

24 Survey-based studies

Sanjun Sun

Definition

The survey is probably the most common empirical research method in the social sciences and the humanities. It is a method designed to gather data about a human population (commonly referred to as a sample) through a sequence of focused questions. One distinguishing characteristic of the survey, according to Marsh (1982) and De Vaus (2011), concerns the form of its data: a structured set of data that forms a rectangle (or variable-by-case grid), in which rows usually represent cases (e.g. respondents, countries), columns represent variables (i.e., questions), and the cells contain information about a case’s attributes (e.g., respondents’ answers). Experiments and tests also use data in this form. The experimental method is different from the survey method in that with the former, “the variation between the attributes of people is created by intervention from an experimenter wanting to see if the intervention creates a difference” (De Vaus 2014: 5).

There are two data collection mechanisms used in surveys: standardized interviews and self-administered questionnaires. Standardized interviews are often conducted in person or over the phone, while self-administered questionnaires are used in group settings (e.g., in a classroom) or in individual settings (e.g., postal survey) (De Leeuw 2008). Each of these forms has a computer-assisted equivalent, such as Internet surveys, which are now more common than postal ones.

Surveys are often used to collect different types of data by asking questions, including (1) factual questions, regarding the demographic characteristics (e.g., age, gender, mother tongue, level of education) of the respondents, to help interpret the findings of the survey; (2) behavioral questions, regarding such things as personal history and language learning strategies; (3) and attitudinal questions, concerning the respondents’ attitudes, opinions, beliefs, interests, and values (Dörnyei and Taguchi 2010: 8–9).

Based on the format of the responses, there are two broad types of questions in a survey: open-ended and closed-ended questions. Open-ended questions do not provide specific answer alternatives and ask the respondent to provide his or her own answers. They can elicit rich information, but they are not easy
to code or quantify. As a result, survey researchers often use open-ended questions to help design closed questions or pilot studies and pretests. They also use open-ended questions sparingly in formal questionnaires (e.g., Holyk 2008). Closed-ended questions contain a predetermined set of answer alternatives for the respondent to select, and can be grouped into two classes: structured answer (dichotomous and multiple choice) and scales (McNabb 2010: 118). Scales are collections of items that measure the level of an underlying variable (DeVellis 2012: 15), which is placed along a quantitative continuum (e.g., from being very favorable to being very unfavorable in attitude or opinion).

Scales are usually used in the measurement of attitudes. There are many types of measurement scales, which fall into two broad categories: comparative and non-comparative. Comparative scales allow the respondent to compare two or more items, while a non-comparative scale allows the respondent to evaluate only a single item. The former category includes paired comparison, rank order, and others; the latter includes, among others, the Likert scale (Likert 1932), which is the most widely used rating scale (see Reddy and Acharyulu 2008: 101). The Likert scale consists of multiple items that typically are summed or averaged to yield an overall score (Brill 2008), and usually includes five to seven response categories, e.g., Strongly Agree, Agree, Neither Agree Nor Disagree, Disagree, and Strongly Disagree. An important consideration for the response categories is whether or not to include a middle position (e.g., “Neither Agree Nor Disagree”) or a “No opinion” option. There have been studies supporting both possibilities (e.g., Maitland 2008).

The steps in conducting survey research include: (1) determining the survey purpose and objectives; (2) defining and operationalizing key concepts; (3) developing specific research questions (and hypotheses); (4) determining the sampling procedure; (5) creating and pretesting the instrument; and (6) collecting, reducing, and analyzing data (e.g., Newman and McNeil 1998; Bartlett 2005). Questions/items need to be reliable and valid (Fowler 2014). Reliability is the extent to which an instrument yields consistent results upon testing and retesting. There are four common types of reliability estimates: test-retest, parallel forms, internal consistency, and inter-rater or inter-observer. In order to increase the reliability of a scale, one can include the number of items in the scale and eliminate items that have lower-than-average correlation with the other items, or have low inter-item consistency. These facets plus item difficulty are tested through item analysis (Angelelli 2004b). Validity refers to the extent to which an instrument measures what it has been designed to measure. Major measures of validity are face, content, construct, and criterion-related validity (see Brown 2001). Determination of face and content validity evidence is often made by expert judgment; a typical method includes several judges who rate each item in terms of its relevance to the content (Angelelli 2004b: 47–63; Kaplan and Saccuzzo 2013: 137). To think about the likely construct validity of a measure, the best way is “to see the full wording, formatting, and the location within the questionnaire of the
question or questions that were used to gather data on the construct” (Lavrakas 2008: 135).

In terms of research purpose, surveys are mainly used for descriptive or explanatory purposes (e.g., De Vaus, 2006). Marsh (1982: 6) claims that “[s]urveys and experiments are the only two methods known to me to test a hypothesis about how the world works.” This, however, is debatable. Dumont (2008: 25) argues that “the survey method does not provide empirical evidence that proves the existence of a causal relationship (only the experimental design can do this)” although it “can provide empirical evidence that a causal relationship between two (or more) variables does not exist.” In TIS, many kinds of research questions exist, including descriptive and explanatory ones. This makes survey research one of the most frequently used methods in TIS (e.g., Toury 2012: 263).

Origin

The roots of survey research can be traced back to population count (census) in ancient times for raising taxes or enlisting soldiers. The first surveys that resemble the modern social survey are believed to be Charles Booth’s surveys of the lives and occupations of the working classes in London, conducted in the early 1890s and reported in a multivolume book entitled Life and Labour of the People in London (e.g., Converse 1987/2009; Bulmer, Bales, and Sklar 1991)

Four basic developments in survey research method set early surveys apart from modern surveys (Wright and Marsden 2010: 3–4): (1) the development of sampling methods for drawing representative samples from human populations; (2) the theory of statistical inference for estimating population parameters from sample statistics; (3) the development of question design strategies for eliciting valid and reliable answers; and (4) the development of data analysis techniques for estimating complex statistical relationships among many variables. Major advances in such aspects as the idea of probability sampling, the use of structured questionnaires, and the basic tools of statistical analysis were already made before the First World War, but they did not spread overnight or were not applied to survey research (Marsh 1982; Alastalo 2008).

According to Groves (2011), the history of social survey research can be divided into three stages: (1) the first era (1930–60), during which the basic components of survey research and related tools were developed, and surveys were conducted in the private, academic, and government sectors; (2) the second era (1960–90), which witnessed significant growth in the use of surveys and quantitative information; and (3) the third era (1990–today), which witnessed the growth of alternative modes of data collection and of continuously produced process data from digital systems in all sectors and the Internet.

Early surveys did not involve sampling; investigators simply questioned everyone in a certain area. The idea of sampling was introduced in 1895,
when Anders Kiaer presented a report on “Representative Method” at a meeting of the International Statistical Institute (see Adèr and Mellenbergh 1999: 112). However, according to Biemer and Lyberg (2003: 10), “[i]n the 1930s and 1940s most of the basic survey sampling methods used today were developed.” In 1934, Jerzy Neyman delivered a paper entitled “On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection,” which established the foundations for modern sampling theory. Based on Neyman’s inference theory, survey sampling pioneers developed optimal methods for sample design and other statistical theories (see Heeringa, West, and Berglund 2010: 4). In 1936, George Gallup’s polling company, based on a small quota sample of 50,000 people (versus over 2 million returned questionnaires in the poll by Literary Digest magazine), correctly predicted that Roosevelt would win the presidential election, which proved that a small but representative sample yields better inferences than a large but unrepresentative sample with a low response rate.

Question wording did not get much attention from investigators in early surveys, as their focus was often on factual information, such as age and education. Things changed when researchers started to investigate people’s opinions, attitudes, and feelings. They found that minor changes in question wording, format, or the context could have a major impact on respondents’ answers (e.g., Cantril 1944). The first book devoted to question design was The Art of Asking Questions (1951) by Payne, which gave a concise checklist of 100 considerations in question formulation. Today, nobody would deny that questionnaire design is an art. Most researchers, however, would also say that it is a science. Since the early 1980s, survey researchers, cognitive scientists, and statisticians have started to investigate questionnaire design from a psychological perspective (e.g., Jabine et al. 1984; Tourangeau, Rips, and Rasinski 2000). They have developed tools for improving questionnaire design, such as cognitive interviewing for questionnaire pretesting (see Willis 2005).

Among data collection methods in social surveys, the face-to-face interview was the most common method until the end of the 1970s (Corbetta 2003: 142). Due to concerns over the escalating cost, telephone surveys became a standard survey practice by the 1980s. Mailed paper questionnaires have been used since the early twentieth century for measuring literate populations. Compared with the face-to-face interview, it is cheaper, but the turnaround time may last months. Since the world wide web became widely available around 1995, e-mail and web surveys have become increasingly common, enabling the expanded use of visual and audiovisual instruments, as well as innovations in sampling (Wright and Marsden 2010: 22). Computerization has also increased the availability of survey data and made secondary analyses attainable (Alastalo 2008: 34). Face-to-face interviews, despite the decline in volume, are still the mainstream method for complex academic surveys (Fu and Chu 2008: 290).
Uses

Surveys can be used for studying most, if not all, of the topics in TIS, such as quality evaluation (e.g., Kurz 2001), training (Chmiel 2010), roles of translators and interpreters (Angelelli 2004b), their social status (Dam and Zethsen 2008), and translation/interpreting strategies (PACTE 2008). To date, there have been hundreds of survey-based studies in the field of TIS published in English and many more in other languages.

Many survey-based studies have focused on the translation profession. For example, the special issue of Translation and Interpreting Studies, “Profession, Identity and Status: Translators and Interpreters as an Occupational Group” (4.2 [2009]) contains reports from survey studies. The survey has been one of the most productive methods in quality assessment research. Kurz (2001), for instance, offers an overview of 17 questionnaire-based studies on quality in conference interpreting, which focused on user expectations and/or user responses. Many research questions in translator and interpreter training can be addressed using surveys. For example, “How Effective Is Teaching Note-Taking to Trainee Interpreters?” (Chmiel 2010). Surveys are often used to investigate audience reception, e.g., Widler’s “A Survey among Audiences of Subtitled Films in Viennese Cinemas” (2004).

The most notable surveyor in TIS is probably Common Sense Advisory, an independent Massachusetts-based market research company with a research team that focuses on localization, translation, and interpreting practices. Since 2002, they have published on commonsenseadvisory.com over 500 reports, and their typical methods are questionnaire and interview. Their surveys of buyers and suppliers of translation services and technology vendors from the language industry perspective have covered various topics, such as translation quality, language services market (e.g., “Wages of Translation,” “Trends in Translation Pricing”), and best practices (“How to Buy Translation”). There are over a dozen reports on translation quality, including “The Buyer-Supplier Quality Gap: How Customer and Language Supplier Views of Translation Quality Differ,” and “Selling Different Levels of Quality.” Proz.com, a translation portal, frequently conducts quick polls, in which often over 1,000 translators and interpreters throughout the world participate. All kinds of questions are asked in these polls, such as “Do you have professional indemnity insurance?” and “Do you find translation theories useful for your professional practice?”

Sample studies

Surveys have become part of our life. Throughout the world, thousands of surveys are being undertaken every day (De Leeuw, Hox, and Dillman 2008). While some surveys are complex and may require years of development, some are simple and easy to conduct. A common type of survey is the public opinion poll (e.g., the Gallup Poll), which seeks opinions regarding social and political issues. Official statistics (e.g., censuses) are another common type, produced
by government agencies to facilitate decision making. Also, in the business world, surveys (e.g., the Job Attitude Scale) have been frequently used. According to Biemer and Lyberg (2003: 7), the survey industry employs more than 130,000 people in the United States alone. There are many professional organizations and societies (e.g., the International Association of Survey Statisticians) and academic journals (e.g., *Survey Research Methods, Journal of Survey Statistics and Methodology*) dedicated to improving survey work. Many universities have survey research centers or institutes, which are housed in the departments of statistics, sociology, psychology, education, communication, or business (Biemer and Lyberg 2003).

For examples of survey-based studies, one can refer to compendiums of questionnaires, scales, and tests. In the field of second language research, for instance, Dörnyei and Taguchi (2010) present a list of published L2 questionnaires on language anxiety, attitudes, language course evaluation, language learner beliefs, language learning motivation, language learning strategies, language learning styles, needs analysis, self-evaluation, teacher evaluation, etc. Of these questionnaires, a notable example is the Strategy Inventory for Language Learning (SILL) (Oxford 1990), which is the most widely employed second language learning strategy questionnaire. It has several versions. SILL Version 5.1 is for native English speakers and contains 80 items assessing the frequency of strategy use, while Version 7.0 is geared to students of English as a second or foreign language and contains 50 items. SILL uses a 5-point Likert scale (for all versions) ranging from “Never or Almost Never” to “Always or Almost Always.” It has been extensively field tested from a psychometric viewpoint. For example, the internal consistency reliability of a slightly earlier, 121-item version of the SILL using Cronbach’s alpha is .96 for a 1,200-person university sample; content validity is .95, based on classificatory agreement between two independent raters, who “blindly” matched each of the SILL items with the strategies in the comprehensive list; concurrent validity of the 121-item form is found in strong, statistically significant relationships between SILL results and self-ratings of target language proficiency and motivation in the 1,200-person university sample; no evidence of social desirability response bias appeared in three samples (Oxford 1990: 255). SILL Versions 5.1 and 7.0 have also been evaluated and are high in reliability and validity (Oxford 1996). They have been translated into various languages and tested (e.g., Demirel 2009). In the field of business training, Cook (2007) provides nearly 100 assessment questionnaires and checklists regarding communication, performance management, personal effectiveness, teamwork, training and development, and others.

Many research topics in translator and interpreter training can be studied through a survey. Li (2000), for example, conducted a questionnaire survey of practicing translators in Hong Kong in an attempt to study ways of tailoring translation programs to social needs. Ulrych (2005) investigated translation teaching practices and curricula in 41 academic institutions in Europe and North America. With an online questionnaire, Katan (2009) investigated
890 respondents on how translation should be taught, on the role and status of the profession (ideally and in practice), and on personal satisfaction. As admitted by the author, however, issues of validity and reliability were not given adequate attention. Another issue pertains to the Likert scale. Katan (2009: 189) mentioned that for many questions “a 5-point Likert rating scale was used.” One of the items in his questionnaire, for example, was “How satisfied are you with your present job in comparison with your initial expectations regarding the field of translating/interpreting” followed by five response options: Extremely, Pretty, Fairly, Not Very, and Not at All. However, this is not a Likert item. First, in a Likert scale, it is usually a specific statement (e.g., “As an interpreter, I should adhere to the conversational conventions established by the speakers”) expressing an attitude or opinion that is followed by the response points (Strongly Agree, Agree, Neither Agree Nor Disagree, Disagree, Strongly Disagree); it is not a question. Second, a Likert scale contains multiple items to measure the same latent variable, which need to be summed or averaged to produce a more reliable measure. According to Brill (2008: 427), the previous example item from Katan (2009) is just “an ordered-category rating item with the Likert-like properties of a bipolar and balanced response set.”

Angelelli (2004b) provides an outstanding example of measurement instrument development in the field of TIS. The book-length report details how she developed the Interpreter’s Interpersonal Role Inventory, and describes ways to improve and test the reliability and validity of the measurement scale. For example, she took advantage of surveys, feedback from peers at a seminar for measurement instrument design, literature reviews, and interviews with interpreters in composing the 80 initial items; expert opinion and focus groups were then used to establish the content validity of the scale. In that study, Angelelli recognized five sub-components of the construct of visibility: alignment, trust/respect, affect, cultural gaps, and communication rules.

PACTE Group (2008) presents a typical study using mixed methods. In this study, the researchers designed an experiment for the validation of their translation competence model. Their data collection instruments included: (1) texts and translations, (2) translation protocols, which were recorded using software programs, (3) direct observation, (4) questionnaires, and (5) retrospective interviews. Three questionnaires were used: (1) an initial questionnaire to ensure that the participants selected for inclusion in the experimental groups fulfilled selection criteria; (2) a questionnaire eliciting information on translation problems encountered during the translation process; and (3) a questionnaire designed to obtain information on the participant’s knowledge of translation, which was conceptualized in terms of seven dimensions.

**Conclusion and potential applications**

Investigating the attitudes, opinions, beliefs, interests, and behavior of stakeholders in the field of translation and interpreting through a survey can bring
new perspectives to existing research questions in TIS, raise awareness of new concerns, help researchers better understand the profession, and reach beyond the ivory tower.

Yet, in using the survey method, many existing studies have methodological inadequacies. Kurz (2001), in her review of 17 questionnaire-based studies on quality in conference interpreting, found that there was little comparability among the individual surveys. Similarly, Pöchhacker (2009), after reviewing 40 survey studies on the profession of conference interpreting, observes that most of the findings are not solid enough in terms of sampling and there is a lack of thoroughly tested instruments (cf. Angelelli 2004b). This has several implications. First, researchers need to pay more attention to reporting guidelines or best practices. For example, survey-based studies should provide access to the questionnaire items, describe participant characteristics and sampling procedures in detail, and provide information on the validity and reliability of the instrument (see Kelley et al. 2003; Angelelli 2004b; Bennett et al. 2011). A detailed description of the methods used enables the reader to evaluate the methods and results, and facilitates research replication. Second, TIS researchers need to develop standard questionnaires (e.g., Angelelli’s Interpreter’s Interpersonal Role Inventory) and replicate previous studies to determine whether there are any significant differences between cultures or socioeconomic conditions.

It should be noted that the survey method can be used alone or with other methods in a study. In TIS, especially in translation process research, the idea of triangulation has been adopted. For example, process researchers have typically used questionnaires, verbal protocols, keystroke logging, and/or eye-tracking in an experimental setting (see Alvstad, Hild, and Tiselius 2011), while interpreting researchers have used a combination of surveys, observations, and interviews (see Angelelli 2001 and 2004a).

Further reading

Fowler (2014) is probably the most widely read and quoted book on survey research methods; it provides a comprehensive introduction. Dörnyei and Taguchi (2010) detail how to produce and use questionnaires in second language research, and include a list of published L2 questionnaires. Fishman and Galguera (2003) introduce in an accessible way how to construct tests for research purposes. Angelelli (2004b) provides an excellent example of measurement instrument development in the field of TIS.

References


