Questionnaires in language teaching research

Peter Yongqi Gu  
Victoria University of Wellington, New Zealand

Questionnaires are often used to examine people’s attitudes, beliefs and behaviours in language learning and teaching. The data we get from questionnaire research can be especially insightful and satisfying when patterns emerge from a large number of respondents, when apparent differences or similarities are found among groups, or when relationships are ascertained among variables. As researchers, we feel empowered when we make recommendations for learning and teaching if the pattern we find is not only salient and strong, but also attested on a large scale.

However, by nature, survey research is exploratory and shallow. It often does not go beyond pattern finding or relationship mapping. The patterns themselves are long on description and short on explanation. And of course, with every pattern there are exceptions. Moreover, just like any other research tool, the questionnaire we use will need to be valid, and consistently so, in terms of eliciting what it is designed to elicit. Most consumers of our research only care about our conclusions and pedagogical implications. As researchers, however, it is our responsibility to make sure that we have selected the right tool, and that the tool we have used to reach our conclusions is indeed trustworthy.

The questionnaire as a research instrument has been a concern for applied linguistics for some time. As early as 1990, Reid (1990) traced the processes in developing her learning styles measure and boldly revealed the ‘dirty laundry of ESL [English as a second language] survey research’. Luppescu and Day (1990) attempted to develop a questionnaire to canvass the attitudes and beliefs of Japanese learners and teachers towards the learning of English, but found it difficult to validate the student questionnaire. In fact, instead of arriving at any conclusions about student beliefs, they presented a ‘more important conclusion’, that ‘questionnaire data should not blindly be accepted or considered meaningful unless they have been properly validated’ (p. 125).

Corresponding author:  
Peter Yongqi Gu, Victoria University of Wellington, Wellington, New Zealand.  
Email: peter.gu@vuw.ac.nz
Compared to other fields such as the social sciences, education, marketing, and medicine that extensively make use of questionnaires, applied linguists have paid only cursory attention to the questionnaire tool itself. Despite an occasional paper on questionnaire validation (e.g. Brown, 1997) or issues involved in using the Likert scale (e.g. Busch, 1993; Gu, Wen, & Wu, 1995), questionnaire design and validation remained a topic rarely touched upon until the end of the 20th century. Even up until this day, the only book-length volumes on questionnaire design are Brown (2001) and Dörnyei and Taguchi (2010; first edition published in 2003). Only a few papers (e.g. Petrić & Czárl, 2003) showcase how a questionnaire can be validated.

With these precious efforts, novice researchers today have much clearer guidance in the process of questionnaire design and validation. However, applied linguists have barely explored the major issues involved in the analysis of questionnaire data. One example would be the analysis of Likert scale data. While researchers (e.g. Broca, 2015) have discussed a range of potential rater responses, the Likert scale which involves bipolar scales with normally 5 or 7 categories is the de facto response format for the majority of questionnaires. The most often seen opinion scales normally range from one end (strongly disagree) to the other end (strongly agree); whereas the most often seen behaviour scales range from ‘never’ to ‘always’. The problem is: the resulting data is ordinal data and, as such, should not be treated as interval data in statistical analyses. For one thing, when the verbal responses are coded into numbers, e.g. never = 1 and always = 5, the interval between 1 and 2 does not equal that between 2 and 3.

Hatch and Lazaraton (1991) offer a potential solution and reassure us that ‘wider scales encourage more precision in rating and thus approach equal intervals’ (p. 57), but human judgment and ranking are by nature imprecise; and we have no guidance as to how many categories are optimal. Likewise, Brown (2011) reminds us of the distinction between individual Likert-type items and Likert scales that combine a number of items. He contends that the resulting scale scores can be taken as interval data, in the same way as we treat test scores that result from individual items being added up. To me, this explanation is partially satisfactory only because it does not deny the non-continuous and ranking nature of the data. Moreover, factor analysis (both exploratory and confirmatory), the most common way of revealing the construct validity of Likert scale questionnaires, makes use of individual items.

It is now time for the field to pay more attention not just to what a questionnaire study reveals, but also to how the questionnaire is designed, validated, and analysed. In this sense, this issue of Language Teaching Research is an excellent example. All five articles in this issue involve the use of a survey to ask teachers or learners about their beliefs and practices (Graus & Coppen; Buss), classroom willingness to communicate (Pawlak, Mystkowska-Wiertelak, & Bielak), classroom anxiety (Dolean), and self-regulation strategies (Seker).

Graus and Coppen’s study of trainee teacher beliefs on grammar instruction is a true gem in questionnaire research. It documented the clear procedures in theoretically conceptualizing the construct of ‘teacher beliefs on grammar instruction’, in operationalizing the construct into questionnaire sections and items, in validating the instrument, and in presenting the analysis and findings. With a meticulously crafted questionnaire tool
administered among 832 student teachers from nine universities in the Netherlands, the findings sound very convincing to me. For example, the student teachers in this study became increasingly aware of the importance of focusing on meaning (as opposed to form-focused instruction), although this preference was found to be dependent on the level of target students and the complexity of the structures under focus.

Seker developed a ‘self-regulated language learning questionnaire’ and administered it among 222 Turkish university students at Level A2 on the Common European Framework of Reference. The process of conceptualizing ‘self-regulated language learning’ and operationalizing the construct into the questionnaire was also described in detail. Exploratory factor analysis reduced the original 43 Likert-type items into a 30-item questionnaire, and the five resulting factors largely corresponded to the five sub-constructs in the conceptualization. Spearman’s rank-order correlation and stepwise multiple regression were then used to link different components of self-regulated learning to English language achievement.

Buss studied teacher beliefs and practices regarding pronunciation. Sixty Brazilian EFL teachers answered a 74-item questionnaire adapted from two existing questionnaires on the teaching of pronunciation. The participants’ practices in teaching pronunciation, their beliefs, and their self-reported training were reported in terms of frequency tallies and percentages. This is a typical exploratory survey that shows a pattern, which should in turn inform teachers and teacher trainers in terms of where the perceived difficulties lie and where the teaching foci currently rest.

Questionnaires were the main research instruments in the previous three studies. The next two studies in this issue made use of questionnaire results as only part of the data. Dolean examined the effects of teaching songs on a group of teenagers’ anxiety level in their learning of French as a foreign language during a 5-week intervention programme. In this sense, the questionnaire Dolean used to elicit the students’ anxiety levels before and after the intervention programme can be seen as pre- and post-tests. The 33-item Foreign Language Classroom Anxiety Scale (FLCAS) was developed by Horwitz and her colleagues in the 1980s (Horwitz, Horwitz, & Cope, 1986). The only ‘adaption’ that was done in this study was changing the wording of ‘foreign language’ to ‘French’.

Pawlak, Mystkowska-Wiertelak, and Bielak’s study focused on the changing nature of classroom willingness to communicate (WTC) during English conversation classes among 60 Polish university students. Their main research tool was the students’ self-ratings of their own WTC every five minutes, on a scale ranging from −10 (total unwillingness) to +10 (total willingness). A questionnaire was used at the end of each class, with a main aim of eliciting the respondent’s reactions to the factors affecting their increase or decrease in WTC.

Questionnaire research is often seen as ‘quick and dirty’. While the administration of a questionnaire does seem quick, the development, validation, and analysis of the questionnaire are far from a quick and easy process. Another common practice that makes it feel easy is the adoption/adaptation of existing questionnaires. Again, this is not as straightforward as it seems. In commenting on Park’s (2014) recent validation of the FLCAS, Horwitz (2016) emphasized the point that the factor structure of the foreign language anxiety scale may vary from group to group, and that:
future empirical efforts to understand the components of FLA [foreign language anxiety] should clearly delineate the specific learner population and learning context being examined to determine if, and if so, how the construct of FLA differs across learning populations and situations. (p. 4)

I quickly add that this wisdom rings true in all questionnaire research, no matter what the construct is. Research findings are as good as the tools we use to obtain our data. This is especially true for questionnaire research.

Arguably, the five articles in this issue show where we are as a field in terms of questionnaire use and questionnaire research literacy in language teaching research. On the one hand, we should be proud of what we have achieved. On the other hand, we have every reason to work harder.

References


