Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool*

PING LI
FAN ZHANG
ERLFANG TSAI
BRENDAN PULS
Pennsylvania State University

The language history questionnaire (LHQ) is an important tool for assessing the linguistic background of bilinguals or second language learners and for generating self-reported proficiency in multiple languages. Previously we developed a generic LHQ based on the most commonly asked questions in published studies (Li, Sepanski & Zhao, 2006). Here we report a new web-based interface (LHQ 2.0) that has more flexibility in functionality, more accuracy in data recording, and more privacy for users and data. LHQ 2.0 achieves flexibility, accuracy, and privacy by using dynamic web-design features for enhanced data collection. It allows investigators to dynamically construct individualized LHQs on the fly and allows participants to complete the LHQ online in multiple languages. Investigators can download and delete the LHQ results and update their user and experiment information on the web. Privacy issues are handled through the online assignment of a unique ID number for each study and password-protected access to data.

Keywords: LHQ, self-rated language proficiency, language background, language dominance, questionnaire, web interface, online data collection

Introduction

The language history questionnaire (LHQ) is an important tool for assessing the linguistic background of bilinguals or second language learners, the context and habits of language use, proficiency in multiple languages, and dominance and cultural identity of the languages acquired. Outcomes from such assessments have often been used to predict or correlate with learners' linguistic performance in cognitive and behavioral tests. For example, Dunn and Fox Tree (2009) developed and validated a bilingual dominance gradient that accounted for use, acquisition, and restructuring of both languages. Bedore, Pena, Summers, Boerger, Resendiz, Greene, Bohman and Gillam (2012) showed that bilingual dominance and proficiency classifications vary according to the types of measures employed in assessments, such as semantics versus morphosyntax tasks. Gollan, Weissberger, Runnqvist, Montoya and Cera (2012) examined the correlation of bilingual dominance and proficiency indices given scores from self-ratings, interviews, and picture naming tests. Although these assessment tools have been independently developed for different purposes (e.g., Bedore et al. focused on bilingual children), they also share some common features and have similar question items.

Despite the large amount of language assessment data from various measures and scales, there has been no standardized language history questionnaire available that integrates the various measures of language background, proficiency, usage, and dominance. Investigators still tend to design their own questionnaire for each particular study they conduct, making it difficult to compare their results due to the different questions, measures, and scales used. Recognizing this problem, Li, Sepanski and Zhao (2006) developed a generic LHQ by examining 41 published studies and identifying the most commonly asked questions in these studies. Li et al.'s LHQ was among the first attempts at providing researchers with comparable

Address for correspondence:
Ping Li, Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA 16802, USA
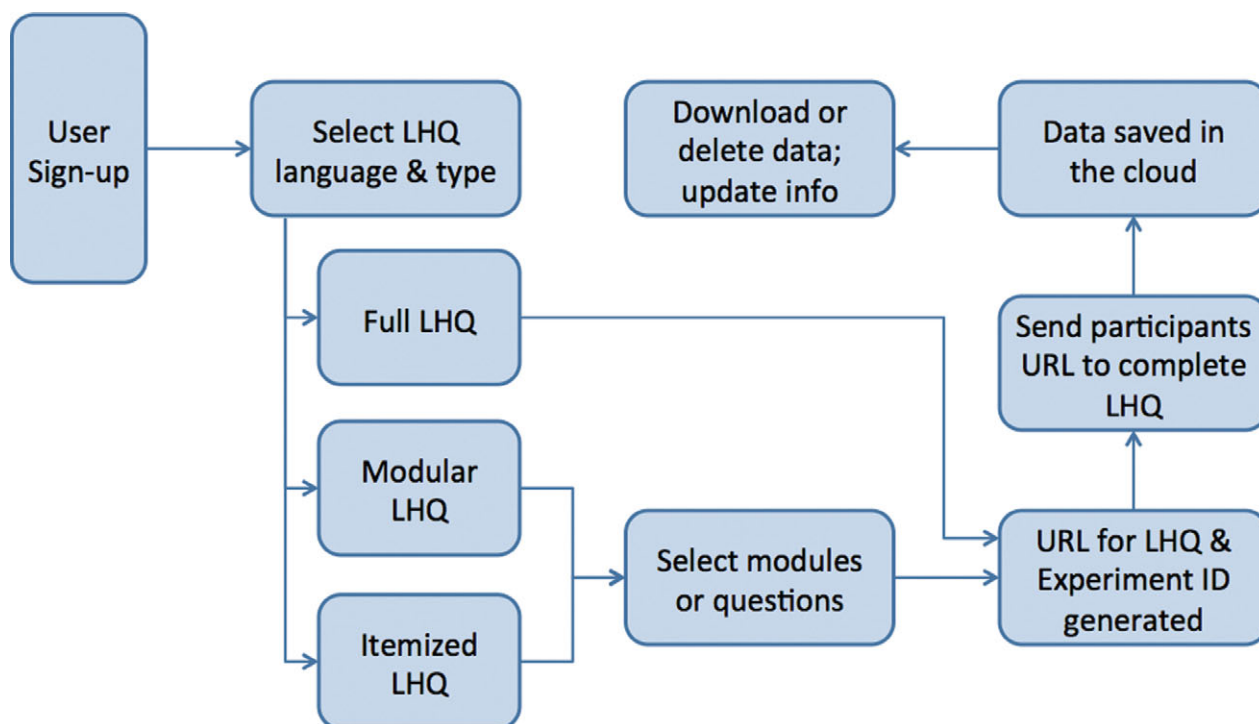*pul8@psu.edu*

Figure 1. (Colour online) Schematic diagram of the LHQ 2.0 process. See text for further explanations.

standards and has since been used in several dozens of published studies (e.g., Chandrasekaran, Krishnan & Gandour, 2009; Crinion, Green, Chung, Ali, Grogan, Price, Mechelli & Price, 2009; Krishnan, Gandour & Bidelman, 2010; see also Google Scholar, 2013).

The importance of web-based questionnaires for bilingualism and second language acquisition research has been clearly recognized by researchers (see Wilson & Dewaele, 2010). Realizing the advantage of web-based data collection, Li et al. (2006) made their LHQ available to the research community through a web-based interface. However, researchers have so far preferred to download and print out the original LHQ rather than use the online version. A number of obstacles may have prevented researchers from making full use of the online version of the original LHQ. First, the original LHQ's online version was based on web technology of the early 2000s and did not have the functionality, flexibility, and usability as a website developed today. Second, researchers could not easily adapt the questionnaire to their own needs or focuses of research. Third, the RTF format of the output data could not be easily computed in aggregated form. Fourth, the online LHQ was available only in English, which could have limited its usefulness in countries where English is not the native language.

Considering these problems, we have implemented a new web-based interface that overcomes the limitations of the original LHQ by using dynamic web-design features. The new dynamic web-based LHQ allows researchers to collect and store LHQ results in the cloud, and it can greatly facilitate data collection and data analyses in the context of language background and language assessment.

Figure 1 presents an overview chart of the process associated with the new LHQ (henceforth LHQ 2.0). LHQ 2.0 is freely available to researchers at http://blclab.org/language-history-questionnaire/.

As shown in this figure, LHQ 2.0 consists of several components:

(i) the investigator provides some basic information about the study in which LHQ data are to be collected;

(ii) the investigator selects the type of LHQ that is needed, either the full LHQ, a modular LHQ, or itemized LHQ;

(iii) the investigator chooses the language of the LHQ for his or her participants to complete;

(iv) the investigator receives the URL and ID numbers associated with his or her experiment, and then passes the URL onto participants to complete the LHQ;

(v) LHQ data are automatically and cumulatively saved in a spreadsheet as individual participants complete the LHQ; and

(vi) the investigator, at any time in the LHQ process, can access or delete the LHQ data, or update their user or experiment information through an account management page (see Figure 7 and discussion below).

Figure 2. (Colour online) Screenshot of an example sign-up page. More languages will be added to the "Select LHQ language" tab in the near future.

In addition to a new user-oriented and user-friendly interface, LHQ 2.0 is built around three key design features: flexibility, accuracy, and privacy.[1] We highlight these features below from the user's perspective.

LHQ creation and user sign-up

There are two types of LHQ users: the investigator or experimenter (henceforth investigator) of a study and the participants involved in the study. The original LHQ was accessible online to the latter users anytime, but the investigator or experimenter had no control of when the participants would complete the LHQ, and could not check the data online. The new LHQ 2.0 gives the investigator full control of both the data collection process and the actual data. To achieve this, the system requires the investigator to sign up at the beginning to provide basic information about his or her study.

Figure 2 shows a screenshot of an example sign-up page. Specifically, the investigator provides the following type of information: his or her name, email address, institutional affiliation, a password, a short title of the experiment, the number of participants expected to complete the LHQ (the system does not set a limit on the number of participants; this information is used by the system to create the number of lines in the experiment roster), and the date the experimenter expects to receive all participant responses to the LHQ (after which the LHQ data can be retrieved from the cloud within 60 days). Such information is then used by the system to generate automatic and unique ID numbers and URLs for the study, so that the investigator can keep track of the associated

LHQ data with the experiment. To maximize usability and flexibility, the sign-up page (as well as the LHQ form itself) provides help icons (question marks within blue dots) to explain what is required of each field for completion.

The final steps in the sign-up process are to select one of the three types of LHQ: the FULL, the MODULAR, or the ITEMIZED LHQ (see more details in the next section) and to choose the language of the LHQ for the participants to complete. Once the sign-up process is complete, the investigator is provided with a unique Experiment ID number (which is needed for accessing the LHQ data online) along with a URL that is also uniquely assigned to the particular study, as shown in Figure 3. The investigator can also download an experiment roster that simply contains a column of numbers (Participant IDs) to be matched with participant names (see Figure 3), which allows the experimenter to keep track of which participant has completed the LHQ and which participant is matched to which Participant ID number. Since no personal names will be collected on the LHQ, it is the investigator's responsibility to keep track of the Participant ID and the corresponding participant names. The sign-up completion page also provides a web link to the account management page, where the investigator can download or delete their LHQ data (see Figure 7 and discussion below).

LHQ tailor-made for specific needs

A key motivation behind the design of LHQ 2.0 is the consideration that different investigators may wish to look at different aspects of bilingual learners' language background due to the nature or the focus of their specific research; sometimes investigators may be interested in only a subset of the questions from the full LHQ (which contains 22 questions), and as such, asking the

---

[1] Following feedback from users and researchers, we have also made several wording and organization changes to the original LHQ, including the merger of several questions that were related or redundant.
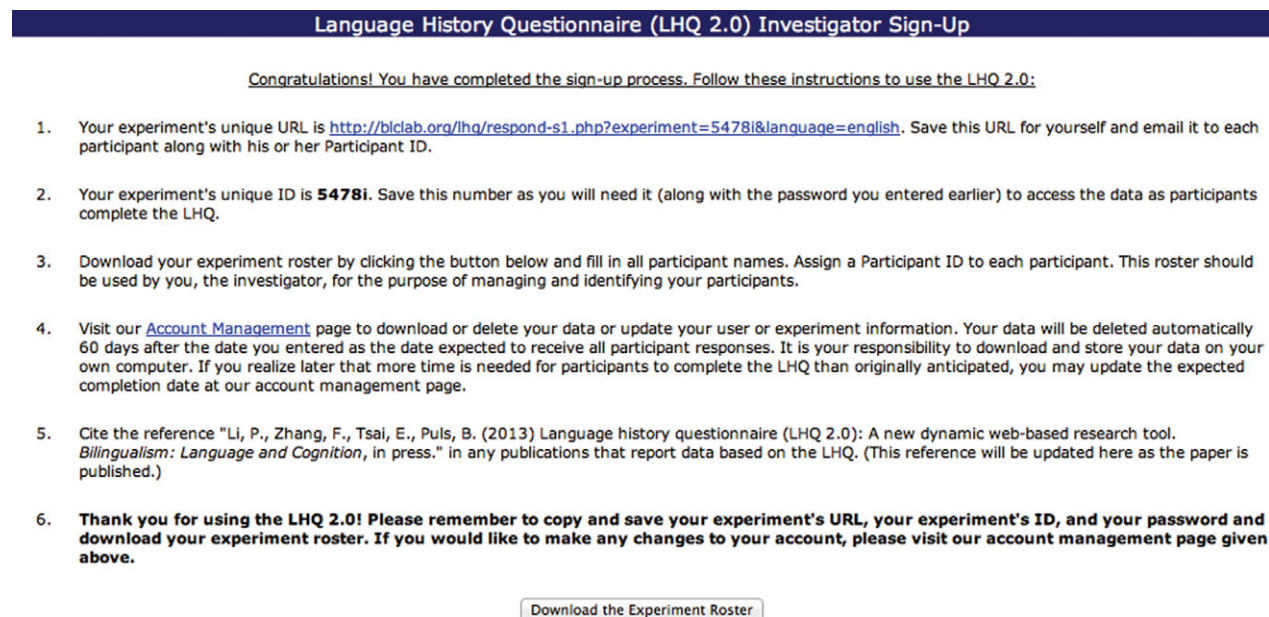
Figure 3. (Colour online) Screenshot of a completed sign-up page, with information about the use of LHQ 2.0 and access to LHQ data.

participants to complete the whole questionnaire could be a waste of time on both the investigator's part and the participants' part. The LHQ 2.0 interface now offers two ways for investigators to dynamically construct their own individualized LHQ on the fly to suit different research needs: the modular LHQ and the itemized LHQ.

For the modular LHQ, we have considered the prevalent uses of LHQ in the literature and implemented four different modules. Each of the four modules contains a subset of questionnaire items pertaining to the users' linguistic history (BACKGROUND), proficiency in first, second, or multiple languages (PROFICIENCY), context and habits of language use (USAGE), and dominance and cultural identity of the languages acquired (DOMINANCE). When the modular LHQ option is selected at sign-up (see Figure 2 above), the user will be brought to a new page to choose history, proficiency, usage, or dominance, or any combination of these modules. The subset of question items based on the selected module (or modules) is then automatically generated, along with a few items related to the basic demographic information of the participant (e.g., age, gender, and education). Figure 4 presents a snapshot of the background module of the LHQ. As is shown, such questions have been used previously by most researchers who construct this type of questionnaire (see Li et al., 2006, p. 203, Table 1) given that they relate to self-reported listening, speaking, reading, and writing abilities.

For the itemized LHQ, very much the same steps as the modular LHQ are involved, except that here the researcher has even more flexibility in controlling what question items are to be included in the LHQ. That is, the

researcher can select questions on an item-by-item basis rather than by modules, again plus a few items related to the basic demographic information of the participant (e.g., age, gender, and education) that will be automatically included in all LHQs. As of yet, LHQ 2.0 does not allow investigators to dynamically add their own specific questions that are not available in the full LHQ, but plans are in place for implementing this function in the next version of the LHQ.

To increase the usability of the LHQ for participants whose native language is not English, we have also implemented a multilingual function which allows researchers to choose the language of the LHQ for their participants to complete. Figure 5 presents a side-by-side view of the basic questions for all participants in English, French, and Chinese (available in simplified or traditional characters). Versions in these three languages are now available in both web and paper forms, and users will also see a gradual addition of other languages into this new function (e.g., Spanish, Italian, and Portuguese, which already have paper versions in PDF format).

In all cases, as already noted, the investigator will receive a unique Experiment ID and a unique URL associated with their customized LHQ constructed online. The investigator can then send the URL to the participants and instruct them to complete the LHQ on the web.

Data output for easy data analysis

Another significant improvement of LHQ 2.0 over the original LHQ is that LHQ 2.0 provides more

Figure 4. (Colour online) Screenshot of example question items from a modular LHQ (background).

accuracy and convenience for data output and subsequent analyses. Importantly, data are stored cumulatively as the participants complete the LHQ, in a format (Excel spreadsheet) that facilitates further data aggregation or analyses. The spreadsheet is organized in the order of (i) time of LHQ completion by participants (rows) and (ii) question items in the selected LHQ (columns), and all data are converted to numerical values for easy data analysis.

Figure 6 presents an example page of the data spreadsheet. As can be seen, the first two rows contain the study information entered on the sign-up page (e.g., the investigator's contact information, the title of the experiment, the expected date of completion), the fourth row contains the complete LHQ questions, and the fifth row the individual data fields for each question item of the LHQ. The third row is left empty for visual clarity. Each row starting from the sixth row contains data from

a given participant (with the first column containing the Participant ID), which extends from column B onward up to column IP (depending on how many items participants have responded to).

Once any participant has completed the LHQ, the researcher can use the Experiment ID along with a password (entered at sign-up; see Figure 2) to access the data. The researcher has up to 60 days (starting from the date when all participants are expected to have completed the LHQ) to download and remove the LHQ data from the cloud. The date when the researcher expects all participants to have completed the LHQ is specified at the sign-up (see Figure 2). If the researcher realizes later that more time is needed than originally anticipated for all the participants to complete the LHQ, he or she can amend the expected completion date through the account management page (see Figure 7).
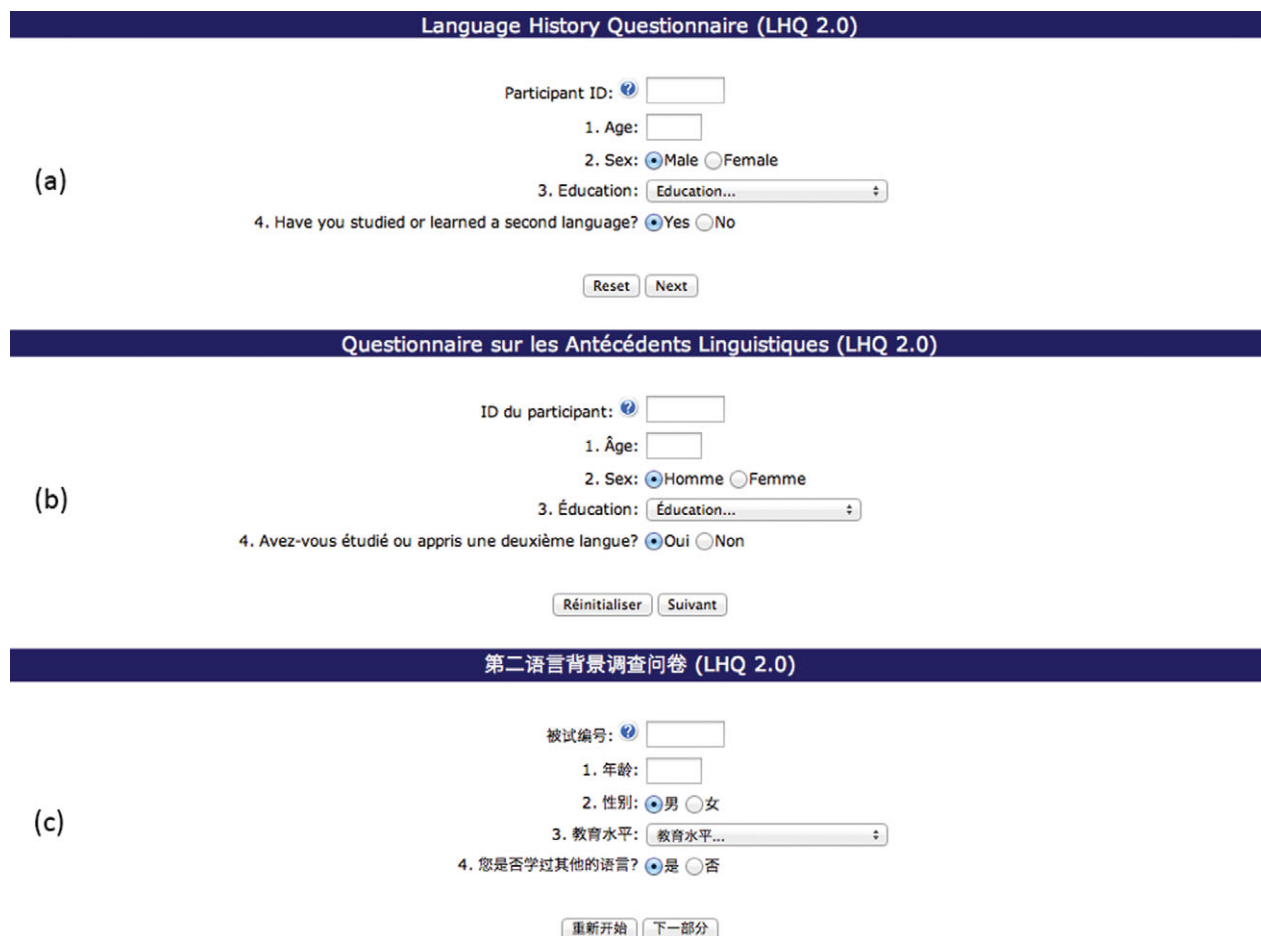
Figure 5. (Colour online) Multilingual function of LHQ 2.0: screenshots of the (a) English version, (b) French version, and (c) Chinese version (simplified characters).

Figure 6. Screenshot of an example output data file.

The automatic data storage and retrieval system implemented in LHQ 2.0 also has the advantage over the old LHQ of producing more accurate data. First, as participants are completing the LHQ online, it eliminates potential errors associated with participants' hand-written results; in most cases, the participant clicks on the relevant buttons, or uses drop-down menus to answer the questions, instead of writing verbal responses (see Figure 4 for an example). Second, the experimenter no longer needs to spend time manually coding and transcribing LHQ results into a spreadsheet or other data-format sheet, as the data are automatically saved in a spreadsheet file. This eliminates potential errors associated with the manual coding and transcribing process, which involves examining and transferring data points from one form (on paper or in individual participant files) into aggregated digital form for further data analysis. Finally, the streamlined web process is not only more accurate

Figure 7. (Colour online) Screenshot of account management page for LHQ data download and deletion, and for user and experiment information update.

but also time-saving. On the basis of our experience, we estimate that an average of 40–50 hours may be saved by using LHQ 2.0 for an experiment with 20 participants (e.g., one hour spent on bringing the participant into the lab for data collection and one hour on coding and transcribing the data).

Privacy of users and data

A final important issue for consideration in the development of LHQ 2.0 is the privacy of participants and data. Some researchers may prefer to use the paper-and-pencil version of LHQ because of privacy concerns. The LHQ 2.0 now fully considers user and data privacy issues. As mentioned earlier, the collected LHQ data are accessible only to the investigator or experimenter with a valid Experiment ID and a uniquely assigned password (generated through the sign-up process). In addition, the LHQ form will also only be accessible to participants via an individualized URL generated for each study after the investigator completes the sign-up. Figure 7 shows an example of how the investigator can access the LHQ data and control the removal of the data through an account management interface. This interface also allows the investigator to update user and experiment information when changes occur with the user (e.g., the investigator has a new institutional affiliation) or with the experiment (e.g., more time is needed for LHQ data collection).

To maximize the privacy of participants, LHQ 2.0 will not collect personally identifiable information. Rather, data identity is handled through the online assignment of an Experiment ID randomly generated for a given study, and each participant will complete the LHQ with the Experiment ID along with his or her Participant ID (as outlined earlier). The investigator (rather than the LHQ interface) is responsible for the identity of their LHQ data and participants. In this way, LHQ data can be correctly and accurately recorded and retrieved and at the same time, privacy protected. Finally, all data from a study will be removed from the cloud 60 days after the expected completion date (which is specified by the researcher during sign-up and can later be amended). Not storing LHQ data permanently in the cloud allows as many LHQ datasets as possible to be collected, while at the same time preserving data privacy and giving the investigator full control of their own data.[2]

Conclusions and summary

Researchers in language studies often find it important and necessary to assess the linguistic background and self-reported proficiency, usage, and dominance of their participants, and web technology has much to offer for this purpose. We had previously provided a generic LHQ (Li et al., 2006) to the research community, and over the years we have received colleagues' comments and feedback on the need of improved web functionality and usability as well as more flexibility, accuracy, and data privacy. Keeping these design features in mind, we have developed a new language history questionnaire, LHQ 2.0, using currently available web technology.[3]

LHQ 2.0 incorporates dynamic web design features for enhanced data collection and has significantly improved in functionality over the original LHQ, including its flexibility in dynamically generating customized LHQs on the fly, its accuracy in data recording and retrieval through individualized URLs, and its preservation of privacy of participants and data through automatically generated ID numbers and password-protected access and removal of data. To accommodate researchers who are interested in collecting data from participants of various bilingual backgrounds, we have implemented a multilingual function that allows researchers to choose the language of the LHQ for their participants to complete.

---

[2] The LHQ system will neither store nor own the data in any fashion. All data collected belong to the user and the investigator.

[3] One researcher asked whether the LHQ 2.0 features can be incorporated into Google Forms. Our response is that use of Google Forms can achieve some of the accuracy reported here but provides neither the flexibility of LHQ 2.0 nor the privacy for users and data.

Additionally, we have provided paper LHQ forms in Chinese (in simplified or traditional characters), French, Italian, Spanish, and Portuguese, as well as English, downloadable from the LHQ 2.0 website.

The new LHQ is easy to use and mostly self-explanatory. To recapitulate Figure 1, investigators can follow three simple steps in deploying LHQ 2.0 for their study, available at http://blclab.org/language-history-questionnaire/:

Step 1: The investigator or experimenter completes the sign-up process (by clicking on Investigator Sign-Up under LHQ 2.0 Functions), and receives a unique Experiment ID and an individualized URL associated with his or her experiment.

Step 2: The participants complete the LHQ online through the individualized URL, and data (along with participant numbers) are automatically and cumulatively saved.

Step 3: The investigator accesses the account management page (by clicking on Account Management) to download or delete data, or to update user and experiment information, through his or her unique Experiment ID and self-generated password (from Step 1).

References

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition, 15,* 616–629.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108,* 1–9.

Crinion, J. T., Green, D. W., Chung, R., Ali, N., Grogan, A., Price, G. R., Mechelli, A., & Price, C. J. (2009). Neuroanatomical markers of speaking Chinese. *Human Brain Mapping, 30,* 4108–4115.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition, 12,* 273–289.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition, 15,* 594–615.

Google Scholar (2013). Language history questionnaire: A web-based interface for bilingual research. http://scholar.google.com/scholar?hl=en&q=language+history+questionnaire&btnG=&as_sdt=1%2C39&as_sdtp= (retrieved September 8, 2013; cited by 73).

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics, 23,* 81–95.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods, 38,* 202–210.

Wilson, R., & Dewaele, J.-M. (2010). The use of web questionnaires in second language acquisition and bilingualism research. *Second Language Research, 26,* 103–123.