positive but otherwise relatively small (see Section 4.1.4). They are extensively used in non-probability samples in general, in order to compensate for non-coverage, nonresponse and self-selection (Callegaro et al., 2014c).

- An active approach for coping with non-coverage involves more expensive recruitment strategies to attract the missing segment. Typically this means involving a mixed-mode system, where we approach the missing units with some alternative mode, as discussed in Section 2.1.2. Most typically for web surveys, this means combining them with mail surveys, but can also involve dual frames, where independent and also overlapping samples are used (e.g. sample of mail addresses and sample of emails).

Non-coverage is also one of the important challenges in web survey methodology which is being increasingly explored in the scholarly literature (e.g. Bethlehem & Biffignandi, 2012; Smyth et al., 2010), in national and cross-national general population surveys such as the European Social Survey (e.g. Ainsaar et al., 2013), in various research networks such as the NCRM network[6] or Cost Action WebDataNet,[7] and in the various national probability online panels that are being increasingly established (see Section 5.2).

To summarize, we should keep in mind that we have discussed non-coverage only in a probability sampling context. With respect to non-probability samples, these principles can be applied only as an approximation. Of course, we should be aware of the fact that non-coverage already constitutes an important cause for the non-probability nature of a certain sample. Also, we need to emphasize that non-coverage issues in web surveys are especially critical when surveying the general population. Despite increasing Internet penetration, even in developed countries, in 2014 we still typically have 20–30% of Internet non-users and very few countries have surpassed the 90% benchmark. However, we believe that in a few years this specific problem of Internet non-coverage will disappear in developed countries (e.g. OECD countries). In the rest of the world, this process will be much slower.

## 2.3 QUESTIONNAIRE PREPARATION

Preparation of the questionnaire is usually the lengthiest stage of pre-fielding and typically also requires the largest amount of resources. As with other stages, we predominantly focus on aspects specific to web surveys; however, we cover related general issues to a larger extent in this section, because they are essential for the quality of web questionnaire preparation.

We first provide a general overview of questionnaire development issues (Section 2.3.1), then discuss question types (Section 2.3.2), questionnaire structure, computerization and layout (Section 2.3.3) and engagement aspects (Section 2.3.4). We conclude with questionnaire testing (Section 2.3.5) and integration of questionnaire preparation activities (Section 2.3.6).

### 2.3.1 General Issues

In this section we present a general introduction to survey questionnaire development by reviewing the main definitions, classifications and concepts.

#### 2.3.1.1 Typologies of Survey Questions

The literature often structures survey questions according to their *content* (Bradburn, Sudman, & Wansink, 2004; Groves et al., 2009; Presser, Rothgeb, Couper, Lessler, Martin & Singer, 2004). When a respondent is asked about observable information, we talk about *factual questions*, ranging from simple demographics and other characteristics to complex issues related to behaviour. On the other hand, *non-factual questions* relate to attitudes, opinions, intentions, expectations, beliefs, perceptions, self-classifications, assessments, evaluations, etc. In addition, there are *psychographic* and *knowledge questions*, which are often treated outside the survey context, within psychology and educational testing. With respect to content we can also

distinguish questions according to their *sensitivity* or *threatening nature*. That is, in some cases questions can make respondents uncomfortable, because they invade their privacy, confidentiality or intimacy (e.g. income, sexuality, health) or relate to *socially undesirable* behaviour (e.g. lying, stealing, cheating, etc.).

Another typology often met in the literature is based on the *response format* (Alwin, 2007; Krosnick & Presser, 2010; Saris & Gallhofer, 2014). With the *closed-ended* question format the respondent selects an answer from a list of predefined response categories (e.g. gender, region). In contrast, the *open-ended* question format requires the respondent to type a text (e.g. open comment on some service) or numeric response (e.g. entering a certain value for money).

An essential distinction relates also to *measurement level*, which is determined by the nature of response options offered to the respondent. Due to its importance, this is discussed in almost all methodological and statistical textbooks. For example, respondents can express their degree of happiness – which is a concept related to their inner state – with certain response values which may be expressed with different measurement levels. The corresponding options can range from offering respondents a very unstructured open-ended text format to highly structured measurement level, where respondents select some number, say, on a scale of 1–5. The corresponding measurement level of the question can be nominal, ordinal, interval or ratio:

- For questions with *nominal measurement,* responses can only be distinguished among themselves, but there is no intrinsic ordering in the measurement level itself, so they cannot be sorted. Examples are closed-ended question formats asking about gender, region, religion, as well as open-ended text formats asking for some descriptions.
- *Ordinal measurement* allows response categories to be sorted (e.g. we can say that daily usage is more frequent than weekly usage, or that the response category 'happy' means more happiness than 'unhappy'), but

does not allow comparing the differences between categories.
- In *interval measurement,* distances between categories of responses can be compared (e.g. we can say that the difference between years 1994 and 2004 equals the difference between 2004 and 2014).
- *Ratio measurement* in addition includes a zero point, enabling the calculation of ratios (e.g. 40 years is twice as much as 20 years).

The measurement levels are extremely important for our discussion. Readers who are unfamiliar with these concepts can find further explanations and illustrations in *Measurement levels*, Supplement to Chapter 2 (http://websm. org/ch2). They can also consult books on basic survey methodology (e.g. Groves et al., 2009) or the statistical literature (e.g. Kirk, 2007).

We should add that the measurement level of the survey question, which is related to the process where respondents are providing answers, is usually reflected in the corresponding measurement level of the variables in the analysis stage, where it is often labelled as the *scale of the variable*. However, the measurement levels of the question and the scale of the variable are still two separate issues and they do not match automatically. For example, responses obtained with questions using the nominal measurement level can be coded and analysed as ordinal scale variables, or responses obtained with questions on the ordinal measurement level can be treated as interval scale variables in the analysis stage.

Understanding different types of questions according to content, response format and measurement level is essential for the survey questionnaire development process and also for the preparation and organization of the survey *datafile* (Section 4.1), which is a term we use for denoting the file with responses. Usually this is a rectangular matrix, where the respondents (units) are in rows and responses (variables) in columns.

### 2.3.1.2 Question Development Process

The question development process starts with the conceptualization and operationalization

of the theoretical constructs that we wish to measure with the survey. While textbooks from different fields (social, marketing, health research, etc.) approach and structure this process in different ways (Bradburn et al., 2004; Iacobucci & Churchill, 2015; Neuman, 2009; Saris & Gallhofer, 2014), we refer to the very general and simplified description of this process provided by Hox (1997). The questionnaire development process starts with an elaboration of a certain *theoretical construct* or *concept* and the definition of its sub-domains, which we also call *dimensions*. We refer to this process as *conceptualization*. For example, in an employee survey the concept would be 'job satisfaction', while further dimensions (sub-domains) could be 'satisfaction with salary', 'satisfaction with managers', 'satisfaction with working conditions'. This is then followed by *operationalization*, where *empirical indicators* for each concept or dimension are searched for and then translated into the actual *survey questions*.

More complexity occurs when several empirical indicators for each dimension are defined. For example, the dimension 'satisfaction with the salary' can be further operationalized with different empirical indicators, such as 'satisfaction with the actual amount received', 'satisfaction in relation to average salary in organization', etc. In the language of survey questions, we typically name these indicators as *items*. Of course, in case we only have one indicator for a certain dimension, such a dimension matches with a corresponding item. In fact, for sake of simplicity in the discussion, we will assume hereafter that dimensions have only one indicator.

Another complication arises when each concept or its dimension is measured for different *subjects*; for example, respondents are answering questions on job satisfaction for themselves and for their partners. In such cases we have separate items for each dimension of each subject.

Questions can be visually presented to respondents as stand-alone questions, or as questions with several sub-questions, typically grouped into tables. The notion of item here covers both stand-alone question and sub-question, whenever it measures a single dimension of a certain concept for a single subject. In statistical terminology the item corresponds to the concept of the variable introduced in sampling (Section 2.2).

Several questions or tables of questions can be further structured into *blocks*, which usually denote a set of questions that form an entity of content-related questions (e.g. a block of socio-demographic questions, a block of job satisfaction questions, etc.). In self-administered survey modes, including web surveys, the questionnaire usually spans across several pages. Each *page* can contain one or more questions, which is a very important methodological decision.

## 2.3.1.3 Question and Questionnaire Design Principles

The basics of question wording and questionnaire design are covered extensively in the literature (e.g. Brace, 2008; Bradburn et al., 2004; Krosnick & Presser, 2010; Saris & Gallhofer, 2014). We summarize here some recommendations from Krosnick & Presser (2010), who provide general suggestions regarding question wording: use simple words (rather than technical terms, jargon or slang), simple syntax, specific and concrete wording, and comprehensive and mutually exclusive response options. Questions should measure one dimension at a time, rather than being double barrelled. Words with ambiguous meaning, leading questions, negations – single and particularly double negatives – should be avoided.

More specific suggestions are related to question types according to their content, which can be found in corresponding topic-specific literature. For example, with behavioural questions, attention to memory and recall problems is needed. Many specifics exist also for sensitive issues, proxy responses (where the respondent answers for another person), longitudinal research and specific groups (e.g. children).

Another set of principles deals with the design of the questionnaire as a whole, from the

structure of the questions, issues of context and the order of the questions to formatting, writing instructions, visual layout, etc.

### 2.3.1.4 Measurement Process and Measurement Errors

When the questionnaire is implemented in the measurement stage (i.e. when responses are filled in) random or systematic *measurement errors* can appear (Biemer, Groves, Lyberg, Mathiowetz & Sudman, 2004; Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz & Trewin, 1997). Measurement errors are defined as the differences between an observed value and a true value. The correlation between the two values is often expressed by the concept of *validity*, which relates to the problem of whether we truly measure what we aim to measure (Saris, 2012, p. 537). The concept of *response bias* is closely related to validity, which reflects the systematic difference between the responses and the true values (Groves et al., 2009, p. 279). On the other hand, the concept of *reliability* expresses random oscillations related to the stability of the responses across repeated measures (e.g. are the attitudes on general life satisfaction stable if asked again in 20 minutes or the next day?), which is also closely related to response variance (Groves et al., 2009, p. 282). Standardized information about validity and reliability scores often exists, for example for questions in the ESS (Survey Quality Predictor tool)[8] or for marketing research questions (Bearden, Netemeyer, & Haws, 2011).

### 2.3.1.5 Cognitive Aspect of the Response Process

When answering a survey question, respondents need to perform specific cognitive processes to provide an answer. Research on these processes started in the 1950s (Biemer & Lyberg, 2003, p. 123) and since then numerous models that structure the response process into several components have been proposed (for an overview see Tourangeau & Bradburn, 2010). The most widely recognized is the model by Tourangeau, Rips, & Rasinski (2000), which comprises four cognitive components:

*comprehension* of the question; *retrieval* of the relevant information from memory; *judgement* of the retrieved information; and finally *response*, with the selection and reporting of the answer. These components also form a basis for studying measurement errors due to respondents and for the development of questionnaire testing approaches, such as cognitive interviewing.

An important descendant of the cognitive approach models is the *satisficing* model (Krosnick, 1999), which deals with various deviations that may occur during the response process. According to this theory, respondents can *optimize* their response behaviour by carrying out all cognitive steps with the necessary effort to come up with an appropriate answer. On the other hand, *satisficing* occurs when respondents perform one or more of the steps superficially. *Strong* satisficing implicates ignoring retrieval or judgement steps, so respondents interpret the question superficially and provide an answer that appears reasonable in the situation, selecting an easily defensible response, such as a status quo option, no opinion, random selection or responses without much differentiation. *Weak* satisficing implicates carrying out all of the four cognitive steps, but less carefully and thoroughly, such as selecting the first acceptable option or showing acquiescence (tendency to agree).

### 2.3.1.6 Specifics of Non-Substantive Responses

In addition to the open-ended response format or substantive response options in the closed-ended questions, we can also offer respondents the option of selecting a *non-substantive response* (e.g. 'don't know', 'no opinion'). However, this can fuel satisficing and other forms of low-quality responding (Krosnick, 1991; Krosnick, Holbrook, Berent, Carson, Michael Hanemann, Kopp, ... Conaway, 2002; Thomas, Uldall, & Krosnick, 2002), with no evidence of an increase in reliability (e.g. Alwin, 2007, p. 199). We thus follow here the recommendation of Krosnick & Presser (2010) that non-substantive responses should in principle be avoided, particularly with non-factual questions. Exceptions can be made when such a response is considered an important

and legitimate option in a specific setting (Manisera & Zuccolotto, 2014) or a question is related to the respondent's knowledge (Sturgis, Allum, & Smith, 2008). Similarly, non-substantive responses may be needed when questions are mandatory, so that respondents are required to select answers to closed-ended questions. When a non-substantive response option is used, it must be visually separated from the substantive response categories (Tourangeau, Couper, & Conrad, 2004), otherwise respondents can mistakenly perceive it as another substantive response category. In web surveys the decision about non-substantive responses closely interacts with the strategy related to reminding or even forcing respondents to reply to certain questions.

## 2.3.1.7 Question Banks

Before starting our question development, it is important that we thoroughly check the work done in the past in the related area. Data archives have thousands of questions from official and academic surveys that have already been used and tested. This is particularly handy for demographic questions, where there is not much new that we can invent that would prove more valid and reliable. In addition, various academic and government agencies have developed resources dedicated to this issue (e.g. Q-Bank[9] or the Survey Question Bank[10]). Similarly, numerous sources and handbooks with lists of questions exist in the field of marketing (Bearden et al., 2011) or psychometric measurement (Spies, Carlson, & Geisinger, 2010), where validity and reliability scores are reported in addition to the wording. We may add that web survey software suppliers are increasingly developing libraries with rich question banks, sometimes even in multiple languages.

## 2.3.1.8 Specifics of Web Questionnaires

Web surveys have few specifics in the early stages of questionnaire development, where we deal with conceptualization and operationalization. However, when it comes to question wording and the layout of the questionnaire,

the specifics of the web survey context become increasingly important, especially the issues of self-administration and computerization.

Due to *self-administration*, the web questionnaire is the main communication tool in a researcher–respondent interaction. Its task is not only to measure, but also to convey the legitimacy and importance of the survey, to provide instructions and to ensure motivation. Self-administration also generally increases the sense of privacy, with potentially positive effects on disclosure of the required information. Web questionnaires require general and computer literacy, which may be a problem for certain population segments. The visual communication is very powerful, so careful elaboration of instructions, format and layout is needed (see Dillman et al., 2009).

Another specific of the web questionnaire arises from *computerization*, which brings numerous advantages, including interactivity and the option of integrating multimedia elements. Specific aspects are related to reading habits: on the web we tend to scan web pages, picking up individual words and sentences (Nielsen, 1997) and read in an F-shape pattern (Nielsen, 2006). This also applies to web questionnaires, as eye-tracking studies report that respondents spend more time at the top of the questionnaire page (Garland, Chen, Epstein, & Suh, 2013), at the beginning of the response options list (Galesic, Tourangeau, Couper, & Conrad, 2008), and answer in a more dispersed way compared with a P&P questionnaire (Fuchs, 2003).

The changing nature of the *technological environment* is indirectly related to computerization, where devices, interfaces and browsers are being continuously transformed, with important consequences for the user experience. Web survey software increasingly incorporates these advances, which makes web questionnaires even more specific. This software is thus particularly important for web questionnaire development, because it may or may not offer elaborate question types, advanced features, fine formatting options, question banks, guidance throughout the process, etc.

In summary to Section 2.3.1, we can say that the development of a survey questionnaire

can be a very complex process, where simplifications, shortcuts, rushing and a lack of professional care can severely damage the entire project. We should thus carefully consider the above-mentioned general issues, as well as the particularities of web surveys, which we systematically review in the next sections.

## 2.3.2 Question Types

We first present key question types used in web questionnaires, together with recommendations on when to use them and in what layout format. We address the general methodological aspects and not the content of questions, which is otherwise an alternative approach for discussing question types, used in various content-specific literature (e.g. Bradburn et al., 2004).

Throughout our discussion we restrict ourselves to the basic web survey mode, as defined in Section 1.1.1, which means the visual display of the questions on the screen and the limited use of multimedia and animation. It also means that respondents provide answers using a keyboard, mouse, touch screen, or some other manual electronic device (e.g. pointer, stylus, etc.). We further limit the manual electronic output of respondents to words and numbers (i.e. to letters and digits). This excludes situations where the respondent is asked to draw something, as well as other alternative methods of survey measurement (see Groves et al., 2009, p. 150) where the respondent is asked to attach a file, take a photo, enclose scans, receipts, etc., This also excludes automatic measurement (e.g. GPS) and various recordings, such as the inclusion of data from external administrative (e.g. results of exams) or business records (e.g. financial data). We briefly reflect on these alternative survey measurements at the end of the section, not because they are unimportant, but because they are not subject to the classification discussed here.

Within this context we understand a survey question as a specific survey measurement method, based on set of words forming a sentence, which then serves as a stimulus for the respondent to address the corresponding concept.

As a result, the respondent then provides an answer through the related cognitive process.

Our approach is to classify questions according to methodological criteria. We thus initially (first level) separate questions with respect to their methodological complexity:

- *Single item questions* address one dimension of one concept for one subject at a time. In this case we have one survey item also matching one column in the datafile and one variable in the statistical analysis. An example would be a question on the age of the respondent in years.

- *Questions with multiple items* can be more or less complex. In the simplest case, we have a table with several sub-questions of the same format addressing more dimensions for a single subject. For example, we can measure the respondent's general satisfaction on three dimensions, such as 'satisfaction with job', 'satisfaction with family life', 'satisfaction with health'. In this case we have three survey items, which then match with three columns in the datafile and with three variables in the statistical analysis. A similar situation occurs when respondents evaluate one dimension of more subjects (e.g. 'satisfaction with health' for three family members). Again, this results in three survey items, three columns in the datafile and three variables in the statistical analysis. More complex combinations are also possible. For example, a table can consist of sub-questions addressing several dimensions of a certain concept for several subjects, such as three dimensions of job satisfaction for the respondents and for their partners. This results in $3 \times 2$ survey items matching six columns in the datafile and six variables in the statistical analysis. Of course, even more complex combinations are possible, with more dimensions and more subjects. However, we still assume here that the respondent deals with each item separately in a sequential manner.

- In addition, *other question types* also exist. Some of them require the respondent to process more items simultaneously. For example, a question may address a concept or several subjects (e.g. ranking more products across some quality dimension). Specific questions also appear in relation to various other combinations, as well as with the extensive use of graphics.

We will further classify single and multiple item questions (second level) according to the measurement level, where we separate the nominal, ordinal, interval and ratio measurement levels as discussed above.

In the next step (third level) we structure each measurement level with respect to key layout implementations of the questions (hereafter denoted simply as *layout*), which depend on the response format, interactive features and graphical appearance of the question. For example, the question about gender on the nominal measurement level can be asked by offering respondents: (a) open-ended text entry to describe their gender; (b) two radio buttons (male/female); (c) a drop-down menu with two options (male/female); (d) numeric entry (e.g. '1' for woman, '2' for man); as well as (e) a graphical layout (e.g. selection of the picture of a man or woman). It would be less appropriate for the situation of the nominal measurement level where we seek dichotomy (male/female), to use (f) scale. However, in some research we may ask respondents to express the perception of their own gender on some scale (e.g. 1–100, 1–10 or 1–5), between the extremes of male and female self-perception.

The literature on web surveys often begins the classification of question types according to the standard HTML elements (e.g. Bethlehem & Biffignandi, 2012; Couper, 2008; Tourangeau, Conrad, & Couper, 2013), which are involved as response options to certain question: radio button, checkbox, drop-down menu and text input. We acknowledge the importance of these HTML elements too, but we consider them only at the final (i.e. third) level of question type classification, when we discuss the layout. Here, due to the prevailing methodological approach of this book, we have rather decided to conceptualize single item questions primarily according to their measurement level. There are four main reasons for this:

- Firstly, it fully corresponds to the substantive activities of the question development process, where the measurement level is initially selected according to the research aims and according to the content of the question. Often, the research aims and the content already determine the measurement level. Only when the latter is selected can we start discussing the layout, which then includes a dilemma about the response format (open- or closed-ended questions), the selection of HTML elements and other graphical aspects.

- Secondly, the measurement level is important because it also reflects the essential specifics of cognitive processes on the respondent's side. The respondent needs first to understand what kind of measurement precision is required (nominal, ordinal, interval or ratio). Each of these levels then requires a specific cognitive process, while the layout has relatively little impact.

- Thirdly, the measurement level determines to a considerable extent the format and structure of the related variables in the datafile, and consequently also the nature of post-fielding processing. All this is extremely important for automatic analysis in preliminary reporting. For example, the means cannot be calculated for nominal measurement.

- Fourthly, the measurement level has permanent conceptual features, rooted in the nature of human cognitive processing, while the role of HTML elements changes with usability and technological developments. This is particularly true for devices that use smaller screens, where the role, importance and perception of HTML elements are modifying.

In the following discussion we identify key question types, classified as described above, address their essential methodological aspects and provide guidance for their use in web surveys.

We are fully aware how difficult it is to follow a discussion on question types without corresponding examples. We have provided an interactive overview of this in *Questions and layouts in web surveys*, Supplement to Chapter 2 (http://websm.org/ch2), which also includes a default statistical analysis related to certain question types.

## 2.3.2.1 Single Item Questions

The simplest type of survey question addresses a single dimension of a single concept for a single subject, which then results in one item. According to the response format, these questions can

be open- or closed-ended. With open-ended questions, the respondent is in principle free to write any kind of response in *text* or *numeric* form, while with the closed-ended format, the respondent needs to select one answer from a list of categories which should be exhaustive and mutually exclusive. The selection can be done with a click (touch), but also by entering the corresponding number (or letter) that denotes the selected option.

Closed-ended questions, where the respondent selects one answer from a list, are also called *single answer questions*. They are different from *multiple answer questions*, where respondents can choose more than one answer from the list of categories, often in the format of checkbox. However, this is essentially a question with multiple items, because each of the response options counts as an item and needs to be treated by respondents separately. Each response option also gets a separate column (i.e. it is a variable) in the datafile (typically with values '1' when selected and '0' when not selected). As such it is thus addressed within the discussion of questions with multiple items.

### 2.3.2.1.1 Questions With Nominal Measurement

With nominal measurement, responses belonging to a question can be distinguished, but they cannot be sorted according to some substantive dimension, which would be intrinsic to the question. As a consequence, when transferred into the measurement level of the corresponding variable, very limited univariate statistical analysis can be done besides compiling a frequency distribution. The separation between open-ended text entries and the closed-ended format (with different layouts) is particularly critical here.

**Open-ended text entries** Open-ended text entry questions allow for the collection of responses in text format. We can ask for a *longer narrative answer* (e.g. comments, explanations), for a *shorter narrative answer* (e.g. name) or even for *non-narrative answers* (e.g. URL, email address, CAPTCHA codes, etc.).

Open-ended text entries are the least structured measurement level of survey responses.

We could even say that formally they hardly belong to the nominal measurement level.

As a consequence this type of question can also be exempted from the discussion on the nominal measurement level (Saris & Gallhofer, 2014). That is, it is difficult to talk about the nominal level of measurement because when the respondent is providing answers, there are no categories related to nominal measurement to assist the measurement process. Similarly, when we are dealing with data in a raw format (before cleaning and coding), the responses to the open-ended text entry questions can only be distinguished among themselves. Usually, every response is different, so the number of different responses simply equals the number of respondents (i.e. number of units). Thus, we do not obtain nominal values, which can be statistically analysed as a frequency distribution. This is only possible after a researcher has cleaned and coded the answers into a smaller number of categories in the post-fielding step. The measurement level usually remains nominal, but it can also change. For example, if we rate the open-ended responses according to certain criteria (e.g. less to more favourable) during coding, we may end up with ordinal measurement. However, this occurs only in post-fielding: at the measurement stage itself the question offers no measurement levels to the respondent.

The coding process for open-ended text entries means we assign the same number code to the same characteristics of responses. For example, in questions asking the respondent to comment on a new service, all responses which praise the speed of a new service receive the same code. We discuss coding in Section 4.1.5 of the post-fielding step. Below we only present a few approaches which can be undertaken when formulating and formatting the open-ended text entry questions and which can facilitate the coding:

- We should ask the open-ended questions in a structured way whenever possible. For example, we ask respondents to fill in their first name and their family name as separate entries. Similarly, it is advisable to ask about advantages and disadvantages separately.

Sometimes it is also useful to ask respondents to write down, say, three key advantages and three key disadvantages.

- Explicit requests for structuring can also be formalized with separate entry fields, in particular when we look for lists (e.g. for naming three top restaurants, we present the respondent with three separate entry subfields). However, this is then already a multiple item question.
- Open-ended text entries can suggest a certain structure with additional specifications. For example, the respondent can be asked to complete a sentence.
- For text entries related to specific non-narrative answers, we can impose further structuring and formatting with symbols and subfields (e.g. two fields with @ in between for the email address), various input character restrictions, labelling, instructions and validations, including external ones (e.g. checks with a database of ZIP codes or URL domains). As respondents are sensitive to the visual layout of questions, a careful design of visual layout encourages answers in the desired format. Couper, Kennedy, Conrad, & Tourangeau (2011) and Dillman et al. (2014) provide further practical guidelines for the presentation of open-ended text entry fields, such as labelling of input fields, structuring questions into components, providing templates, etc.
- Open-ended text entry can also be used as a supplement to closed-ended questions. Unless we have a list of all possible categories (e.g. administrative regions), it is useful to add the open-ended option 'Other, please specify'. For example, when asking about the respondent's favourite fruit, we identify 15 of the most popular fruits and capture others with the open-ended option.

The entry fields for open-ended text questions can be of different sizes. When only one or a few words are expected (e.g. name, favourite fruit), the entry field can be shorter and one full line is enough (e.g. 30 characters), which is called a *text box*. A text box is also used with the 'Other, please specify' option, offered as a final category in various closed-ended response formats. When we expect longer text (e.g. comments, explanations), more width (e.g. 60 characters) and more lines

(e.g. five) can be offered. The enlarged text box is commonly known as a *text area*. Web survey software usually allows for the specification of the number of rows of the text entry field, its width and a maximum number of characters allowed.

The size of the text entry field is important for the respondent. It has been determined (Dennis, de Rouvray, & Couper, 2000; Smyth, Dillman, Christian, & Mcbride, 2009) that larger entry fields in web questionnaires increase the length of the narrative response. However, when a respondent is faced with a larger text field, the perception of the burden and the related nonresponse may increase (Zuell, Menold, & Körber, 2015). Alternatives to this are the auto-adjustment of the text field size, also called scrollable boxes (where additional lines automatically appear when the respondent comes to the bottom), or the option for respondents to expand the size of the text field by dragging its lower right corner. The size of the box can be also customized according to previous answers (Emde & Fuchs, 2012b). Unfortunately, these options may not be supported by all software and have not been extensively tested.

Other strategies to obtain richer responses in web questionnaires are strong motivation statements (Smyth et al., 2009) or follow-up probes (Holland & Christian, 2009; Oudejans & Christian, 2010). For example, after providing an open-ended answer, the respondent receives a message acknowledging the response and at the same time is also asked a follow-up question about whether the respondent would like to add any other issues.

Studies have confirmed the advantage of text entries in web questionnaires, which produce longer, more detailed and more revealing responses compared with self-administered P&P questionnaires (Barrios, Villarroya, Borrego, & Ollé, 2011; Kiernan, Kiernan, Oyler, & Gilles, 2005).

We should be aware of the fact that open-ended text entries generally increase the respondent's cognitive burden and require a longer time to respond. This may then impact on the number of low-quality responses (e.g. invalid answers), breakoffs (preliminary terminations of surveying) or item nonresponse (omission of response to certain question), although evidence shows that

this does not happen more often with web surveys than with P&P self-administration (Barrios et al., 2011; Kiernan et al., 2005; Kwak & Radler, 2002). Furthermore, text entries limit the quantitative analysis and involve coding, which is inconvenient, subjective, time consuming and expensive.

Whenever an exhaustive list of response options can be specified – and respondents know all of them – we recommend using a closed-ended format. Of course, there are cases when open-ended text entries are unavoidable. Typical situations include too many response options (e.g. names), a degree of complexity that is too high to be structured (e.g. complex sequence of certain actions), or when the options are impossible to foresee (e.g. preliminary enquiry about the most important problems in a certain context). Thus, the decision to use this question type requires a researcher to carefully consider substantive and methodological circumstances; a useful discussion can be found in Krosnick & Presser (2010) and in Saris & Gallhofer (2014). We also caution that this decision should be taken very responsibly, because it can affect different outcomes. For example, Reja, Lozar Manfreda, Hlebec, & Vehovar (2003) found considerable differences when the question of the most important Internet-related problems was asked with an open-ended text entry type compared with a closed-ended one.

**Radio buttons** Radio buttons are the default option with closed-ended questions requiring one answer from a list of answers for nominal measurement; this is also the default option to be considered with nominal measurement level in general. Radio buttons are the visual equivalent of the P&P questionnaire format and are thus preferred for the *unified design* used in several survey modes (Section 2.1). An additional advantage is that this is an HTML element which appears in many other web contexts (e.g. forms, bookings, payments, etc.) and respondents instantly recognize its function for the selection of only one option. The problem with the HTML radio button format is its

rather small size, which cannot be increased or changed. The selection can require a high degree of precision with the mouse, which can be problematic for certain purposes and populations (Lumsden, 2007). With respect to further layout possibilities, radio buttons can be presented horizontally below or alongside question text, with response option labels usually placed to the right (at least in Western cultures), but can also be placed below or above the radio buttons. The radio buttons can also be placed vertically, which is usually the only option for small screens. To maintain comparability across screens this is then the default recommendation whenever possible. Research has shown that these variations (i.e. horizontal–vertical) have few effects (Toepoel, Das, & van Soest, 2009) even with ordinal measurement, so with nominal ones we can expect them to be even fewer. The same research also showed that grouping response options in more columns of radio buttons is not beneficial; however, we may not be able to avoid it with a large number of categories. The alternative in such a case is an *answer tree*, which means structuring the responses in more steps (e.g. we first select a continent and then the corresponding country).

With radio buttons it is advisable to have a feature that allows unchecking (deselection) when needed. However, not all software supports this, so once the respondent selects a certain radio button, it may not be possible to unselect it and leave the question unanswered.

We may add that when we have items with only two response options, also called *dichotomy* (e.g. YES/NO), a specific alternative of a checkbox layout appears. Here, marking the checkbox means YES and leaving it unmarked implicitly means the answer is NO. While a stand-alone checkbox can be used in registrations and administrative forms, it is highly inappropriate as a layout for web survey questions, because of ambiguities when left unchecked. On the other hand, this can be a specific layout alternative for certain series of dichotomous questions. We therefore discuss this layout later, in the context of multiple item questions.

**Drop-down menu**   A drop-down menu, also known as a pull-down menu, drop-box, drop-down list or pull-down box, is another popular format for displaying a list of possible responses in closed-ended questions with nominal measurement. It appears in many CASIC modes, but has no comparable option in P&P modes. It is also a standard HTML element commonly used in everyday web browsing, so respondents are familiar with it. It is especially suitable for handling very long lists (e.g. hundreds of countries, year of birth) and should be considered when a radio button layout cannot fit on one typical screen.

Drop-down menus can be time consuming, especially when they include very long lists (B. Healey, 2007). The *autocomplete function* – also called *database lookup*, which is a feature predicting a word or phrase from the list in the drop-down menu after typing in the first few letters – can help here with rapid access to the desired response. For example, Couper, Zhang, Conrad, & Tourangeau (2012) successfully implemented this function in selecting medication from a list of thousands of entries.

When designing questions in the drop-down menu format, it is very important that the initial visible part of the responses is limited to the first line which is labelled 'click here' or 'select from the list', rather than being empty or filled with dashes '---', while offering the first response option should not be used at all.

In case of a very large number of response options the drop-down menu can use a series of related questions where each subsequent question is displayed conditionally on the answer of the previous one. For example, to select a model of car, the respondent first selects the brand, then the model within the selected brand. Here, the drop-down menu of the 'model' adapts interactively to the options selected in the 'brand' menu. This can be done in a single line, instead of tens of lines with an alternative 'answer tree' layout of radio button questions, which requires a sequence of conditional pages. This layout is sometimes referred to as a *drill-down* menu.

In comparison with the radio button layout, drop-down menus have several limitations. When used as a standard HTML element, we cannot add the open-ended option 'Other, please specify' nor the *content search*, which is very valuable if we have many options and the respondent knows the answer (e.g. country of residence). Additional scripts and browser functionalities (such as a combo-box for editable drop-down menus, where respondents can add an option that is not on the list; Couper, 2008, p. 120) can handle these deficiencies, but it is sometimes unclear if users always recognize these extra options.

A more serious problem with this layout format is related to the fact that the respondent does not see all the options in advance before the menu opens, which can be critical when the respondent is not very familiar with all the options.

Studies show a series of disadvantages with drop-down menus in terms of data quality: from stronger primacy effects, by which respondents more often click on response options at the top of the list (Couper, Tourangeau, Conrad, & Crawford, 2004), to increased response times due to the additional click needed to display the list (B. Healey, 2007). In common situations, when we have enough space and not too many response options, drop-down menus provide no advantages over radio buttons (Tourangeau, Conrad, & Couper, 2013, p. 93), so there is very little justification to use them.

**Numeric entry**   Sometimes we can use closed-ended questions with a numeric entry, where the respondent enters a number that corresponds to a certain response category at the nominal measurement level from a closed set of options. An example would be entering a certain well-known number that denotes a region, postal number or ZIP code. We may also ask respondents to put corresponding numbers (e.g. '1' for male and '2' for female), but very good reasons should exist for this, because it is much less convenient than the alternative with radio buttons. Nevertheless, specific benefits may also exist in situations when the mouse cannot be used effectively, or when the web questionnaire serves for the purpose of recording observations with an external code list. This is commonly used with data entry of P&P questionnaires, as well as in CAPI and CATI surveys.

Compared with the selection of categories with a click (or touch), entering a number is slower, prone to errors and less intuitive. In addition, entries outside the range of response options should be disabled, which we discuss in more detail with interval and ratio measurement, where an open-ended numeric entry is the default layout format. We may add here that instead of digits, we can also enter letters to denote a selected category (e.g. 'A', 'B' … for school grades, or 'M' for male and 'F' for female).

**Advanced graphical presentation** Closed-ended questions with nominal measurement can be implemented also by using various graphical presentations. For example, radio buttons can be replaced with special visually designed buttons (e.g. larger images of radio buttons) or other images (e.g. stars). Alternatively, verbal descriptions of the answers can be fully or partially replaced by certain graphical or multimedia elements (e.g. audio, video), for example when selecting the best video, the best logotype, etc.

In extreme examples, both the radio buttons and the verbal answers can be replaced with direct visualizations of response options (e.g. pictures of different types of fruits when selecting the best fruit), where the selection is done with a direct click on a specific picture (without the radio button). A specific case is a question with image area selection, *clickable image maps*, where the respondent selects a point on a prespecified area of the picture, for example a region from a map or a body part from a picture of a human. This format is particularly frequent in education (e.g. quizzes, e-learning) and marketing research, where it is sometimes called *hotspot question*.

An advanced graphical option is the use of *drag and drop* functionality, which is very specific and also popular in web surveys, particularly in marketing research. In the case of the nominal measurement level this would require the respondent to select a certain category (subject, usually presented with a picture) with the mouse, and drag and drop it into some preselected location. In this specific example there is an obvious increase in the mouse movements needed, compared with the simple alternative of one-click selection of the corresponding category. The arguments for using it are usually related to increased involvement and engagement of the respondents.

While radio buttons are commonly presented across different technologies (devices, browsers and interfaces) in a standardized manner, the alternative graphical presentations may not be uniformly supported. Nevertheless, we have observed an emerging trend of new graphical question formats, especially with mobile devices.

We will present further examples later on when discussing the extensive use of graphics, where special care is needed with regard to the possible unintended effects of graphics on respondents' answers.

## 2.3.2.1.2 Questions with Ordinal Measurement

There are substantive and also methodological differences between a closed-ended category response at the nominal measurement (e.g. select a country of residence) or at the ordinal measurement, such as frequency (e.g. never, sometimes, often …) or agreement (strongly disagree … strongly agree). In contrast to the nominal measurement, the ordinal measurement enables the calculation of a median, a minimum, a maximum and Spearman's correlation coefficient. In practice, results from ordinal measurement are sometimes even treated as if they come from interval measurement, so that a mean and variance can also be calculated. If we further assume a normal distribution for the responses, a full array of multivariate statistical methods can be applied.

However, this important difference is not necessarily reflected in the visual layout of survey questions. For example, the same question layout with radio buttons is used in both cases. Below, we highlight the methodological specifics of the same layout implementations, but when used for the ordinal measurement.

With the ordinal measurement level, both factual and non-factual questions can be used. With non-factual questions (e.g. attitudes), we frequently encounter questions labelled as *rating scales*, where respondents rate (i.e. assign values

from a closed list of ordered categories) an underlying concept. We distinguish between *unipolar* concepts (e.g. strength, from low to strong) and *bipolar* concepts (e.g. agreement, ranging from extreme agreement to extreme disagreement).

We may add that the rating scales are often labelled as Likert scales, particularly when denoting a 5-point rating scale. However, we avoid this labelling as it is too often used differently from what Likert originally intended (Neuman, 2009, p. 207).

In practice, rating scales with five response categories or more are often treated in the analysis as the interval measurement level. The prerequisite for this is that the visual distances between adjacent categories are equal, so as to resemble the interval measurement level. Within this context, efforts should be made also to obtain a normal distribution. For example, if all respondents tend to 'strongly agree' when asked about agreement with the statement 'The food was very good', we may consider rewording the question to a stronger assertion: 'The food was exceptionally good'.

Our discussion of question layout implementations within the ordinal measurement level will predominantly focus on rating scales; however, it will also cover other possibilities within the ordinal measurement related to factual questions (e.g. frequency, observed ranks, etc.).

**Radio buttons**   Radio buttons are the prevailing layout format for rating scales and for the ordinal measurement in general. Radio buttons share all the characteristics we discussed in the context of nominal measurement. In addition, rating scales also face many general methodological dilemmas that originate in traditional survey modes, but still have certain specifics in the case of web questionnaires. As they are very important for scale constructions, we briefly present their essential characteristics:

- Response categories can be numbered (e.g. 1, 2, 3, 4, 5) or can be verbal (e.g. agree, disagree …), or both labelling principles can be used. In specific situations, when numbers are used, they can also be negative (e.g. −2, −1, 0, +1, +2), although this is generally not recommended

because the respondent may tend to avoid negative values (Toepoel et al., 2009). Other combinations exist, for example only end points can be labelled (strongly agree and strongly disagree), while the remaining points are not labelled. The decision depends on the situation, but in general full verbal labelling takes precedence (Dillman et al., 2014; Krosnick & Presser, 2010, p. 275).

- There is an indication that the vertical layout of response categories is recommended over horizontal ones (Krosnick, 2013), but the question of whether this outweighs the required additional space remains.

- In the vertical format, a decreasing order of response categories (e.g. from good to bad) is recommended, because respondents expect positive things first, and because it is more conventional and thus imposes less cognitive burden on respondents (Holbrook, Krosnick, Carson, & Mitchell, 2000; Tourangeau, Couper, & Conrad, 2013). This would mean in principle that we should start with a positive category also in a horizontal format. However, this is contrary to the analogy of the coordinate system used in mathematics and also in the everyday practice of measuring length, time, weight, etc., where the scale goes from low to high (e.g. from 0 to 100). There is little evidence that the orientation of response categories is important for the horizontal layout. However, Toepoel et al. (2009), for example, found small, yet significant differences, with some evidence of primacy effects for the vertical format.

- An odd number of points is recommended (e.g. 5 or 7) so that a mid-point response option exists (Krosnick & Presser, 2010), except in specific circumstances.

- With respect to the number of categories, research has shown that more than seven categories result in smaller gains in data quality, or even in deteriorated quality: see Malhotra, Krosnick, & Thomas (2009) who recommend 5-point scales for unipolar concepts (e.g. from low to high strength) and 7-point scales for bipolar concepts (e.g. from extremely dissatisfied to extremely satisfied). However, in certain circumstances, other alternatives may be justified, particularly 2-, 3-, 4-, 6-, 10-, 11- and even 101-point scales. Some important surveys (e.g. the ESS)[11] use 11-point scales. The research is still inconclusive, since justification exists for a low

number (e.g. Alwin, 2007) as well as for a high number of points (e.g. Saris & Gallhofer, 2014).

With respect to verbal labels, it often seems very practical to use wording with commonly used scales (e.g. agree–disagree), although research (Krosnick & Presser, 2010) sometimes shows certain advantages (less acquiescence, higher reliability and validity) of content and construct-specific scales. For example, instead of asking for agreement (disagree–agree) with the statement that the food was tasty, we ask the respondent to rate the tastiness from extremely un-tasty to extremely tasty. The same is also true for other question types, so natural metrics should be used instead of general ones. For example, we ask about the frequency of some events specifically (daily, weekly, monthly …) instead of using commonly used categories (very often, often, not so often …).

The above issues originate in the P&P self-administered questionnaire, but are equally relevant for web surveys. Special care is needed when it comes to the visual presentation of questions and response options, because respondents also consider spacing between categories (i.e. radio buttons) when interpreting distances between them. So, if we plan to use the responses from the ordinal measurement with radio buttons as interval ones in the analysis, equal spacing between categories should be ensured (Tourangeau, Conrad, & Couper, 2013, p. 79). This can be compromised if web survey software adapts the width of the columns to the width of the response labels and therefore introduces unequal widths across the scale.

There are additional specifics in web surveys as regards questions with non-substantive responses ('don't know', 'no opinion'), due to the potential interaction with real-time validations and prompts, which we discuss in Section 2.3.3.

Various rating scale subtypes exist, such as *compare one against another* (when two options are offered horizontally), *semantic differential* (where only end-point options on a horizontal bipolar scale are verbally labelled, such as ugly–beautiful) or *Stapel scale* (unipolar 10-point vertically oriented scale, numerically labelled from −5 to +5 without a middle point).

**Drop-down menu**   Drop-down menus share the same reasons against their use as outlined in the discussion on nominal measurement, but to an even greater extent. Since the number of categories is usually relatively small in rating scales and other questions with ordinal measurement, one of the key justifications for the use of drop-down menus – the convenience when having a large number of categories – disappears.

**Numeric entry**   As with nominal measurements, respondents can enter their answers in closed-ended numeric entries also for the ordinal measurement. The same limitations and specifics as with nominal measurement apply. However, in some situations, entering the numbers, such as school grades (e.g. 1, 2, 3, 4, 5), is very simple and intuitive. Of course, entries outside the range of response options should be disabled.

**Advanced graphical presentation**   The graphical presentation of questions with ordinal measurement is the same as in the nominal measurement when graphics and multimedia can replace radio buttons, response labels, or both. However, more specific alternatives appear here, such as smileys or thumb fingers. For example, faces with different expressions (from least happy to most happy) can be used to present satisfaction with a service. There are some positive reports on using images in rating scales, particularly in psychological research measuring pleasure, arousal and dominance (SAM scales), where images were found useful (Bradley & Lang, 1994). On the other hand, Emde & Fuchs (2012a) did not find any considerable advantages of using faces in web surveys, just delays and problems in response distributions. We raise further concerns against the use of various graphical presentations in the general discussion on graphics and multimedia later on.

Another graphical option used to present possible responses in ordinal measurement is graphical lines or ribbons, with a discrete number (e.g. five) of points in a graphical format on a line, where values can be selected. By this we do not mean a full continuous scale, because this

is used for the interval and ratio measurement level, as discussed below.

## 2.3.2.1.3 Questions with Interval and Ratio Measurement

Here, we jointly discuss questions with the interval and ratio measurement level. In comparison with ordinal measurement, interval measurement enables us to compare differences among values. Ratio measurement relates to numeric quantities (e.g. age, income, weight, height, speed, time, etc.) where – in addition to the characteristics of the interval level – there is also a zero point, which enables the comparison of response values as ratios. In principle, the ratio is the most preferred measurement level, because it enables the broadest array of statistical methods and because, in addition to general univariate and multivariate statistics, relative measures can be used (e.g. coefficient of variation).

The reason why we jointly discuss the two measurement levels is that additional specifics of ratio measurement over interval one are very small, at least at the measurement stage. There are also almost no methodological specifics and differences between the two measurement levels in terms of question development and layout implementation.

We have already mentioned that interval measurement can sometimes be assumed in the analysis stage for rating scales and other questions on the ordinal measurement level.

**Open-ended numeric entry**   Open-ended numeric entry, where responses are typed in as digits, is the most common layout option for questions at the interval and ratio measurement, because it provides high precision. In principle, we expect here numeric entries of infinite possibilities (e.g. salary). However, in reality the border between open-ended and closed-ended numeric questions – where we actually enter a number from a rather closed set of options (e.g. year of birth) – is sometimes blurred.

The size and format of entry fields are very important here, as in open-ended text entries. Tailored entry fields (e.g. a two-digit entry field instead of a ten-digit field, if we ask for the number of children) can result in more precise answers (Couper et al., 2011; Couper, Traugott, & Lamias, 2001). It is thus useful to structure and elaborate the entry fields by specifying the format (e.g. explicit currency signs $, €, £ ...), decimal points, length and subfields (e.g. year–month–day, hours–minutes–seconds). Real-time validations (see Section 2.3.3) are essential here to prevent numeric entries outside the range (e.g. an age of 572 years).

**Radio buttons and drop-down menus**   Closed-ended response formats, such as radio buttons and drop-down menus, are in principle inferior to the open-ended numeric entry, which is usually more precise and neutral (Krosnick & Presser, 2010, p. 267) when it comes to interval and ratio measurement (e.g. salary). However, when we have a very limited number of response categories (e.g. number of children), the radio button layout can be faster. With a higher number of categories (e.g. age), a drop-down menu can be more error-free. In addition, sometimes we may consider categories with aggregated values across the expected range of the scale. For example, knowing that we will only use the age variable coded into three age groups (e.g. <30, 30–49, 50<) in the analysis (thus only at the ordinal measurement level, which means that we lose measurement level precision) it is better to offer immediately a closed-ended question with the three age groups instead of asking for the year of birth and recoding answers later. In general, closed-ended questions are also considered less burdensome for respondents. However, any grouping of values in advance needs to be done very carefully, since there is plenty of evidence on how improper categorization of numeric quantities in the closed-ended format can skew the results (Dillman et al., 2014, p. 161).

**Continuous scale** A continuous scale can be used in principle only when the response values formally take an infinite number of values on some continuum, instead of a finite number of closed-ended response categories. More specifically, the respondents need to be able to observe and express answers on very detailed interval and/or ratio measurement levels, which is however rarely the case. With attitudinal questions this can be a very controversial issue, so a researcher has to decide whether certain concepts belong to the ordinal or interval measurement. Of course, as soon as we implement a continuous scale, we assume that we have an item on interval or ratio measurement level.

Presentation of the continuous scale requires a simple graphical option where respondents denote their position on a line. The approach originated and raised controversies in traditional surveys in the form of the so-called *line production* (Saris & Gallhofer, 2014, p. 109), with the aim of providing better precision and particularly for avoiding the rounding of numbers (e.g. multiplies of 10 or 5), which often occurs when we ask for numbers in the open-ended numeric format. The advantage in web surveys is that the graphical response is automatically transformed into a numeric value (e.g. 1–100), which can be optionally displayed to respondents in real time.

In web questionnaires the continuous scale most frequently appears as the *visual analogue scale* (VAS) – predominantly in health or marketing research – with a horizontal line and two end-point verbal labels (e.g. agree–disagree). The respondent clicks with the mouse on the selected point (Reips & Funke, 2008). When we add labels (e.g. weak pain, medium pain, strong pain) to other parts of the line, or divide it into numbered segments, this is called a *graphic rating scale* (GRS). The selected position can be additionally marked with a handle so that the respondent drags it left or right. This is sometimes called a *slider bar*. Presentation of the slider bar can be problematic if the handle is positioned anywhere in the scale before the respondent selects an answer (e.g. in the middle of the scale). In such cases the intentional middle response cannot be distinguished from the situation where the respondent omits a response (Funke, Reips, & Thomas, 2011, p. 223).

Continuous scales are sometimes reported to be superior to radio buttons (Funke & Reips, 2012; Reips & Funke, 2008; Torrance, Feeny, & Furlong, 2001), because they are more precise and provide fewer extremes and mid-point answers. Continuous scales are also advocated by Saris & Gallhofer (2014), however, they also advise caution when it comes to implementation, where care should be taken whenever respondents differ strongly in their perception of the question. There is also research that shows the disadvantages of continuous scales. For example, Couper, Tourangeau, & Conrad (2006) showed that continuous scales took longer to complete than using a radio button layout, had higher breakoffs and a higher level of missing data, as well as a higher level of rounding when numeric feedback was provided. Similar problems were found in psychological research (Flynn, van Schaik, & van Wersch, 2004) and in Funke et al. (2011), especially with less educated respondents.

It thus seems that continuous scales may work well only with fully computer-literate respondents and for concepts where respondents can truly separate very fine nuances. Other than that, in general, continuous scales have few advantages and various disadvantages, so their usage requires very explicit justification.

**Advanced graphical presentation** The continuous scale itself already involves graphics; however, in addition to simple VAS, GRS and the slider bar, described above, many other graphical presentations can be included, from pictures replacing verbal descriptions and figures that appear on the scale to various animations and multimedia. In general, as with other question types we advise caution when implementing graphics.

With this we conclude our review of key layout formats for the corresponding measurement levels with single item questions. Table 2.1 summarizes the possible combinations. Filled circles mark the options which in our discussion are to be considered as default ones, while the open

**Table 2.1  Key layouts for single item questions across levels of measurement**

| Level of measurement | Key response layouts | | | | | |
|---|---|---|---|---|---|---|
| | Open-ended text entry | Radio button | Drop-down menu | Numeric entry | Continuous scale | Advanced graphics |
| **Nominal** | ○ | ● | ○ | ○ | | ○ |
| **Ordinal** | | ● | ○ | ○ | | ○ |
| **Interval/ratio** | | ○ | ○ | ● | ○ | ○ |

circles mark the options that may be suitable in certain circumstances but their use requires some explicit justifications. Other options are not considered possible or reasonable (e.g. continuous scale for nominal level of measurement).

Some of the combinations related to open circles are very rare, such as numeric entry for the nominal level, while others are more frequent, such as open-ended text entry for the nominal level. With the latter we may repeat that, so far, we have discussed only single item questions, which excludes any series of open-ended text entries (e.g. for collecting the lists). Similarly, we referred here to the measurement level at the stage of obtaining answers from respondents and not to the measurement level at the stage of statistical analyses (i.e. scale of the variable) where, for example, responses in open-ended text entry questions can be further coded and treated as the ordinal scale. The same situation occurs when data are collected with ordinal measurement, but the corresponding variables are then treated in statistical analysis as being on an interval scale.

## 2.3.2.2 Questions with Multiple Items

In practice we often group questions with single items to save space or to speed up the respondent's task. The idea of saving space originates from the P&P mode, with the goal of reducing printing costs. In web surveys, space itself is no longer a direct limitation. The increased speed, together with the alleged decrease in the respondent's burden when similar questions are grouped together, remains the main argument for grouping questions. In addition, we may group questions to add the same context to them. The potential advantages need to be weighed against the danger of lower data quality, which might result from increased complexity, compared with sequences of single item questions.

The grouping of single item questions results in questions which contain sub-questions; we thus talk about questions with multiple items. Such groups of questions are sometimes labelled as a *matrix* (Dillman et al., 2009, p. 179) or *grid*. However, this latter term often denotes a narrower combination of items with the same response options (Tourangeau, Conrad, & Couper, 2013, p. 72). A closer look at the discussions on grid questions reveals that they predominantly address the rating scales with the radio button layout. Another alternative labelling for a series of items with the same response option is a *battery* of questions (Alwin, 2007; Saris & Gallhofer, 2014, p. 86). We prefer to use here the most general notion of a *table* to denote any grouping or combination of questions.

Tables present many methodological challenges. We first discuss *simple tables* with a series of homogeneous questions for several dimensions or for several subjects, where individual questions measuring one item at a time, using the same implementation layout, are grouped together. The term roughly matches practical usage of the notions of battery and grid mentioned above. We thus discuss simple tables with homogeneous questions. For the sake of simplicity – unless explicitly denoted differently – we will talk simply about tables. Later we will reflect on more complex table types.

Again, examples and illustrations for the tables discussed below are presented in the Supplement to Chapter 2, *Questions and layouts in web surveys* (http://websm.org/ch2).

## 2.3.2.2.1 Tables with Questions at Nominal Measurement

We first observe the grouping of similar single item questions which use nominal measurement. The layout is very important here, as it determines the format of the tables.

**Tables of open-ended text entries**  When similar items with open-ended text entries are combined into a single row (or column), we get a one-dimensional table, so we prefer to talk about a *series* or an *array* of text entries. For example, we can put three text entries in one line for name, family name and address. In this way we measure three items, resulting in three columns in the datafile and three variables in the analysis. The advantage is that this saves space and eliminates the need to repeat the introductory question text. The labels (e.g. name …) can be added close to the corresponding entry field. When a series of text entries is repeated for five family members (subjects), each in a separate row, we have a two-dimensional table of text entries. In the case of five persons, we have $5 \times 3 = 15$ items measured (15 columns in the datafile and 15 variables in the analysis), packed into a neat interface, which saves a lot of space and is intuitive and user-friendly. Such layouts often appear in household rosters or in social network measurement as the *alter-wise* layout, because we take each alter (e.g. friend) and then assign corresponding answers (values) for all items (name, family name, address). An alternative is the *item-wise* layout, where we have rows for each item (e.g. name, family name, address) and we then assign values for all alters (e.g. friends). Unfortunately, it is not entirely clear whether – and for what complexity – the table layout with open-ended text entries provides an overall higher or lower data quality compared with the corresponding series of individual questions for each family member (Dillman et al., 2009, p. 180).

**Tables of radio buttons**  When respondents select their employment status from three available categories (e.g. employed, unemployed, retired) in a horizontal radio button layout, this is a single question measurement for one item. Reporting the same category for more subjects, for example five family members, can be conveniently done in a table, with family members listed in separate lines. We then need only one introduction and one set of response labels on top. Although this appears as a two-dimensional table with $5 \times 3 = 15$ radio buttons, only one dimension (employment status) is measured on five subjects (family members) and thus only five entries are required. Therefore, only five items are measured, resulting in five columns in the datafile and five variables in the analysis. Compared with five separate questions, this approach seems faster for the respondent and saves space in the questionnaire. Nevertheless, these advantages might be deceptive, because increased complexity can reduce the attentiveness of the respondents. We return to this issue when discussing tables with ordinal measurement. Of course, this danger also increases if we squeeze in more dimensions to obtain so-called *double tables* or *triple tables* of radio buttons, for example by adding radio button questions for gender for five family members to the same table. This is then a real two-dimensional table measuring $5 \times 2 = 10$ items (two socio-demographic characteristics for five family members).

**Tables of drop-down menus**  If we put employment status in the above example into a drop-down menu layout, this will rarely be advantageous, as already discussed with drop-down menus measuring only one item. However, if we need to save space, the drop-down menu layout enables clear presentation of additional dimensions in a row. For example, we can have a row with drop-down menus for employment and gender for each subject (e.g. family member). As above, this results in a two-dimensional table measuring $5 \times 2 = 10$ items. This is a very complex layout with

a realistic fear of excessive burden and reduced attentiveness among respondents, so good justification is needed.

**Tables of numeric entries**   Employment status in the above example can also be entered with the corresponding numeric entry (e.g. '1' for employment, '2' for unemployment, '3' for retirement) in a series of closed-ended numeric entries for each family member. However, this could be difficult to justify, due to the objections we have already raised in the discussion of questions with single items. To a lesser extent – due to effective space-saving – the same disadvantages hold true for 5 × 2 table of closed-ended numeric entries (two dimensions for five subjects), where the above discussion on drop-down menus also applies.

**Series of dichotomous questions**   When we have a set of questions with a dichotomous status (YES/NO) – typically asking about evidence or possession of certain goods, characteristics, experience or agreement – it is particularly convenient to group them together. For example, we can ask whether a respondent visited each of the four countries listed by displaying the question text only once and listing the four countries as sub-questions. There are several possible layout options for doing this: a series of YES/NO radio buttons, a series of checkboxes, or a multiple selection box.

The first possibility is the *series of YES/NO radio buttons* in the form of a simple table of radio buttons discussed above. In this case, each country is put in a separate line, while two (YES and NO) radio buttons are positioned on the right. This seems much simpler than repeating the same question and response options separately for each country. However, with a question about last year's visit for 28 EU countries or for the 50 US states, this is not a good solution because a high number of redundant clicks on the option NO is usually needed.

The *checkbox* is another layout alternative for a series of radio buttons in the case of dichotomy. Similarly to radio buttons and drop-down menus, the checkbox is also a standard HTML element. As already mentioned, a stand-alone (single) checkbox is rarely used in web questionnaires, because the unchecked option with the implicit meaning NO is unclear compared with an explicit selection of NO in the corresponding radio button layout. That is, the unchecked option can be the result of an omission or refusal. Rather than for survey questions, checkboxes can be thus used in web surveys to get informed consent from respondents (e.g. 'Check if you agree to the conditions of the study and are willing to participate') or allow them to sign up for results (e.g. 'Check if you would like to receive the results of the study').

On the other hand, putting a *series of checkboxes* together is very common in web questionnaires. This layout is often labelled simply as a *checkbox question*, a *check-all-that-apply* or a *multiple answer question*. In the above example, a list of countries with a checkbox adjacent to each country would appear. The respondent selects (checks) only the countries visited, but is spared responding on the unvisited ones. Such questions do not belong to the group of single item questions, as might appear at first sight. Here, one concept (visit) across several subjects (countries) is measured, resulting in as many items as there are subjects, which also includes the corresponding number of columns in the datafile and variables in the analysis.

We may add that a situation with a checkbox question where the number of options is limited (e.g. only two selections are allowed) brings much more complexity (because, to select only two options, all options need to be considered simultaneously) and already belongs to ranking, which we discuss later.

Despite the checkbox being an HTML element – so respondents generally know that it gives the possibility of selecting more than one option – an explicit instruction that several answers may be selected is strongly recommended, as this may not be clear to some respondents.

A potential advantage of the checkbox is that it is quicker to complete in some situations (Callegaro, Murakami, Tepman, & Henderson, 2015) and also takes up less space than the

alternative of YES/NO radio buttons. In the latter case, the respondent is explicitly faced with all options and needs to select an answer for each of them, so this is sometimes also called a *forced choice question.*

However, checkboxes also have serious disadvantages (Bradburn et al., 2004, p. 171). Firstly, problems can appear with mixed-mode surveys, since the standard way of asking multiple answer questions in auditory modes (like telephone surveys) is in the form of YES/NO items. Conversion to checkboxes for the web mode actually changes the context of the question and may introduce methodological differences among modes.

Secondly, an unchecked item in a series of checkboxes can have multiple interpretations. We usually assume that an unchecked item denotes the answer NO. However, the respondent might have missed that option or refused to answer, which in fact denotes item nonresponse. Likewise, respondents might not be sure, so the unchecked item actually means 'don't know'. This limitation can be avoided by adding the option 'none of the above' or 'other, please specify' to the end of the checkbox items, which may help in identifying some item nonresponse.

A meta-analysis of several randomized experiments that compared the layout of a series of checkboxes and YES/NO radio buttons (Callegaro et al., 2015) found that the YES/NO format uniformly provides higher endorsement rates. Nevertheless, the rank and the relative ordering of the items still remain the same in both formats.

To summarize, we prefer a series of YES/NO radio buttons, since it gives us a clearer interpretation of item omissions, respondents might consider each item more carefully, and it is more comparable across survey modes. Explicit reasons should thus exist for using the checkbox question. One such situation is when we have a few very clear and understandable categories (e.g. race origin question). Another is when our main focus is on ranks and relationships between items, and not so much on their precise absolute shares (i.e. exact endorsement rates or market shares). In such situations, the pressure

to minimize the respondents' survey time can make the checkbox question more favourable. The same is true when we have a large number of categories and only a few of them relate to each respondent.

The *multiple selection box* is another alternative for a series of dichotomous YES/NO questions. Sometimes it is also called the multiple selection drop-down menu, multi-select or even combo-box question. Visually, it resembles an opened drop-down menu; however, the respondent can select several responses from the list, by holding down the CTRL key, which is not possible with the standard drop-down menu. Visually, a multiple selection box differs from the drop-down menu because more options are immediately visible without any action from the respondent. In the case of several options, a clearly visible scrollbar is available, and is used to see other options. In the above example of countries visited, instead of a series of YES/NO radio buttons or a series of checkboxes, one single multiple selection box could be used to select all countries visited. An explicit instruction for pressing the CTRL key is strongly recommended here, since this option might not be familiar to all respondents. Due to the complicated interface and reduced familiarity, explicit justification is needed to use this layout. Similar to drop-down menus, it might be practical for a large number of categories and extreme space limitations. A very important improvement here arises when all the potential categories appear in one box on the left, and are then moved into another box on the right with the drag and drop function.

**Tables of checkboxes and multiple selection boxes**   In all the above examples of layout implementation of a series of dichotomous questions, we discussed the measurement of a single concept across several dimensions. In the case of country visits, we have four countries, resulting in four measured items (four columns in the datafile and four variables in the analysis). In a more complex situation,

we may have the same question, but repeated for more subjects (e.g. five family members). Using a series of checkboxes (a *table of checkboxes*) we obtain a two-dimensional table with $5 \times 4 = 20$ items.

Using a multiple selection box for each of the subjects – which means a *table of multiple selection boxes* – further increases the main advantage of a multiple selection box: it takes up even less space, since a single column with five multiple selection boxes will do, one for each family member. We could even add another column of multiple selection boxes: for example, which of the listed languages (English, Italian, Slovenian) each family member speaks, producing a total of $5 \times 4 + 5 \times 3 = 35$ items. Needless to say, the disadvantage of multiple selection boxes in terms of unfamiliarity among the respondents becomes even more pronounced, particularly because a layout with drag and drop into another box is not possible in this case.

## 2.3.2.2.2 Tables with Questions at Ordinal Measurement

Similar to single item questions, our discussion of tables with questions at ordinal measurement predominantly focuses on rating scales. Here, this is even more justified, since a lot of research has been conducted with tables of rating scales, also labelled as matrices, batteries or grids.

Let us stress that many aspects (e.g. number of rows and columns) as discussed here in relation to tables with rating scales relate to the layout of tables in general.

**Tables of rating scales with radio buttons** The radio button layout prevails in tables with ordinal measurement, where each item – that is, each dimension (sub-question) of the measured concept(s) or each subject – is presented in a separate line. For example, we can ask the respondent to rate an overall travel experience for four countries on a 5-point rating scale. Typically, the question wording is placed on the left and the repeating response options (i.e. the radio buttons) on the right, while a reverse orientation – lines belonging to response categories and dimensions in columns – is generally inferior (Galesic, Tourangeau, Couper, & Conrad, 2007). Labels for response options are displayed only once, in the header at the top of the response columns. As already shown in the discussion on tables with nominal measurements, we have here only a series of single rating scales.

As with items on rating scales, we should be careful about column spacing among the response options in tables, because uneven spaces may affect response distributions (Tourangeau, Conrad, & Couper, 2013, p. 79).

Care is needed when the number of columns and rows is high, which is a general issue for all types of tables. We should avoid horizontal scrolling (left–right), even if this requires a reduction in the number of scale points (e.g. from 7 to 5). Vertical scrolling (up–down) is less critical, but is still best if it is avoided, due to disappearing top column labels when scrolling down a long table. To display the entire table on a typical computer screen, we can limit the number of dimensions in rows (items) per table to 8–10. Another solution is a programming feature where the header of the table is fixed and only items are scrolled, or the header labels are repeated after a certain number of items. However, as long tables also increase the respondent's fatigue, they are generally detrimental.

Marking the lines in the table with alternate shading is recommended for better separation among them. For example, Crawford, McCabe, & Pope (2005) suggested using light-grey background shades in alternating rows to improve readability. Dynamic shading – the font colour of the row changes or shading is provided after a response has been selected – was also successful in reducing item nonresponse (Galesic et al., 2007), though pre-selection shading and mouse-over highlighting can be detrimental (Kaczmirek, 2011). In general, however, when formatting

the tables we should avoid any redundancy and visual clutter, so that the design of the table can fully serve its basic purpose: enabling respondents to provide quality answers in a user-friendly manner.

As summarized in Lozar Manfreda & Vehovar (2008), tables of rating scales with radio buttons save space, make the questionnaire look shorter, and are convenient to create. They often require less effort from the respondents, as there are fewer instructions and response labels. Mouse movement is also reduced, as well as the response time. On the other hand, such tables can have serious disadvantages, as shown by various studies:

- When the complexity of tables increases the respondents' cognitive burden, they can react with various satisficing strategies (e.g. smaller differentiation of answers), resulting in lower data quality (Fricker, Galesic, Tourangeau, & Yan, 2005; C. Zhang, 2013).
- There is evidence (e.g. Lozar Manfreda, Batagelj, & Vehovar, 2002; Toepoel et al., 2009) and concern (e.g. Couper, Tourangeau, Conrad, & Zhang, 2013, p. 113) that tables increase item nonresponse in comparison with a sequence of single item questions.
- Tables may change the nature of questions, because the items are placed in a comparative framework, which may result in context effects. These effects are weak but still present in web surveys (e.g. Tourangeau et al., 2004).
- Tables are often reported as being the most critical point at which respondents abandon the survey (e.g. Henning, 2011; McMahon & Stamp, 2009).

We lack systematic research where potentially negative effects (e.g. lower data quality) of tables of rating scales with radio buttons would be compared with actual disadvantages (e.g. increased time and length) of a series of single radio button questions as the alternative format. One reason for this deficit is the perplexing factors that can appear. For example, the effect of a table layout on breakoffs can differ at the beginning and the end of the questionnaire. Similarly, the effects of table layout on satisficing strategies can differ for 'professional' respondents in online panels and

for novices. We also advise caution regarding research on response times, which is often influenced by the effects of page design, where each item is placed on a separate web page as an individual question. Minor differences in response times were found when tables were compared with a sequence of single item questions on the same page (Bell, Mangione, & Khan, 2001), despite the fact that the latter option resulted in a questionnaire that was doubled in length. On the other hand, having each question on a separate page can cause an increase of over 50% in the response time (Callegaro, Yang, Bhola, Dillman, & Chin, 2009), which is mostly due to the effect of page breaks and not directly related to the specifics of the table layout.

**Unfolded tables – horizontal scrolling matrix** The *horizontal scrolling matrix* (HSM) is an alternative layout for presenting sub-questions in tables with rating scales using radio buttons. The sub-questions (i.e. items) appear on the screen one by one, with responses (radio buttons) in either a vertical or a horizontal layout. The respondent sees only one sub-question at a time. The next question is automatically presented after the previous one is answered; there is no need to click on the 'Next' button. Navigation with an item counter and a visual bar provides control over the number of questions answered and progress. This format shows promising results with respect to context effects and reduced complexity (Klausch, de Leeuw, Hox, de Jongh, & Roberts, 2012). The approach is particularly suitable for devices with small screens (e.g. smartphones) and certain web survey software automatically transforms tables into such separate questions whenever a small device is detected (see Section 5.1).

**Other tables with questions at ordinal measurement** Drop-down menus, open-ended numeric entries and advanced graphical layout are alternative implementation layouts not only for simple tables with the rating scales, but also for other questions with ordinal measurement. The discussion of their limitations was covered in single

item questions and with tables of rating scales, and they are also fully relevant here. Similarly, all the general problems with table layout apply too.

### 2.3.2.2.3 Tables with Questions at Interval and Ratio Measurement

The specific features of radio buttons, drop-down menus, numeric entries and continuous scales, serving as potential layouts for single item questions in the interval and ratio measurement levels, apply to corresponding tables as well. With respect to the layout itself, the general principles already discussed in relation to tables also hold true. We may only add that the cumbersome appearance of tables of continuous scales – compared with a more familiar radio button layout – can potentially reduce participation in general population surveys.

The *open-ended numeric entry format*, as the default table layout for the interval and ratio measurement levels, is closely related to those issues concerning open-ended text entries, as discussed with tables at the nominal measurement level. Namely, a series of open-ended numeric entries (e.g. height, width and length) strongly resembles a series of open-ended text entries (e.g. name, family name, address).

A date entry, which is basically a simple series of three numeric questions, often serves as a typical case for discussing how variations in the layout implementation of questions affect the responses. To summarize, a simple series of three numeric entries for the year, month and day with symbols 'DD', 'MM' and 'YYYY', added close to the entry field, can be a good solution (Christian, Dillman, & Smyth, 2007; Couper et al., 2011b), but drop-down menus can be used as an alternative too (Couper et al., 2011b).

### 2.3.2.2.4 Combined Tables

In all the above cases, questions of the same measurement level and of the same implementation layout are combined in the form of tables, resulting in what we call simple tables. However, the complexity of tables can increase with *combined tables* (sometimes also called 3D tables), which combine more measurement levels and layout formats. An example is household rosters, where for each person we may have an open-ended text entry (name), an open-ended numeric entry (age) and a drop-down menu (gender) in one line. In addition, such lines are combined together in a table for members of the whole household. In general, we should be cautious about adding such complexity and should consider splitting such a table into sets of simpler tasks. However, situations exist where the savings in space and time may outweigh the potential disadvantages. This is particularly true for factual questions, which are extensively used in business surveys (e.g. financial sheets with entries for items by years). Another situation is where combined tables are required to reflect fully the parallel P&P forms, although this may not be the optimal solution for web questionnaires. Certain general usability principles were developed for this setting by Morrison et al. (2010), and should be carefully implemented in web questionnaires.

In conclusion, should tables in web questionnaires be used or not? As usual, there is no uniform solution, but we can still summarize certain conclusions. Although these are predominantly based on the most commonly used simple tables with rating scales, we believe they can be largely generalized to other table layouts as well. In the majority of situations, a well-designed table can save space. But an improper implementation can seriously affect data quality. Even when properly designed, there are indications that tables can cause problems (e.g. context effect, item nonresponse, satisficing). On the other hand, there are no clear indications about the advantages of using tables, with the exception of the unverified belief that – due to the obvious reduction in space – tables directly save response time and allegedly increase response rates and data quality. Consequently, a general recommendation would be to avoid – or at least minimize – the use of tables whenever possible. As tables importantly make the questionnaire more difficult, they may cause more damage to data quality than slightly lengthier alternatives with simpler questions (Burdein, 2014). In any case, despite considerable past research, we still lack a comprehensive study that compares tables

with the alternatives (i.e. sequences of single questions) and simultaneously addresses all key aspects of data quality.

## 2.3.2.3 Other Question Types

In addition to the above question types, which relate to single item questions or their combinations in (simple) tables, various other question types exist. They are relevant when the respondent needs to consider more items (dimensions, subjects) simultaneously or when extensive graphics are involved, as well as when questions are combined and implemented in very specific contexts. Due to their complexity, we do not structure them systematically, but we illustrate below some common examples.

### 2.3.2.3.1 Ranking

*Ranking* mimics the card sorting technique used in F2F interviews, where respondents sort a set of cards (subjects) according to a single criterion (conceptual dimension), for example brands according to their preference. A similar approach is *grouping*, where respondents group certain subjects, for example grocery products, clothing products, etc. Respondents can be asked to select or rank only certain subjects, for example the top three subjects. The ranking task can also be split into several steps, as in *paired comparisons*, where not all, but pairs of subjects are compared/sorted at a time, or in *maximum difference scoring* (MaxDiff), where only the extreme subjects are identified, such as the most important and the least important.

Usually we have only one dimension; ranking according two dimensions is also possible, but very rare (Bradburn et al., 2004, p. 176).

Ranking questions can be implemented using the list of short numeric entries, into which a respondent enters a rank for each subject. Another possibility is to drag and drop subjects from one list to another in a preferred order, or to reorder the subjects within the same list. Blasius (2012) found the drag and drop layout superior to all alternatives (numeric entry, movement with arrows, most–least selection) in terms of the response time and item nonresponse. However,

the drag and drop technique may be problematic for some users, due to technical limitations, particularly if mobile devices are used to access the web questionnaire.

The ranking task is cognitively very demanding, though it appears as a simple ordinal measurement question. However, an important specific of ranking is that two subjects cannot be assigned the same rank value from the available ordinal categories. As a consequence, the respondent needs to consider more subjects simultaneously, which is already demanding with five subjects, and becomes very difficult with more than ten.

Serious controversies about ranking vs rating already exist in traditional survey modes, but have been reinforced with the potentials of drag and drop in web questionnaires (Neubarth, 2010). However, the dilemma remains relatively under-researched. The rating ensures independent responses for each subject, higher validity, richer statistical analyses and lower cognitive burden compared with ranking. In addition, once we have the ratings, we can always sort the subjects to obtain the ranks. Nevertheless, ranking may be preferred when we have very small differences in ratings or when we cannot accept more subjects with the same rank. For example, a respondent may endorse four products with the highest rating (e.g. score 5 on a 1–5 rating scale), yet we would like to know the respondent's decision about purchasing only one or two. For this purpose, we need them to be sorted with the ranking approach.

### 2.3.2.3.2 Constant Sum

A *constant sum* question – sometimes also called a *running tally* – denotes a series of open-ended numeric entry questions (e.g. hours spent daily on certain activities), where responses are additionally restricted to a certain fixed sum (e.g. daily number of hours to 24). This is another example of a cognitively demanding question, where respondents need to consider and process more dimensions or subjects (items) at the same time. Implementing a real-time automatic sum, accompanied by certain prompts, in this case clearly increases the data quality (Callegaro,

DiSogra, & Wells, 2011; Conrad, Tourangeau, Couper, & Zhang, 2010).

## 2.3.2.3.3 Extensive Use of Graphics

Graphics are a very powerful feature of web surveys. Although we mostly focus on the basic web survey mode, where the use of multimedia (images, audio, video, animations) is limited, we nevertheless present below some of the possibilities of using extended graphics for survey questions, and discuss them briefly in relation to data quality. We return to this issue later when we address the entire visual layout of the web questionnaire and gamification (Section 2.3.4).

**Graphics for illustrating survey questions.** The use of graphics to illustrate survey questions in web surveys is attractive, easy and inexpensive, compared with other modes. However, a general concern exists in academic (e.g. Couper, 2008), as well as in marketing research (e.g. Poynter, 2010, p. 55) about the weak level of control researchers have on corresponding effects. Firstly, various technical difficulties may slow down download time, produce a different appearance of colours and fonts, and prevent appropriate functioning on certain devices and browsers. For example, Lozar Manfreda et al. (2002) report that when brand logos were added to questions in web questionnaires, the breakoff rate increased. Secondly, while technical problems are diminishing with increased Internet connection speeds and technical standardizations, this is not the case with unpredictable effects on responses.

When graphics are the essence of the question – as in a question on logotype preference – little can go wrong. However, when they are used merely as an additional questionnaire element (e.g. as an attempt to make the questionnaire more attractive), unpredictable side effects may occur. This is particularly problematic when graphics are inconsistent with the verbal context, which generally takes precedence over the visuals (Toepoel & Couper, 2011). A typical example is a picture of a healthy or a sick person next to a question on respondents' health, which then generates differences (3.4 vs 2.5 on a 7-point scale) in responses, with additional

effects depending on the size and position of the picture (e.g. previous page, header, question side) (Tourangeau, Conrad, & Couper, 2013, p. 88). There is little evidence that adding non-essential pictures will improve data quality or increase satisfaction with the questionnaire (Toepoel & Couper, 2011), which is otherwise a frequent idea behind the use of graphics. No advantages of pictures were found also by Deutskens, de Ruyter, Wetzels, & Oosterveld (2004) and Ganassali (2008).

In general we should use additional graphics very carefully and thoroughly think about the possible side effects. In case of doubt, a conservative approach is recommended, particularly in prevailing situations when there is no research evidence to support the advantages of additional graphical elements. Such an attitude may seem unusual, conservative and even out of date, because graphics and multimedia are essential advantages of the web survey mode. Nevertheless, without evidence of advantages and corresponding guidelines, it is better to be on the safe side.

**Graphics for simulating social presence.** Pictures or video of interviewers may be used in web questionnaires to simulate social presence with the idea of increasing motivation and data quality, as successfully shown by Półtorak & Kowalski (2013). Experiments on this issue have replicated the effects from interviewer-administered surveys, particularly gender. (Fuchs, 2009; Tourangeau, Couper, & Conrad, 2003) and the ethnicity of the interviewer (Krysan & Couper, 2006); thus, their use should be treated with caution.

**Graphics forming new question types.** Apart from the use of graphics as an additional element illustrating survey questions, graphics also provide opportunities for new question types and alternative layouts to existing ones. We have already presented some examples in our discussion (drag and drop, images replacing radio buttons, hotspots and continuous scales) and below we provide some more illustrations of questions in web questionnaires which take functional advantage of graphics:

- *Calendar layout* as a graphical interface for date format is an alternative layout asking for

the date. Pop-up calendars are used in this case, and selecting a date from the calendar fully replaces the need for numeric entries or drop-down menus. Their use in various other contexts (e.g. online booking) makes this format increasingly advantageous for many situations in web surveys. More complex versions of online calendars are used in life history measurement, with graphical representations of dates, including certain landmarks, in order to simplify complex recall tasks (Glasner & van der Vaart, 2013).

- Selection of subjects in a *virtual reality environment* is an extension of ranking; for example, in a virtual supermarket products are picked up, inspected and placed in a certain order into a shopping basket using the drag and drop function (Brace, 2008, p. 169).

- *Heat-map* questions rely on a graphical interface, where respondents click on a point in the picture, according to a certain dimension (e.g. most important, attractive or unattractive part). There are no predefined areas here as with graphics discussed in the nominal measurement level (e.g. selecting regions). As each point matches the two-dimensional coordinates at the ratio measurement level, this enables very fine analyses. The results of corresponding heat-map analyses are density focuses presented by coloured graphs, similar to reporting on eye tracking or mouse movement. The majority of applications of this layout relate to usability and evaluation studies.

More examples of graphics forming new question types can be obtained on various marketing research websites (e.g. GMI).

## 2.3.2.3.4 Content-specific Implementations

We initially restrict our discussion on question types in relation to methodological aspects and exclude various content-specific questions (e.g. job satisfaction), which are sometimes also protected by copyright (e.g. Q12 questions for measuring employee engagement by Gallup Consulting; Harter, Schmidt, Killham, & Asplund, 2008). We also exclude specific combinations of questions closely related to certain research methods. These aspects essentially bring no new question types, but they do create

very specific implementations. We thus highlight a few examples which typically illustrate their extensions in the web survey context:

- **Ego-centric social network questions** collect lists of people (alters) built with name generator(s). Respondents (egos) then answer the same questions (e.g. on age, gender, relation) for each alter from their network (e.g. friends). The *name generators* are formally a series of open-ended text boxes, where respondents identify the names of network members. The same layout can be used to generate lists of products, services, destinations and other subjects, especially in marketing research. Research on the most appropriate and standardized implementation layout of such questions in web questionnaires is still in progress (e.g. Lozar Manfreda, Vehovar, & Hlebec, 2004; Vehovar, Lozar Manfreda, Koren, & Hlebec, 2008). Major challenges are related to the graphical interface and format of the name-collecting text boxes, where the sequential appearance of text boxes, with the next box displayed only after the previous one is filled in, seems to be preferred (Hogan, Carrasco, & Wellman, 2007). For collecting data on alters, a series of questions for each alter can be used, either item-wise or alter-wise (Coromina & Coenders, 2006; Vehovar, Lozar Manfreda et al., 2008). The potential of the graphical interface to replace survey questions for social network measurement presents a further challenge (Hogan et al., 2007; Koren & Hlebec, 2006; Lackaff, 2012).

- **Conjoint analysis questions** integrate ranking questions into a series of questions with potential decisions (e.g. the purchase of a product). A carefully designed setting is used here, with different levels (ranks) of characteristics. This layout then enables us to apply a conjoint analysis approach. For example, respondents need to decide – among other variations – between a fast, cheap and small car vs a slow, expensive and large one. This approach can also use rating scales.

- **360°** or **multi-rater feedback** is a variation of the social network questionnaire used in human resource management (HRM), where each employee rates (evaluates) a list of persons from the higher (managers), lower

(subordinates) and peer (coworkers) levels. The web is particularly convenient for this task, but raises numerous methodological challenges (LeDuff Collins, 2009).

### 2.3.2.3.5 Questions Including Observations and External Data

Data from respondents in web surveys can also be collected by other methods (Groves et al., 2009, p-150), using recordings, plug-ins or file uploads. In this way audio recordings, drawings, signature scans (receipt payment), barcode or QR code, photographs (e.g. selfies), videos, bio-markers (e.g. measuring weight), media channel recognition (e.g. radio, TV) or GPS location can be collected. These possibilities represent an extremely important advantage of web surveys, particularly with the use of mobile devices. However, we will not go into more detail because we initially restricted the discussion to basic web survey mode.

With this we conclude our overview of question types in web questionnaires (Section 2.3.2). We followed the classification according to the methodological conceptualization, starting with complexity and then with measurement level and layout variations.

However, when we are creating web questions, we typically encounter the structure of question types provided by the web survey software used, which is determined to a large degree by the technical nature of selecting sub-settings and sub-options. At the very first level, web survey software usually separates single item questions and tables, open and closed-ended questions, questions with more answers, etc., which are closely related to HTML elements. This first level navigation for the question-selecting process in web survey software also depends on the frequency of use of various question types. Thus, the most frequent types (i.e. radio buttons, checkboxes, open-ended text and numeric entries, simple tables of rating scales with a radio button layout) are usually more easily available. See details in *Frequency of appearance of question*

*types in web surveys*, Supplement to Chapter 2 (http://websm.org/ch2).

In sum, selection of the most appropriate question type is often a very complex decision that needs to be guided by considerations of question content, clarity of presentation, technical requirements, and the task difficulty it imposes on the respondent. It is especially important that a specific type is not used simply because it is available or because it looks innovative and interesting – the primary criterion of use should be the data quality it provides for our research purpose.

## 2.3.3 Questionnaire Structure, Computerization and Layout

A survey questionnaire is much more than just a sequence of questions. Especially in the self-administered modes, such as web surveys, it can be regarded as a medium of conversation between respondents and researchers (Schwarz, 1996). The questionnaire's role is to ensure the flow of this conversation. This is achieved through the questionnaire's structure, interaction with the respondent and visual layout.

### 2.3.3.1 Structure of the Questionnaire

Decisions related to the structure of the questionnaire involve a broad spectrum of factors that we briefly address here: order of questions, distribution of questions across pages, inclusion of non-question pages, navigation, division of the questionnaire into sections and blocks of questions, and the use of special layouts for specific purposes.

#### 2.3.3.1.1 Question Order and Context Effects

It is well known that the order of questions within the questionnaire is important and may guide their interpretation and the provision of answers. When one question affects the processing and answering of other questions, we talk about *context effects*. These effects have few specifics for web surveys and many studies report on

their occurrence (Couper et al., 2004; Ester & Vinken, 2010; Malhotra, 2008; Nielsen & Kjær, 2011; Siminski, 2008), including questionnaires on mobile devices (Mavletova, 2013; Peytchev & Hill, 2010). The general principles of survey methodology can be thus directly applied to web questionnaires (Krosnick & Presser, 2010, p. 264): questions should be grouped by topic, starting with those mentioned in the invitation, and then proceed from the most to the least salient; within each topic there should be a flow from general to more specific questions; questions should also be grouped by format and logic (e.g. chronology of events); starting questions should be simple and attractive, while demographic and sensitive questions should be left to the end.

Context effects can be prevented by avoiding tables and by increasing the number of page breaks to separate the questions visually. Furthermore, question order effects can be handled with randomization, which we describe further in this section. More information on context effects can be found in general textbooks, such as Krosnick & Presser (2010, p. 291), Dillman et al. (2009, pp. 157–165) and Tourangeau et al. (2000, p. 197).

### 2.3.3.1.2 Page Breaks and Paging vs Scrolling

Questions can be distributed across questionnaire pages in different ways. Two extreme approaches regarding the number of questions per page are a *one-page design* (also named *scrolling*), where all questions are presented on a single page, and a *one-question-per-page design* (also named *paging*). The comparison of these two approaches was the focus of some of the first web survey experiments (Lozar Manfreda, Batagelj & Vehovar, 2002; Vehovar & Batagelj, 1996). These experiments found no differences in break-off rates for a 7-minute survey on general topics, but the scrolling design had more questions that were left unanswered. On the other hand, it was faster and showed lower response times. The latter difference was in large part due to the slowness of the Internet connections at the time. Nevertheless, even today it seems that these essential findings still hold true, although weakly.

The paging design has the advantage of resembling interviewer-administered questionnaires, where respondents are presented with only one question at a time. It also has an advantage of easier and more robust server-side implementation of interactive features, which are only executed after the respondent moves to the next page. However, the need to move to the next page after each question increases the burden for the respondent and typically expands response times slightly (Couper et al., 2001; Lozar Manfreda, Batagelj & Vehovar, 2002; Thorndike, Calbring, Smyth, Magee, Gonder-Frederick, Ost & Ritterbrand, 2009; Toepoel et al., 2009; Tourangeau et al., 2004).

Weak support for interactivity is one of the key deficiencies of scrolling. For example, branching, which means that the respondent skips over some questions (e.g. if gender is male, than the questions on childbirth are skipped), can be implemented by including hyperlinks, accompanied by written instructions (e.g. Peytchev, Couper, McCabe, & Crawford, 2006), which is rather awkward and burdensome for respondents; it also increases the response times. Alternatively, client-side scripts to perform *dynamic branching* – also called *hybrid* (Dillman et al., 2009, p. 202) – can be used, so that a click on a certain response option instantly (on the same page) invokes an additional set of questions. However, Peytchev et al. (2006) found that skips in scrolling design can cause avoidance of those response options that lead to a display of a large number of additional questions. Another disadvantage of scrolling is that the majority of web survey software saves answers only when proceeding to the next questionnaire page. Scrolling thus requires completion of the questionnaire in one session and does not save responses in case of break-offs. Furthermore, since all questions are visible to the respondent at once, the likelihood of context effects may increase.

On the other hand, the absence of page breaks in scrolling reduces the number of required clicks and also the response times. It also provides the respondent with insight into the entire

questionnaire and thus more closely resembles a P&P self-administered format.

With respect to differences in substantive results, studies generally found no additional effects of paging vs scrolling design (Thorndike et al., 2009; Toepoel et al., 2009). It is true that in the case of simple web questionnaires, certain disadvantages of scrolling design disappear, but this still does not mean that scrolling has any advantages. According to Toepoel et al. (2009), differences in response times among four 10-item-per-screen pages and a scrolling design with all 40 questions per screen were negligible, while the increase in satisfaction with scrolling was very small, from 7.06 to 7.37 on a 10-point scale. The only real advantage of scrolling might appear when the context is required (e.g. all questions should be available on one screen for respondents to see all of them) or when it is essential that it resembles a P&P version.

In practice, we often use approaches somewhere in between these two extremes. We define a *modified scrolling* design as when page breaks appear only when this is necessary for the execution of server-side features such as branching, for reducing the context effect and for intermediate saving of responses. In an alternative approach, which we call a *modified paging*, each page contains a limited number of questions that fit on a typical screen without further scrolling. Given that studies found only small differences between the full paging and the full scrolling design, we can expect them to be even smaller for modified paging and modified scrolling.

We can conclude that we almost cannot go wrong with the modified paging strategy, while modified scrolling might still suffer from some of the scrolling problems mentioned above, especially from less frequent saving of responses into the database and problems with skipping. In addition to this general conclusion, we need to consider the advantages and disadvantages of each approach within the context of a specific survey.

Finally, we should point out the changing role of scrolling in the last decade, due to blogs, social networks and mouse devices. These all caused major changes in web usability principles (Nielsen, 2000a), which were very much against scrolling in the early years of Internet development. Scrolling has been additionally reinforced lately with mobile devices (smartphones, tablets) (Mavletova & Couper, 2014). All these might soften the disadvantage of scrolling designs.

### 2.3.3.1.3 Non-question Pages and Sections

In addition to questionnaire pages with survey questions, we sometimes include pages or questionnaire sections without questions. They contain introductions, additional instructions and other information relevant for respondents before they proceed with the questionnaire. Certain types of non-question pages listed below are commonly used, while the content of others is more usually presented next to other questions on the same page, especially if it is relevant only for specific questions and is relatively short:

- *An introduction page (splash page, welcome screen)* is a separate first page which introduces the survey to the respondent. It is omitted only in very short surveys (e.g. evaluation forms). The introduction page has the role of convincing the respondent to participate in the survey. Many respondents access the introduction page, but are not persuaded to continue with the survey, which makes this page a common place for major breakoffs. Its content and design are therefore very important: it must be appealing and respectful, professional and polite. It should not jeopardize research ethics, so essential information therefore needs to be conveyed fairly to the participants: namely, the survey sponsor, purpose, content, privacy issues, contact information and expected time needed to complete the questionnaire. It is also important to stress additional encouragement for the respondents' participation, like the importance of the survey, the benefits for respondents, incentives, and so on. We further discuss these aspects of ensuring survey participation in Section 2.5 on nonresponse. At the end of the introduction page, it is also necessary to provide instructions for starting participation in the survey. This may include a simple instruction, 'To proceed click on the "Next" button', or additional

guidance for entering a survey access code. In general, the page should be kept short and simple. Certain research has shown (e.g. Bauman, Jobity, Airey, & Hakan, 2000) that replacing a lengthy introduction page with dense text in 'cover letter' style with a shorter and more concise presentation increased cooperation. When mail or email invitations are used, the content of the introduction page needs to be in line with the invitation in order to avoid unnecessary repetition (Section 2.5.7).

- *A transition page* introduces respondents to a new topic. Some research shows the positive effects of transition pages without raising break-off rates (Callegaro et al., 2009). They can be used to slow down the pace of the questionnaire and ease the transition from one questionnaire topic to another.

- *Instruction pages* are used to provide respondents with the necessary information for completing the more complex survey tasks (e.g. how to perform a ranking task). When the required instructions are not long and complex, it is not necessary to present them on a separate page, and they may be included directly next to the question to which they apply.

- *An incentive* or *raffle page* can be used to give additional information about incentives for completing the questionnaire, for example by showing a picture of the incentive or stating the odds of winning the lottery incentive. Respondents may also collect incentives, such as online coupons and other electronic incentives.

- *A file upload page* or *section* is sometimes used to obtain pictures, text documents or other files from respondents. For example, a respondent can be asked to upload a photo of a defective product to which the questions refer. This can appear as a special page or, more often, as a specific request for an upload listed among other questions, although we can hardly talk about such a request being a survey question.

- *A thank you page* is the last page of the questionnaire and is used mainly to acknowledge the respondent's participation in the survey. It may also contain other elements, such as links to external websites (e.g. the website of the survey sponsor or the website with content related to the topic of the questionnaire) and information about the availability

of results. Sometimes respondents may be asked for contact information in order to receive the results by email, participate in further surveys or join an online panel. The completed questionnaire, as filled out by the respondent, may be enclosed (e.g. in PDF format) so that the respondent has an archive of the responses.

- In addition to non-question pages and sections, we can also have non-question sentences which are part of the questionnaire pages (e.g. introductions, explanations, encouragement, thank you notes).

### 2.3.3.1.4 Questionnaire Navigation

A respondent usually moves back and forth in the questionnaire using dedicated navigation buttons in the questionnaire. Here, the *'Next' button* is obligatory, while the *'Previous' button* is optional. The latter gives respondents some control over the questionnaire and allows them to correct answers (as in a P&P questionnaire). Sometimes this is regarded as undesirable behaviour and the option is removed, but this can then lead to an increase in breakoffs (R. P. Baker & Couper, 2007). We provide more details on the formatting of these buttons in the discussion on visual layouts further in this section.

Sometimes, *automatic forwarding* is used to avoid the need for the 'Next' button. Here, the selection of any responses automatically leads to the next page of the questionnaire. One problem with this approach is that it can create navigational confusion, since it can be implemented only with closed-ended single answer questions. Hammen (2010) also showed an increased tendency to satisfice in case of automatic forwarding, so very specific circumstances must exist to justify it, for example with the horizontal scrolling matrix presentation of tables of items (Section 2.3.2).

### 2.3.3.1.5 Blocks of Questions

The questionnaire is often structured into parts or sections, technically formatted as blocks. In longer questionnaires these can be visually separated by subtitles, running heads, introductions and encouragement, which provide respondents with some

orientation and motivation. Top-level blocks can be presented even as navigation, visible throughout the entire questionnaire in the form of tabs, which then enable respondents to switch between different topics. This approach is more commonly used in business surveys, where different people may complete different parts of the questionnaire. However, navigation through the questionnaire should not be made too complex because it may confuse respondents (Blanke, 2011).

### 2.3.3.1.6 Special Layouts

Web questionnaires usually run in a separate browser window or tab, which can be maximized and resized by the respondent. However, for very specific purposes, alternative questionnaire layouts can be used. One such specific implementation is an *embedded questionnaire*, which is integrated into an existing website and appears as part of a certain web page, such as a consumer satisfaction survey on a certain page of an online store website. The main disadvantage of this approach is the limited space that can be allocated for the questionnaire. It is also less appropriate for more complex questionnaires, due to possible technical problems with some interactive and dynamic features that rely on client-side technologies (e.g. JavaScript).

Another example of a special questionnaire layout is a *split-screen* presentation, commonly used for website evaluations. In this case, a web questionnaire is usually presented in the lower browser window, while the evaluated website can be simultaneously used and inspected in the window above it.

### 2.3.3.2 Computerization, Interactivity and Dynamics

At a certain point in the questionnaire development process, we need to transform the fixed and static content of the draft questionnaire versions – which are often prepared on text processors (e.g. MS Word, Google Docs) or even on paper – into an online version of the questionnaire in the chosen web survey software. Alternatively, we can start developing the questionnaire directly in the dynamic format of the web survey software, on the condition that the software supports the required features (Vehovar, Čehovin, & Močnik, 2014). In both cases, we then face the same essential challenges of dynamics and interactivity: skipping the questions (branching), randomizing questionnaire elements and providing various interactive feedback to respondents.

### 2.3.3.2.1 Branching

We talk about *branching* – also called *routing, filtering, skipping, conditions, IF sentences* – when a certain question is asked conditionally, based on previous responses (e.g. number of births is asked only if the reported gender is female). Compared with written instructions in P&P questionnaires, web surveys are superior and can automatically handle very complex conditions. A specific type is *loop branching*, where we repeat the same set of questions for each of the previously reported items. For example, for each country which was checked as being visited, a set of further questions appears. As in paging and scrolling discussed above, in server-based web surveys the branching process runs on a server and the conditional questions can only appear on separate pages. An alternative that can be implemented without interacting with the server is dynamic branching, allowing a dynamic appearance of questions within the same page. It relies on client-side technologies (JavaScript). As mentioned in the discussion on scrolling design, this enables the next question to appear on the same page (e.g. a question about the number of births appears on the same page as the gender question, immediately after clicking on the female gender option). This might work well for a few questions, but the danger for nonresponse and satisficing exists once the respondent realizes that some responses lead to larger sets of additional questions.

Support for branching is one of the important aspects where web survey software considerably differs in its capacities and even more in the usability of the user interface through which the branching logic is defined.

## 2.3.3.2.2 Randomization

Randomization enables manipulation of the presentation of questions, response options, layouts, blocks of questions, etc., between different respondents, according to some random mechanism. For example, we can use odd and even record numbers of the respondents accessing the web questionnaire to form two random groups.

The first type of randomization affects the order of questions, items or response options presented to the respondent. This can be done by random ordering of (a) response options in closed-ended questions with the nominal measurement, (b) sub-questions in a table, (c) questions on a page or within a block of questions, and (d) the blocks of questions themselves. Of these, the random order of response options is most commonly used, because it can handle response order effects. In web surveys this is especially important due to potential primacy effects, where respondents first consider the few options on the list more thoroughly (Galesic et al., 2008), and are also more likely to choose answers at the beginning of the list. There is also some evidence that respondents think more positively about the first options (Tourangeau, Couper, & Conrad, 2013). Such randomization does not change the problematic behaviour of respondents, but only spreads the effect randomly across all response options or sub-questions. In order to truly prevent primacy effects, other strategies need to be considered, such as increasing the respondent's motivation by appropriate instructions and tooltips (Kunz & Fuchs 2013).

Another type of randomization is an experiment where – different to the above randomization of the order where all respondents receive all options – each respondent is randomly assigned to one of the predefined questionnaire elements (response option, question or block of questions, implementation layout, etc.). A typical example is a so-called *split-sample* experiment with two random groups, which allows us to explore which question wording or implementation layout provides better data quality. We have already discussed experiments in Section 1.3.6, while further recommendations are offered by Reips (2007).

Similar principles of randomization can be used to reduce the respondent's burden. For example, certain questions can be randomly assigned to only half of the respondents and the remaining questions to the other half (see also discussion in Section 2.2.2). This approach, sometimes called *matrix sampling*, can increase the number of questions in the questionnaire without increasing the burden on the respondents. Of course, splitting can be used only for questions that will not be analysed together. Matrix sampling can be extended to *modular surveys* (Johnson, Siluk, & Tarraf, 2014) and *split-questionnaire designs* (Raghunathan & Grizzle, 1995), where additional imputations using data fusion (statistical matching) are then used in an attempt to complete the missing part of the responses.

Randomization can also be used in many other contexts which are otherwise beyond the questionnaire development context, for example the random selection of persons in a household.

## 2.3.3.2.3 Real-Time Validations and Prompt Messages

Web questionnaires can react to the respondent's answers and other actions by performing real-time validations and providing feedback in the form of prompt messages. This can be used for various purposes (Peytchev & Crawford 2005), including (a) prompts about items that are left unanswered and thus present item non-response, (b) controls to assure proper provision of answers according to question types, formats, length and range, and (c) consistency validations, where the consistency of responses is verified against some other data from the same survey (e.g. consistency of reported education and age), previous surveys of the same respondents, or external data from the sampling frame or other sources. Other types of validations and prompts also exist, including feedback based on paradata, such as notifications about responding too quickly. Certain system messages issued by the operating system or the web survey software may be relevant as well, for example

notifications about lost Internet connection or notifications of an expired browser session in the case of long inactivity.

Interaction based on the validation is achieved by error messages and other forms of feedback to the respondent. Technically, the process can run on a server, so the feedback is provided when the respondent attempts to continue to the next page. Alternatively, client-side technologies can be used to enable real-time feedback on the same questionnaire page. Validations can generate three possible reactions: (a) *no prompt* for the respondent, who can normally proceed, but some indicator of a potential invalid answer may be recorded in the datafile; (b) a *soft prompt* where the respondent is notified about the error, but can still proceed without making the requested correction; and (c) a *hard prompt*, where the respondent cannot proceed without correction.

Since validation messages interfere with the respondent's completion of the surveying task, they may be considered intrusive and annoying. It is therefore important to implement them by relying on professional standards, as well as design and usability principles (Couper, 2008; Nielsen, 2000a). It is most important to keep the messages polite and respectful. Their content needs to state clearly what the problem is and how to fix it. They also need to be tested thoroughly.

Below, we take a closer look at the item non-response prompt as one of the most typical, important and frequent validation examples. The main principles remain the same for other types of validation messages.

### 2.3.3.2.4 Item Nonresponse Prompts

When a respondent decides not to answer a certain question (we call this an item nonresponse), a researcher can use a soft prompt, a hard prompt or do nothing. In practice, hard prompts are often used, especially if respondents obtain incentives for participation. However, academic researchers (Couper, 2008, p. 266; Dillman et al., 2009, p. 309) strongly advise against the use of hard prompts if they are not essential for further surveying, for example when a question is a key question on which branching in the remainder of the questionnaire depends. The first reason for avoiding hard prompts is ethical, since survey participation is generally voluntary and respondents should not be forced to answer any question (AAPOR, 2010). The second reason against hard prompts is a methodological concern that they may lead to breakoffs or false responses. Unfortunately, we lack convincing empirical evidence to support this. Although both hard (Albaum, Roster, Wiley, Rossiter, & Smith, 2010) and soft prompts (e.g. DeRouvray & Couper, 2002) decrease item nonresponse, the effect of hard prompts on breakoffs was found to be insignificantly higher compared with soft prompts (Albaum et al., 2010; Couper, Baker, & Mechling, 2011; Heerwegh, 2005). Even more lacking is research on the effects of hard prompts on false responding and other aspects of response quality. Nonetheless, this does not mean that negative effects do not exist.

On the other hand, hard prompts apparently have great benefits: they eliminate item non-response, permit immediate analysis and save lots of resources, compared with situations where we need to deal with missing data in the post-fielding. Hard prompts may also discipline respondents at the outset, so they might take the survey task more seriously. It is thus not surprising that in online panels – and in marketing and commercial research in general – hard prompts are almost uniformly used, except perhaps for open-ended text entry questions. Special attention is needed with hard prompts to avoid generally inappropriate situations where respondents are forced to select only from response options that do not apply to them.

Item nonresponse prompts are closely related to approaches for dealing with non-substantive response options ('Don't know', 'No opinion', etc.) which we have partially discussed already in the introduction to questionnaire development (Section 2.3.1). There we followed Krosnick & Presser (2010) in not recommending the inclusion of such options unless explicitly needed. Another alternative is that non-substantive response options are displayed only after the

respondent does not answer a question, so they are offered together with the prompt for an item nonresponse. If these three basic alternatives for handling non-substantive options (offered, not offered, offered after a prompt) are combined with the three approaches to item nonresponse prompts (none, soft, hard), a researcher would theoretically have nine alternative strategies. In addition, an important insight into the matter – which further increases the number of combinations – can be obtained with the *familiarity pre-question*, which then enables us to pose a question only for respondents familiar enough with the topic, or with the *certainty post-question*, which asks how sure respondents were about the provided responses.

We cannot discuss all of these combinations here, since the research evidence is very limited. In general, we can say that the prevailing practice in marketing research seems to use hard prompts for item nonresponse. On the other hand, academic and governmental research seem to prefer soft prompts or no prompts for item nonresponse, with or without non-substantive responses. When we want to mirror interviewer-administered surveys (F2F and telephone modes) – where non-substantive response is initially not offered, but interviewers record it if explicitly volunteered by the respondent – displaying non-substantive response only after the corresponding item nonresponse soft prompt might be the appropriate selection (Ainsaar et al., 2013).

Practical importance and inconclusive results place these issues among the top priorities for future investigation. Ideally, future research will consider all combinations of item nonresponse prompts and approaches to handling non-substantive options in various essential settings: that is, prompting for all questions or only for certain questions, comparing different types of respondents (general population, specific population, or trained online panellists), various levels of topic salience, different question types, as well as the importance and position of questions. It is also crucial to evaluate different aspects of data quality, namely reliability, validity, item nonresponse, breakoffs, satisficing, response times, as well as engagement and satisfaction levels. Furthermore, it is important to consider the comparability of these approaches with standard procedures in interviewer-administered surveys instead of solely optimizing the interactive advantages of the web mode. Conceptual perception towards non-substantive responses as a legitimate response category – instead of being treated as 'lazy responses' or 'masked nonresponses' – is also important. The familiarity pre-question and the certainty post-question should be included in such an investigation, because they address the essential substantive aspects of the problem.

Additionally, capable software support is also very important here. For example, adding the 'don't know' category option to the substantive responses on the same page only after an item nonresponse occurs can pose a problem for a lot of the software.

### 2.3.3.2.5 Data-Piping

*Data-piping* (or *fills*) enables the use of responses from previous questions, previous surveys or external databases (e.g. sampling frame). It is commonly used to establish clearer instructions about what the question demands from the respondent (e.g. 'What is your relationship with the person you named Mark?'). In addition, it can provide personalized introductions to increase the respondent's engagement. For example, if the name of the respondent is known, it can be included in motivating statements. Use of this feature can be advantageous, but certain care with its implementation is required, especially with personalized messages included directly in the questionnaire, which may undermine the respondent's sense of privacy.

### 2.3.3.2.6 Progress Indicator

A progress indicator offers respondents orientation about how much of the questionnaire they have already completed. It is often used with the aim of keeping them engaged in order to prevent breakoffs. Its position and format vary greatly:

it can be immediately visible or presented on demand/click; displayed on every page or only at certain key transition points; in graphical (e.g. typically a progress bar) or text format (e.g. 60% or 5 out of 10 questions/pages/sections). There is little research evidence on the performance of these variations.

The problems with progress indicators appear in complex questionnaires, with skips of a large number of questions, which cause the progress indicator to advance inconsistently. For example, if several pages are skipped due to branching, the progress indicator may 'jump' from 10% to 35%, but then move very slightly across subsequent pages without skips. These effects of branching are computationally very hard to calibrate (Kaczmirek, 2009, p. 146).

When a progress indicator is not used, respondents complain about its absence (Lozar Manfreda, Batagelj & Vehovar, 2002), but studies have shown that its use generally does not contribute to a reduction in the breakoff rate, and it can even slightly increase the breakoff problem. This was established in a meta-study of randomized experiments for medium to long web surveys with a median time of 18 minutes (Villar, Callegaro, & Yang, 2013).

We can conclude that the progress indicator, especially when a lot of branching is used, has no benefit for medium or long web questionnaires. However, it can be appropriate for shorter and simpler questionnaires, where it moves linearly at a consistent and detectable rate. The inclusion of the progress indicator also presents an interesting ethical dilemma: from the respondent's perspective, the basic orientation of progress is advantageous, but it brings no gains and may potentially introduce problems for a researcher.

### 2.3.3.2.7 Other Interactive Features

Some examples of other interactive and dynamic features have already been mentioned in the discussion of question types (e.g. dynamic shading in tables), and we present some more in the next section on visual layout. However, web questionnaires can offer a number of other possibilities (e.g. occasional encouragement) that can improve the flow of questionnaire completion and are not mentioned in this review. In general, it is important to consider whether and how the use of these features is beneficial for the respondent and data quality and to weigh this against their potential drawbacks. With a lack of relevant research, reliance on common sense and a somewhat conservative approach may be the best choice.

### 2.3.3.3 Visual Layout

In F2F surveys we expect interviewers to be decently dressed, behave pleasantly and professionally, and provide engagement, motivation and support whenever needed. Similarly, in telephone surveys, we train them to be polite and convey questions in a non leading way, with a neutral voice. In web questionnaires such good practices of interaction with respondents need to be achieved using the visual layout of the questionnaire.

This layout usually follows some basic design principles. For example, in Western culture, the top and left positions are often treated as 'more frequent' or 'more positive', and visual distances may reflect distances in perception (Tourangeau, Couper, & Conrad, 2013). There is also a hierarchy from verbal presentation, which dominates, to numeric and visual presentation. Verbal descriptions are thus preferred. When numbers and graphics are nevertheless added (as in rating scales), consistency is required. We should also comply with basic web design and usability principles, and refrain from writing text in capital letters (Nielsen, 2000a; Schriver, 1997). In addition, we should visually highlight what is important, keeping the rest hidden in the background or shown only upon request.

The majority of web survey software already offers certain predefined professional visual designs, called *themes*, *skins* or *survey templates*. When they are in line with basic web design principles, little or no additional intervention is usually required or recommended. If it is necessary to customize profoundly the overall questionnaire

layout, it is advisable to involve someone with sufficient web design experience. The visual layout is technically defined by a special visual design language, called CSS (Cascading Style Sheets), which ensures a consistent look and format across web pages. Changing it usually requires design as well as some programming knowledge. However, researchers are not expected to be completely competent in the task of visual design, apart from minor adaptations, such as the inclusion of logotypes.

A researcher or programmer can make changes to the CSS file by direct modification of the code or by using a graphical user interface (GUI) if provided by the software. Depending on the software, some interfaces may allow modification of a variety of design elements (e.g. background colours, fonts, page structure) in a user-friendly way.

In addition to direct intervention in the basic visual layout, a researcher can also change structural elements, such as layout settings (e.g. logotype position, display of progress indicator, etc.), page structure (breaks, sections, blocks, etc.), formatting of instructions, definitions, help features, and other non-question sentences or pages. Specific editing of question text is also important, but should be done very carefully, particularly when the text editor allows changes to the font and size of the characters, numbers and symbols. Within this context, the size, structure and settings for pictures are also important.

Below we present some key elements, without going into further detail on the visual principles. More discussion on the subtle role of visuals can be found in Tourangeau, Conrad, & Couper, (2013, p. 88), Couper (2008, p. 84) and Dillman et al. (2009). We should keep in mind that checking and testing the look and feel of the web questionnaire in all key browsers, devices and operating systems is essential, particularly when making modifications to the standard visual layout.

### 2.3.3.3.1 Basic Layout

Elements of the basic visual layout of the questionnaire include font types and size, text width, spaces, colours, structure, and other common visual elements of web pages. They should not deviate much from the conventional styles used on other web pages. A general recommendation is to use a white or lightly shaded background, clearly visible standard fonts (e.g. Arial, Verdana, etc.), a professional look and feel, and keep the design consistent throughout the questionnaire. Couper (2008) provides more detail on all these issues. Despite their importance, there is little research on the basic layout specifics of web surveys, apart from some initial evidence that simple designs outperform 'fancy' ones (e.g. Dillman et al., 1998). Here, we should not forget that the quality of responses is the priority, and aesthetics should be used to foster this rather than increase the possibility of distraction. Within this context, Casey & Poropat (2014) explicitly demonstrated that classic aesthetic quality outperformed expressive aesthetic quality and had a positive correlation with the perceived ease of use of the web questionnaire, as well as with trust in the web survey researcher.

### 2.3.3.3.2 Logotypes

Logotypes increase legitimacy and remind the reader about the survey sponsor, survey organization, research project or online panel membership. They should be consistent with the whole visual layout, properly sized and positioned somewhere in the corner of the pages, in order to prevent a cluttered impression. They should not distract from the response process.

### 2.3.3.3.3 Position and Navigation of Other Action Buttons

The position and navigation of other action buttons that enable respondents to move from one screen to another are important. When Couper, Baker & Mechling (2011) manipulated the position of the 'Previous' button, they found no impact on breakoff rates, but increased use when positioned to the right of the 'Next' button. One explanation is that the increased use is due to mistakes, since such positioning opposes the general approach in surveys, other web pages and devices where 'Backward' is commonly to the left of 'Forward'. There are

certain arguments that it may be beneficial to position the 'Previous' button below the 'Next' button or to the right of it in the form of a hyperlink. However, to be on the safe side, prevailing practice is to have (a) both buttons, (b) close to each other, (c) placed at the bottom of the questionnaire on the right or in the middle of the page, (d) with sufficient space between them, and (e) with 'Next' on the right.

Sometimes other action buttons are used for special purposes, such as the 'Save and continue later' button that allows respondents to answer the web questionnaire in multiple sessions, or the 'Print' and 'Save' buttons, used particularly in business surveys. The recommendation is that their visibility and position should be consistent with the navigation buttons, but also reflect their importance.

### 2.3.3.3.4 Position and Formatting of Additional Instructions

Additional instructions about how to answer a particular survey question are sometimes needed. They are usually presented in a format that distinguishes them from the main question text (e.g. by using smaller or differently coloured fonts). The number of additional instructions and the level of detail provided are important dilemmas here. For example, we need to decide if we can assume that respondents understand that checkboxes denote the selection of more answers, or whether this should be explained in further instructions. In this specific case, the latter option may be more appropriate (i.e. a brief note in every question), although we generally recommend reducing instructions to a minimum and to situations where they are really needed. The overuse of instructions can lead respondents to start ignoring them. Of course, this all depends on the context and the target respondents, but in any case we need to balance the importance with the visual exposure.

### 2.3.3.3.5 Position and Formatting of Definitions

Definitions of terms used in the survey questions share the same recommendations and dilemmas as instructions. Peytchev, Conrad, Couper, & Tourangeau (2010) showed that increased visual exposure of definitions also increases their use. The study found the highest use of definitions when they were presented immediately after the question text, followed by a mouse-over appearance, and the lowest use when an additional click was required to display the definition.

### 2.3.3.3.6 Visual Layout of Help and Other Survey-Related Information

It is generally recommended to include a help email address, a toll-free number or a link to a help web page. Additional information, such as copyright, disclaimers, general information and FAQs, is also sometimes needed. In general, the visual layout of these elements should express their importance, for example by using smaller fonts, less prominent colours, and a non-central position at the top or bottom to avoid distractions. Details may be provided on hyperlinked pages.

### 2.3.3.3.7 Adaptability to Various Screen Resolutions

Because a web questionnaire is typically accessed on a variety of computers and other devices, it is important to ensure a robust visual appearance. The size and position of all questionnaire elements (including text, tables and pictures) should be defined using relative specifications and be flexible enough to adapt to the screen resolution. It is also important to ensure a proper appearance of navigation buttons, prevent horizontal scrolling, ensure proper positioning of pictures and prompts, etc. (Callegaro, 2010). A large part of the responsibility for this lies with the web survey software, but researchers should carefully verify its functioning.

To conclude the discussion of possibilities and issues regarding the visual layout of the questionnaire, it is important to remember that a careful and detailed evaluation of the visual elements is essential. This includes the verification of their perception by respondents, as well