

## Open-source software to conduct online rating studies

GERRIT HIRSCHFELD, HEIDRUN BIEN, MEINOU DE VRIES,  
HEIDI LÜTTMANN, AND JOHANNES SCHWALL  
*Westfälische Wilhelms-Universität Münster, Münster, Germany*

Several software programs exist to assist researchers in setting up online questionnaires. Existing tools are of little help for delivering online rating studies, for which it is often desirable to collect data from participants for only a subset of a stimulus set. OR-Vis enables researchers to quickly set up online rating studies by supplying the set of items to be rated, the number of stimuli an individual participant responds to, the number of participants an item is shown to, and the rating questions. The software then generates and delivers unique questionnaires for each participant, while managing the data collection process. The present article describes OR-Vis, its installation process, and how to use it to gather data. OR-Vis is open-source software and can be downloaded from [www.orvis.uni-muenster.de](http://www.orvis.uni-muenster.de).

Online testing has become increasingly important in questionnaire research. This is due to its efficiency in collecting large amounts of high-quality data through automated scripts that prevent missing data, randomize questions, and record answer latencies (Birnbau, 2004; Gosling, Vazire, Srivastava, & John, 2004; Reips, 2002; Solomon, 2001). A major problem in doing online research is that programming online questionnaires can be very time consuming. A variety of methods have been developed, ranging from small, versatile scripts that require some programming skills (Birnbau, 2000; Göritz & Birnbau, 2005) to complete programming environments ([www.limesurvey.org](http://www.limesurvey.org)) that provide graphical user interfaces. Collecting ratings data for normalization (or norming) studies differs from collecting questionnaire data, in that it is frequently desirable to show only a subset of items to a given participant (e.g., to avoid fatigue), while ensuring that an equivalent number of ratings is collected for each item (McRae, Cree, Seidenberg, & McNorgan, 2005; Lahl, Göritz, Pietrowsky, & Rosenberg, 2009).

To assist researchers in conducting online norming studies, we developed OR-Vis, a tool for the online rating of visual stimuli. What distinguishes OR-Vis from other tools is that researchers can set up rating studies for which OR-Vis automatically generates multiple unique questionnaires, each of which may contain a different subset of rating items. Researchers use a graphical user interface to set up the rating study. With this interface, they specify the location of the stimuli on the Web server, the number of stimuli (NS) presented to each participant, the number of ratings (NR) collected for each stimulus, and the rating questions. As an illustration, in order to collect 10 rat-

ings ( $NR = 10$ ) for 500 stimuli, one can either construct five questionnaires of 100 stimuli ( $NS = 100$ ) and recruit 50 participants or one can include the whole set in one questionnaire ( $NS = 500$ ), so that only 10 participants are required. With traditional tools, the former alternative is more time consuming than the latter. In OR-Vis, both cases (and other alternatives) are equally simple. OR-Vis constructs a questionnaire for a particular participant by drawing NS times (without replacement) from those items that have been responded to less than NR times. Note that OR-Vis does not check whether questionnaires that conform to these constraints can be constructed. For example, if 5 stimuli are to be tested, NR may be set to 2 and NS may be set to 5, making it impossible to construct questionnaires of equal length. In these cases, and when the randomization procedure yields an uneven distribution, a slightly higher number of participants may be needed, the last of whom are presented with fewer items.

To capitalize on the strengths of Internet-based data collection, OR-Vis implements features that are commonly available in online research tools, such as missing-data checks, collection of answer latencies for each stimulus (in seconds), assignment of unique verification codes (to confirm participation), a WYSIWYG interface to edit test information, and internationalization (currently in German and English), so that the graphical user interface is shown in the language specified in the browser settings. OR-Vis takes a new approach to online testing by supporting researchers in setting up rating studies rather than questionnaires. Intended users are research groups and individuals who collect stimulus ratings on a regular basis. After installation, the OR-Vis software can be used to conduct an unlimited number of rating studies.

---

G. Hirschfeld, [gerrit.hirschfeld@uni-muenster.de](mailto:gerrit.hirschfeld@uni-muenster.de)

---





Figure 1. The participant interface.

## THE OR-VIS SOFTWARE

In the following sections, we describe the visible parts of the software (participant and researcher interface) before turning to a discussion of technical issues and availability.

### Participant Interface

After logging on to OR-Vis by providing a valid e-mail address, participants are presented a list of all active rating studies. Studies completed by the participant are no longer visible to her or him. When clicking on the name of a study, the participant is presented initial information about the test (e.g., example items, duration) and is asked to provide basic demographical data (currently set to age, gender, occupation, handedness, and language skills). These questions are hard-coded into the program and cannot be changed by using the graphical user interface.

Each questionnaire consists of a series of stimuli displayed one by one, along with a common set of questions that is used for the whole series. If answers are missing, the page is displayed again with the message that all questions need to be answered (see Figure 1). Note that the researcher can construct the question in such a way that it is possible to give a neutral answer, which would allow a participant to skip a page. Each participant is presented with the exact number of stimuli that the researcher had specified. After completion, the participant automatically receives a verification code (both via e-mail and on screen), which can be used to verify participation.

### Researcher Interface

All functions necessary to conduct a rating study are accessible through a password-protected graphical user interface. Here, the researcher can create rating studies, edit existing ones, export collected data, and access the list of the participants' verification codes.

To create a rating study, one needs to fill in the name, e-mail address, and some basic information about the study (title, number of stimuli per participant, number of presentations per stimulus; see Figure 2). Then, one selects the set of stimuli to be rated (i.e., the folder containing the media files on the server). After the basic settings are saved, questions can be added and edited. The available formats are text strings and Likert scales. When entering a text question, one additionally specifies the number of answers to be given. When a Likert scale question is entered, the scale must be defined with start and stop values (only integers are allowed; see Figure 3). Labels can be set for every value or for only a selection of values (e.g., start and stop values). This functionality can be used to add “unsure” or “don’t want to answer” response categories.

The initial study information page can be edited using a WYSIWYG interface (see Figure 4). It can be used to inform the participant about the time it takes to complete the test and to give one or more stimulus examples. The researcher can modify the subject and text of the e-mail sent to the participants, with the verification code being inserted automatically.

Rating studies can either be “closed” or “open,” and only open studies are available for participants. The researcher can close a study at any time; otherwise, it closes automatically as soon as each stimulus has received the specified number of responses.

Collected responses can be exported as comma-separated values (CSV) text files. Each line contains the answer(s) of one participant to one stimulus, together with the response time (in seconds). The answers are displayed in the order in which participants gave them. OR-Vis generates a LaTeX file that can be used to generate paper-and-pencil versions of the rating studies.

### Requirements

Similar to other programs (Görizt, & Birnbaum, 2005; www.limesurvey.com), OR-Vis requires a server running

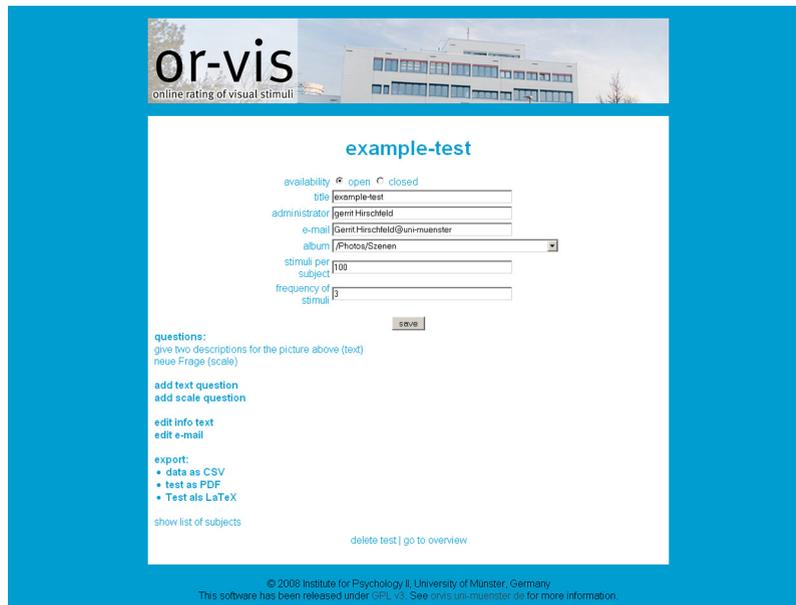


Figure 2. The researcher interface.

PHP and MySQL (e.g., [www.apachefriends.org](http://www.apachefriends.org)), because it is written using the CodeIgniter ([www.codeigniter.com](http://www.codeigniter.com)) Web application framework and uses MySQL for data storage. Stimulus files need to be located on a Web-accessible server, and we suggest using the Gallery2 Web album project ([www.gallery2.org](http://www.gallery2.org)) to manage the files, because it automatically stores stimuli in hierarchical folders that OR-Vis can access. OR-Vis and Gallery2 can run on the same server. OR-Vis accepts media files in several formats (JPEG, PNG, GIF, TIFF, BMP, MOV, and MPEG).

OR-Vis does not come with plug-ins and, thus, relies on the Web browser's ability to display multimedia files. At present, text stimuli (e.g., written words, sentences, ques-

tionnaire items) need to be converted into pictures (e.g., using PowerPoint or ImageMagick [[www.imagemagick.org](http://www.imagemagick.org)]). Compound stimuli, such as two pictures presented next to each other, need to be stored in one file (e.g., using ImageMagick). OR-Vis does not use pop-ups or similar techniques that are problematic with some browsers. The interface for study construction and delivery can thus be accessed from any Internet connection with modern browsers (tested on Firefox, Safari, Camino, and Internet Explorer).

### Installation and License

Basic documentation and all files necessary for installation can be downloaded from the project Web site at



Figure 3. Interface for editing Likert scales.



Figure 4. Interface for editing the initial test information.

www.orvis.uni-muenster.de. OR-Vis is released as a ZIP file. After the folders are unzipped and the files are uploaded to a Web server, two PHP files that parse the contents of individual folders must be copied to the folders containing the stimuli. Subsequently, the connection settings for the MySQL database and paths to the pictures must be entered into a configuration file.

A video tutorial explains all steps necessary to set up a rating study ([www.youtube.com/watch?v=tszZSLL4y8s](http://www.youtube.com/watch?v=tszZSLL4y8s)). The application is available under the GNU General Public License ([www.gnu.org/licenses/gpl.html](http://www.gnu.org/licenses/gpl.html)) and is free for personal, noncommercial use (as a research and teaching tool). Authors using this software are encouraged to cite this article. We would also like to encourage other researchers to adapt the software to their needs and implement further improvements.

### Example: Gathering Picture Norms for Psycholinguistic Experiments

We used OR-Vis to collect names and descriptions for 280 object pictures used in a priming experiment (Hirschfeld, Feldker, & Zwitserlood, 2010). First, the pictures were uploaded in a folder on a Web server using Gallery2. Then, the rating study was set up in OR-Vis. To reduce the workload for individual participants, the number of items per participant was set to 70 and each picture was shown to 20 participants. Thus, a total of 80 participants had to be recruited before the test was closed. For each picture, participants had to type the name of the object (e.g., “duck”) and a label that best described the current state of the object (e.g., “in the air”). Completing the test took about 10 min. We used MS Excel to analyze the resulting CSV file, from which we excluded responses with extremely long latencies (>3 min) from further analysis. The remaining answers were ordered by the pictures’ file names, and the different answers were counted. Setting up the test, advertising the questionnaire, and analyzing the results took about 2 days.

## DISCUSSION

Online testing has changed how researchers collect survey data, in part because of the many tools freely available to assist in data collection. What has been missing is a tool that supports gathering norming ratings across groups of individuals, something frequently done as part of larger norming studies (e.g., McRae et al., 2005). Recent studies reported high correlations between online ratings and traditional paper-and-pencil norming studies (Barenboim, Wurm, & Cano, 2010; Lahl et al., 2009; Paré & Cree, 2009). OR-Vis enables researchers to manipulate potentially interesting variables, such as the number of items shown to an individual participant. It is hoped that the availability of this tool will help advance our understanding of carrying out valid norming studies.

To sum up, we have described the advantages of online testing using OR-Vis, which offers an alternative to traditional paper-and-pencil questionnaires and to other tools available for online research. Once installed on a Web server, it takes about 10 min to set up a rating study that implements stimuli subgrouping, item randomization, and missing-data checks. We hope that it proves useful for individual users and assists researchers in implementing norming studies and collecting online ratings.

## AUTHOR NOTE

The development of this software was supported by the University of Münster research fund. We thank Kati Keuper for rigorously testing the software and Meinald Thielsch and two anonymous reviewers for comments on an earlier draft of the manuscript. G.H. is affiliated with the Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience at the University of Münster. Address correspondence to G. Hirschfeld, Psychological Institute II, Westfälische Wilhelms-Universität Münster, Fliednerstraße 21, 48149 Münster, Germany (e-mail: [gerrit.hirschfeld@uni-muenster.de](mailto:gerrit.hirschfeld@uni-muenster.de)).

## REFERENCES

BARENBOIM, D., WURM, L., & CANO, A. (2010). A comparison of stimu-

- lus ratings made online and in person: Gender and method effects. *Behavior Research Methods*, **42**, 273-285. doi:10.3758/BRM.42.1.273
- BIRNBAUM, M. H. (2000). SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet. *Behavior Research Methods, Instruments, & Computers*, **32**, 339-346.
- BIRNBAUM, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, **55**, 803-832.
- GÖRITZ, A. S., & BIRNBAUM, M. H. (2005). Generic HTML Form Processor: A versatile PHP script to save Web-collected data into a MySQL database. *Behavior Research Methods*, **37**, 703-710.
- GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S., & JOHN, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, **59**, 93-104. doi:10.1037/0003-066X.59.2.93
- HIRSCHFELD, G., FELDKER, K., & ZWITSERLOOD, P. (2010). Manche mögens bildlich—Individuelle Unterschiede im crossmodalen Satz–Bild Priming [Individual differences in crossmodal sentence–picture priming]. In C. Frings, A. Mecklinger, D. Wentura, & H. Zimmer (Eds.), *Beiträge zur 52. Tagung experimentell arbeitender Psychologen* (p. 112). Lengerich: Pabst.
- LAHL, O., GÖRITZ, A. S., PIETROWSKY, R., & ROSENBERG, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods*, **41**, 13-19.
- MCRAE, K., CREE, G. S., SEIDENBERG, M. S., & McNORGAN, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, **37**, 547-559.
- PARÉ, D. E., & CREE, G. S. (2009). Web-based image norming: How do object familiarity and visual complexity ratings compare when collected in-lab versus online? *Behavior Research Methods*, **41**, 699-704. doi:10.3758/BRM.41.3.699
- REIPS, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, **49**, 243-256. doi:10.1026//1618-3169.49.4.243
- SOLOMON, D. J. (2001). Conducting Web-based surveys. *Practical Assessment, Research, & Evaluation*, **7**, Art. 19. Retrieved January 22, 2010, from <http://PAREonline.net/getvn.asp?v=7&n=19>.

(Manuscript received August 7, 2009;  
revision accepted for publication November 14, 2009.)