source and target language) that are identical in one answer.

\(^3\)N of disagreements means the number of different codings in one answer (between source and target language).
countries (coding agreement between 93% and 98%), and both approaches worked well and yielded valid results, we wanted to find out where the differences originated from. Due to the low number of disagreements, we will not discuss any individual codings in what follows, but rather concentrate on typical or systematic problems that might have effects on further analyses of the data at hand, and on how such problems can be avoided in the future. A detailed quantitative analysis is not intended, and qualitative linguistic analysis is beyond the scope of this paper (for linguistic analyses, see Dorer 2021).

**Error source: Missing clarity and context**

Differences between codings of original and translated texts resulted from various error sources. The first major error source was the missing clarity and context in respondents’ answers. In most cases, these written answers expressing respondents’ individual understandings of left and right were just short pieces of text. In several cases, an appropriate translation was not possible, because the context needed to properly understand the intended meaning was missing. In some of these cases, the answers could still be coded in the more general categories of the scheme, at least in the original languages. The respondents were not usually experts in political ideology, and, as their answers showed, many had only a basic idea of the meaning of left and right.

One example of such a problem was found in the U.S. respondents’ answers about the meaning of left, where the most frequent coding disagreements resulted from the answer “democrat.” From the respondent’s perspective, this answer certainly made sense. However, from the research perspective, it could not be interpreted correctly either in the original version or in the German translation. The German translator consistently translated “democrat” as “demokratisch” (“democratic”), and this German term was coded as “democracy.” However, the U.S. coder consistently coded the original “democrat” answers as “Democratic Party.” Without more context—that is, without knowing the true intention of the respondents—neither the decision of the English-language coder nor the decision of the translator can be classified as incorrect.

Another frequent example of missing context was the answer “big government” to the open-ended question on the meaning of left in the U.S. responses, where it was not always clear how to interpret and thus code it. The CICOM project translator was instructed to translate it consistently as “mehr Staat” (“more State”), which could be unambiguously coded. However, the term “big government” in the United States may have several facets, referring either to a government that has too much control over people’s lives or to the centralization of political power. It may also be understood as referring to the existence of too many subsidies or social programs. Each possible nuance resulted in a different coding when coded directly from the original English-language answers.

Although the quality of the answer texts was not always the best in terms of richness and conceptual depth (LaDonna, Taylor, and Lingard 2018) and could not be radically improved, other error sources could be dealt with during the research process. These errors related to the translation and the coding process. There might have been true translation errors and translation issues and flaws in the coding process that could be avoided in the future.
Error source: Translation errors

Some coding disagreements caused by translation errors may have resulted from careless translation. For example, “negación” (“negation”) was translated as “Verhandlung” (“negotiation” – “negociación” in Spanish). In addition, quick and dirty translation may have occurred. For instance, dashes or commas were inserted in the texts translated into German, and these punctuation marks changed the meaning. In other cases, two words were combined to form a compound word. To give an example: The term “radical capitalism” in the original English-language text could be coded as “radical” and as “capitalism.” However, after translation into German, it became one word, “Radikalkapitalismus”, which would be coded with only one code, “capitalism.” Nevertheless, these types of errors were not systematic in our texts and should not therefore result in incorrect interpretations of the general study results.

Error source: Translation issues

“Translation issues” refer to problems where translated texts and original texts resulted in different codings that were not caused by an error on the part of the translator. Rather, translation issues were triggered by the mere fact that translation was involved, because each translation adds a certain layer of dysbalance between the source and target languages. Sometimes the translators added comments in their translation template. And in some cases, alternative translations were listed in the translated answers, for example, separated by a forward slash (e.g., “union” was translated as “Gewerkschaft/Einigkeit”, i.e., “trade union/consensus”). Thus, two ideas were expressed in the translation instead of only one idea in the source text. From the translator’s perspective, this was correct because, as described above, they were instructed to carry out documentary translations. If a forward slash was used by the respondent to separate arguments (e.g., “church/temple”), then the translated text would also have a forward slash, and the coders tended to code all words provided. The issue here was that adding a forward slash and a second term in the translation process had a direct effect on the coding results. This was a systematic and common problem in the translated answers.

Another issue was the cultural background of the translator, who was not always familiar with the meaning of specialist terms in the study country. One example of such an issue was the translation of the Spanish term “clase media alta” (“upper middle class”) to German as “Mittel- und Oberschicht” (“middle and upper class”). The Spanish term referred to people above average, also in terms of income, and meant a class in-between middle and upper class, whereas the German translation referred to two different classes (“the middle class” and “the upper class”). A possible explanation might be that the expert translator was not a social scientist and thus not aware of this difference in sociological terminology between Spain and Germany.

Error source: Coding errors and coding issues

As a last group of error sources, we identified true coding errors and coding issues. Coding errors are typical in all coding processes and will occur irrespective of the
approach applied—that is, whether translation is involved or not. In our study, we corrected for obvious coding errors. Remaining coding errors included incorrect code attribution, incorrect understanding of a category’s definition, ignoring the negation of a statement (normally every code also had a code for its opposite, e.g., “freedom of press” and “missing freedom of press”). These coding errors are typical in all coding processes and could be reduced by well-trained coders and by coder control.

Coding issues were a more important problem. One source of such issues was a culturally specific interpretation of texts. In our study, this was due to the country of origin of the coders. In the case of Spanish, we had a Spanish native speaker. However, he/she did not come from Spain but from South America. In the case of the Canadian data, our English-speaking coder was from the United States. The following is an example of imprecise coding in Spanish: “Independentismo,” which in Spain refers to the important political issue of regional independence, was coded incorrectly, whereas a coder who was a Spanish native speaker from Spain would undoubtedly have been familiar with the term and its interpretation in the context of left and right. This example demonstrates the disadvantage of not using bicultural bilinguals (see Grosjean 2015): The language versions of our coders’ mother tongues differed from the language versions of the original answers (Colombian Spanish vs. Spain-Spanish and U.S.-English instead of Canadian English), and our coders were not constantly immersed in both cultures (Germany and Canada/Spain) in their daily lives. Such systematic coding issues could have been avoided by having true bicultural bilingual coders for each of our three language versions (Canadian English, U.S.-English, Spain-Spanish).

Another very frequent problem was caused by incomplete clarity of some categories in the scheme. For example, the category “right to possess arms” and the opposite category “no right to possess arms” were very important categories for the definition of left and right in the United States. Because a neutral category was not included in the categorization scheme, the English-speaking coder had problems assigning the correct code for a typically neutral answer such as “gun control” that did not include any further information as to whether this meant that the respondent was in favor of or against “gun control.” The meaning of the utterance would either have required clarification, or a neutral category should have been included in the scheme.

The most important sources of systematic error in our study were the missing clarity in the answers “big government” and “democrat,” and the coding issue regarding “gun control,” which resulted in lower coding agreements for the U.S. data because these arguments were frequently used by the respondents.

**Interpretation and conclusions**

In our paper, we investigated issues of coding answers to open-ended questions in a cross-cultural survey project, the impact that the translation of answers had on coding, and the sources of error identified in the process. We used data from the CICOM project, which ran online surveys in five selected countries. For this research paper, we used the answers of respondents in Spain, the United States, and Canada to two open-ended questions about the meaning of left and right. We tested two different approaches: In the one-step approach, the answers to the open-ended questions were coded directly from the
language in which they were collected (English and Spanish). In the two-step approach, these answers were first translated into German and then coded by a German native speaker. Our goal was to find out which of the two approaches resulted in more valid codings. One positive result of our investigation was that both approaches worked well and yielded good results, and that it did not make any difference to data quality whether coding was performed from the original texts or from the translated versions. In our project, all coding agreements between the two approaches were much higher than the minimum level of reliability proposed by Krippendorff (2004:241), thus indicating high data quality.

In addition to this positive result regarding the general approach, another added value of our research was the finding that it is possible to reduce the summed up error resulting from several error sources and to make recommendations on how to avoid or reduce such errors in the future.

Our research showed that some precautions in the project design are necessary to obtain good and valid results. We identified three different error sources: missing clarity and context in respondents’ mostly short written answers to open-ended questions, translation errors and issues, and coding errors and issues. The first of these three error sources—missing clarity and context—cannot be completely avoided. Some answers may be clear and comprehensible from the respondent’s perspective, but the translator or the coder cannot know what the respondent wanted to express. One outstanding example of such an ambiguous answer in our study was the response “democrat” to the open-ended question about the meaning of left. One solution might be to code all “democrat” answers starting with an uppercase “D” as “Democratic Party,” and all “democrat” starting with a lowercase “d” as “democracy.” However, this did not make sense for a web survey like CICOM because the respondents were not instructed to correctly follow orthographic rules in their answers. In a face-to-face survey, in contrast, the interviewer could intervene and ask for a more precise answer. Another possibility for clarifying the meaning of such an unclear response would be to check the respondent’s answer to the parallel question about the meaning of right. Very often, respondents’ answers to the second of two consecutive open-ended questions must be seen in context of the response to the first question, as they often answer these questions in connection with one another. If the U.S. respondent’s answer to the question about the meaning of right reads “republican,” and the answer in the case of left is “democrat,” one can assume that the respondent is referring to the Republican Party and the Democratic Party, respectively. However, because respondents do not always use such pairs of terms to express their opinions, this kind of response interpretation will work in some cases, but not in all. Ultimately, it is the researcher’s decision, not the translator’s or the coder’s task, to decide how to deal with unclear responses and which of the codes, if any, should be used in further analysis.

For some open-ended questions, it could be helpful to minimize such problems of misunderstanding by modifying the question wording. Schuman and Presser (1981:105) commented that “open-ended questions, lacking the additional cues of fixed alternatives, may need to be more clearly focused than closed questions.” In other words, in the case of an open-ended question, respondents must understand very clearly what is expected of them. This may help translators, coders, or researchers to better understand the meaning of an answer and thus result in more valid coding.
The second error source in coding related to the translators and their translations. Answers to open-ended questions are a very special type of texts. In contrast to the longer, continuous texts that translators typically translate, respondents’ answers to open-ended questions are often short and incomplete. In addition, some respondents provide answers that are not always related to the question. Thus, translating answers to open-ended questions requires extra efforts, and there are cases where a correct translation is simply impossible, often because the context needed to clearly understand what the respondent wanted to say is missing. Most translators—also expert translators—are not used to translating texts of this type. To improve the quality of translations of answers to open-ended questions, extensive translator training is recommended that provides adequate information on this text type.

In addition, the translation task should be clarified by providing the translators with clearly formulated translation instructions and rules, explaining that the translators should produce documentary translations—that is, avoid cultural adaptations, optimizing wording, and preserve vagueness and inconsistencies to produce translations as close to the source text as possible. These guidelines could also clarify whether or not word-by-word translations are desired, and how typing errors in the source language texts should be handled. In addition, the translation guidelines should include example lists for translations of frequently used terms relating to the topic that might be difficult for the translator to translate (for instance, how to translate “democrat” or “big government”). Finally, the guidelines should contain instructions for translators how to enter comments and how to distinguish comments from the translations themselves, for example by entering them in a separate comment field. More hints for developing guidelines for translators can be found in Mohler et al. (2016) and in Behr, Braun, and Dorer (2016).

The third error source, coding issues, resulted mostly from unclear definitions in the categorization scheme (Zuell and Scholz 2012). In our study, the coders had some problems with a small number of category definitions in the scheme. In a cross-cultural project, it is important to develop a scheme that is valid for all countries in the study. The scheme we used was carefully developed, and it included categories describing, for instance, all political parties in the countries in our study. However, the developers of the categorization scheme did not realize that, for the U.S. context, the answer “gun control” could not be clearly assigned, because the scheme offered only two categories, one defined as positive (“pro gun control”) and the other defined as negative (“no gun control”). As a neutral category was not provided, neutral answers could not be coded unambiguously, which caused coding problems. Problems such as this in a categorization scheme could be reduced by carrying out coding tests whereby a sample of answers could be coded by coders in all countries included in a study, and the results could then be discussed. Finally, training the coders how to use the categorization scheme and how to apply appropriate codes is an important prerequisite to achieving satisfactory coding results for all countries and languages.

Our experiences from the present study show that, besides well-defined translation and coding rules, well-trained translators and coders are crucial. Thus, the selection of translators and coders is an important issue. Krippendorff (2004:127) already stressed this issue for coders. Besides qualifications and training, coders must have the appropriate cognitive abilities and background for coding the data. They must be capable of
understanding the coding rules and applying them consistently across all codings in the project. Additionally, coders need to have a certain familiarity with the topic under consideration (Krippendorff 2004:128). They should be topical experts as well as experts in the culture of the respective country, not only in its language. Based on the experiences from our study, these considerations are not only true for the selection of coders but also for the selection of translators. The cultural knowledge of translators and coders is very important to guarantee the correct understanding of the specific situation in the country answer texts originate from. In our study, the translators and coders should have been familiar with the political situation and current language use to avoid problems such as the incorrect translation of “independentismo” described above. As our study demonstrates, a native speaker of Spanish who does not come from Spain but rather from South America may not always be familiar with expressions in Spain. Similarly, a native English translator or coder, for instance, from Australia, might not necessarily be familiar with specific expressions in the United States. These findings confirm the pitfalls reported by Peter and Lauf (2002).

It is obvious that not all studies will have the resources for selecting native speakers from all study countries who are also experts in the topic under consideration. In light of these limitations, our results underscore the necessity of clear translation rules and guidelines and a very detailed and well-defined categorization scheme, both of which should be enriched with lists of examples. In addition, the training of both translators and coders should be comprehensive and in-depth.

Ideally, a high-quality multilingual categorization scheme should be developed that covers all the different languages and cultures in a cross-cultural project. In this case, translations would not be necessary. Neither translators nor bilingual coders would be needed as the coders could use the part of the scheme that fits their language skills. However, the development of such a high-quality multilingual scheme requires considerable efforts and resources—in terms of time, money, and staff—which are not usually available in academic projects.

**Outlook**

The two coding approaches applied in our study worked well, and both have advantages. One advantage of the one-step approach (coding original texts) is that only one person (the coder) interprets and codes the texts. In contrast, in the two-step approach (translation of the original texts and coding of the translated versions), two persons are involved: First, the translator interprets the texts while translating; second, the coder interprets the already interpreted texts while coding. This results in two potential error sources instead of one. However, one advantage of the two-step approach is that the step of coding the translated answers can be organized centrally and carried out by a team of equally well-trained coders. On that condition (central organization of coding across all countries), it is also possible to calculate the coding reliability across all countries/languages, which helps to assess the quality of the coding in the project.

In the present study, we used answers to open-ended probing questions about the respondents’ individual understandings of political terms. Other types of open-ended questions (such as sensitive questions or knowledge tests) may cause different problems.
Thus, more research on those types of open-ended questions is needed. In addition, we tested the coding process for two different languages and for three countries. Further tests for other languages and countries could help to determine whether other problems and error sources occur during the translation and coding process, whether the problems identified here refer only to the specific setting of our project, or whether the error sources identified in our project are an inherent part of the process of translating and coding answers to open-ended questions and can thus be found in other studies, too.

Notes
1. https://www.icpsr.umich.edu/web/ICPSR/studies/9581
2. See ESS8-2016 documentation, Appendix A6, “Classifications and Coding Standards”, https://www.europeansocialsurvey.org/docs/round8/survey/ESS8_appendix_a6_e01_1.pdf
3. We are aware that there are several definitions of expert translator. In our view, in order to be accorded the status of an expert in translating answers to open-ended questions, a translator should have several years of experience in the field of survey questionnaire translation.
4. CICOM: Enhancing the Validity of Intercultural Comparative Surveys project was funded by the German Research Foundation (DFG) [grant number BR 908/3-1].
5. Sample size per country: Canada 536, Denmark 559, Germany 1,075, Hungary 559, Spain 559, and the United States 547; 3,835 survey respondents in total.
6. The definition of age groups was as follows: aged 18–30 years, aged 31–50 years, and aged 51–65 years. The mean age ranged from 41 to 42 years. The educational groups were lower education and higher education.
7. https://search.gesis.org/research_data/ZA4602
8. CICOM did not strive to obtain perfectly idiomatic translations. Rather, for its purposes, content transfer was in the foreground of the purpose. Thus, this deviation from best practice requirements for traditional translation was deemed a viable option (Behr 2015:289–290).
9. These local experts are researchers (social scientists and/or linguists) who come from and are at least partly residents in one of the three selected study countries. We are very grateful to them for sharing their cultural expertise and acknowledge their support.

Acknowledgements
The authors acknowledge the contribution of the following local experts: Paul Bandia, Concordia University, Montréal, Canada; Patricia Goerman, U.S. Census Bureau, Washington DC, United States; Ana Villar, United Kingdom; Wiebke Weber, Universitat Pompeu Fabra, Barcelona, Spain. We acknowledge useful information on translations in the CICOM project provided by Dorothée Behr, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany.

Disclosure statement
No potential conflict of interest was reported by the authors.

Notes on contributors
Evi Scholz is a senior researcher at the Data and Research on Society Department at GESIS – Leibniz Institute for the Social Sciences in Mannheim, Germany. She works for the International Social Survey Programme (ISSP), where she is responsible for the German survey and is a
member of several ISSP committees. Her main research interests are cross-national survey design and the measurement of background variables.

**Brita Dorer** is a senior researcher at the Survey Design and Methodology Department at GESIS – Leibniz Institute for the Social Sciences. She specializes in the translation of questionnaires for cross-cultural surveys and is heading the work package on Questionnaire Translation of the European Social Survey (ESS). Her scientific interests include the evaluation and quality enhancement of questionnaire translation, the translatability and cultural portability of source questionnaires, close versus adaptive translation, machine translation, and translation process research/cognitive translation studies.

**Cornelia Zuell** was a senior researcher at the Survey Design and Methodology Department at GESIS – Leibniz Institute for the Social Sciences. Her research focuses on methodological aspects of social science surveys. Her research interests include textual and content analyses in the context of data collection and analysis.

**References**

Bauer-Kaase, P. 2001. “Politische Ideologie im Wandel? - Eine Längsschnittanalyse der Inhalte der politischen Richtungsbegriffe links und rechts.” Pp. 207–43 in *Wahlen und Wähler - Analysen aus Anlass der Bundestagswahl 1998*, edited by H.-D. Klingemann and M. Kaase. Wiesbaden: Westdeutscher Verlag.

Behr, D. 2015. “Translating Answers to Open-Ended Survey Questions in Cross-Cultural Research: A Case Study on the Interplay between Translation, Coding, and Analysis.” *Field Methods* 27 (3):284–99. doi: 10.1177/1525822X14553175.

Behr, D., M. Braun, and B. Dorer. 2016. “Measurement Instruments in Cross-National Surveys.” in *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesissg_en_006.

Behr, D., M. Braun, L. Kaczmirek, and W. Bandilla. 2014. “Item Comparability in Cross-National Surveys: Results from Asking Probing Questions in Cross-National Web Surveys about Attitudes towards Civil Disobedience.” *Quality & Quantity* 48 (1):127–48. doi: 10.1007/s11135-012-9754-8.

Bristle, J., M. Celidoni, Ch. Dal Bianco, and G. Weber. 2019. “The Contributions of Paradata and Features of Respondents, Interviewers and Survey Agencies to Panel Co-Operation in the Survey of Health, Ageing and Retirement in Europe.” *Journal of the Royal Statistical Society. Statistics in Society, Series A* (182):3–35.

Dorer, B. 2021. “Translating Answers to Open-Ended Questions in a Multilingual Survey Context – Challenges and Recommendations.” *trans-kom* 14 (2): 226–56.

Freelon, D. G. 2010. “ReCal: Intercoder Reliability Calculation as a Web Service.” *International Journal of Internet Science* 5 (1):20–33.

Fuchs, D., and H.-D. Klingemann. 1989. “Das Links-Rechts-Schema als politischer Code: ein interkultureller Vergleich auf inhaltsanalytischer Grundlage.” Pp. 484–98 in *Kultur und Gesellschaft*, edited by M. Haller, H.-J. Hoffmann-Nowotny, and W. Zapf. Frankfurt am Main: Campus.

Fuchs, D., and H.-D. Klingemann. 1990. “The Left-Right Schema”. Pp. 203–34 in *Continuities in Political Action: A Longitudinal Study of Political Orientations in Three Western Democracies*, edited by M. K. Jennings, J. van Deth, S. H. Barnes, D. Fuchs, F. J. Heunks, R. F. Inglehart, M. Kaase, H.-D. Klingemann, and J. J. A. Thomassen. Berlin: de Gruyter.

Grosjean, F. 2015. “Bicultural Bilinguals.” *International Journal of Bilingualism* 19 (5):572–86. doi: 10.1177/1367006914526297.

Hadler, P., C. Neuert, T. Lenzner, and N. Menold. 2018. “European Working Conditions Survey (EWCS): Cognitive Pretest.” *GESIS Pretest-Datenbank*. Mannheim: GESIS. doi: 10.17173/pretest72.
He, Z., and M. Schonlau. 2020. “Automatic Coding of Text Answers to Open-Ended Questions: Should You Double Code the Training Data?” Social Science Computer Review 38 (6):754–65. doi: 10.1177/0894439319846622.

Holsti, O. R. 1969. “Content Analysis for the Social Sciences and Humanities.” Pp. 596–692 in The Handbook of Social Psychology, edited by G. Lindzey and E. Aronson. Reading, MA: Addison-Wesley.

International Labour Office (ILO). 2012. International Standard Classification of Occupations 2008 (ISCO-08): Structure, Group Definitions and Correspondence Tables. Geneva: ILO.

Krippendorff, K. 2004. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage.

LaDonna, K. A., T. Taylor, and L. Lingard. 2018. “Why Open-Ended Survey Questions Are Unlikely to Support Rigorous Qualitative Insights.” Academic Medicine: Journal of the Association of American Medical Colleges 93 (3):347–9.

Mohler, P., B. Dorer, J. de Jong, and M. Hu. 2016. “Translation: Overview. Guidelines for Best Practice in Cross-Cultural Surveys.” Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. https://ccsg.isr.umich.edu/chapters/translation/overview/#Guidelines.

Peter, J., and E. Lauf. 2002. “Reliability in Cross-National Content Analysis.” Journalism & Mass Communication Quarterly 79 (4):815–32. doi: 10.1177/107769900207900404.

Riffe, D., S. Lacy, F. Fico, and B. Watson. 2019. Analyzing Media Messages. Using Quantitative Content Analysis in Research. 4th ed. New York and Milton Park: Routledge.

Schuman, H., and S. Presser. 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Singer, E., and M. P. Couper. 2017. “Some Methodological Uses of Responses to Open Questions and Other Verbatim Comments in Quantitative Surveys.” Methods, Data, Analyses 11 (2):115–34. doi: 10.12758/mda.2017.01.

Thrasher, J. F., A. C. K. Quah, G. Dominick, R. Borland, P. Driezen, R. Awang, M. Omar, W. Hosking, B. Sirirassamee, M. Boado, et al. 2011. “Using Cognitive Interviewing and Behavioral Coding to Determine Measurement Equivalence across Linguistic and Cultural Groups: An Example from the International Tobacco Control Policy Evaluation Project.” Field Methods 23 (4):439–60. doi: 10.1177/1525822X11418176.

Zuell, C. 2016. “Open-Ended Questions.” in GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_002.

Zuell, C., and E. Scholz. 2019. “Construct Equivalence of Left-Right Scale Placement in a Cross-National Perspective.” International Journal of Sociology 49 (1):77–95. doi: 10.1080/00207659.2018.1560982.

Zuell, C., and E. Scholz. 2012. “Assoziationen mit den politischen Richtungsbegriffen ‘Links’ und ‘Rechts’ im internationalen Vergleich: Kategorienschema für die Codierung offener Angaben.” GESIS-Technical Reports 2012/03.