Database Design for Dynamic Online Surveys

Roy P. Pargas[1], James C. Witte[2], Kowshik Jaganathan[1], John S. Davis[3]

Department of {[1]Computer Science, [2]Sociology, [3]Management}, Clemson University

Clemson, SC 29631

{pargas, jwitte, jkowshi, davis}@clemson.edu

Abstract

This paper discusses the architecture and implementation of dynamic web-based surveys with an emphasis on the recently completed *Survey2001* project. *Survey2001* was made available at the National Geographic website for several months starting October 2001 and could be taken in four different languages: English, German, Spanish and Italian. This paper discusses surveys in general, the advantages of web-based surveys, lays the background for *Survey2001*, describes the details of the database used and the manner in which transitions were conducted in this dynamic web-based survey. It also lists the results, including other surveys developed using the same database structure, and concludes with a look to the future.

1   Introduction

Web-based survey research represents the most recent addition to a growing repertoire of computer-assisted survey tools—including computer assisted telephone interview (CATI) and computer assisted personal interview (CAPI) systems—dating back to 1971 [2,3]. There are several advantages to web-based surveys. To begin with, as with all forms of computer-assisted survey research, a web-based instrument allows for complicated skip patterns that tailor the survey to the respondent and eliminate redundant or irrelevant questions [6]. For example, a person who is unemployed would not be asked questions about his/her work culture or how many emails he/she received at work. Moreover, as the idiosyncratic performances of individual interviewers are eliminated, customized surveys are implemented with a degree of accuracy and transparency unmatched by CATI or CAPI methods.

As a self-administered survey format, a Web-based survey potentially mitigates interviewer effects and permits a degree of anonymity not found in survey modes that depend on respondent-interviewer interaction [9]. At the same time, web-based surveys may include detailed help functions to guide and assist the respondent to a degree that is not possible with paper and pencil, self-administered formats [3,5].

Web-based surveys are also less expensive to maintain and make it easy to manipulate large volumes of data. Programming a Web-based survey can be costly, particularly if the instrument involves complex skip patterns or elaborate design elements. However, this cost is fixed. Unlike face-to-face or telephone surveys, increasing the sample size is not associated with added interviewer costs. As with other computer-assisted formats, a Web-based survey also eliminates the time and expense of data entry, which is performed by the respondent in the course of the survey [1]. Finally, a Web-based survey may draw on the multi-media capabilities of the Internet to yield an instrument that collects data in an engaging and interactive format.

Web surveys can be either static or dynamic. Dynamic surveys benefit both survey respondent and administrator through the use of interactive forms. With such forms, feedback can be displayed that is specifically tailored to the content of the responses supplied by the user, thereby giving the respondent instant feedback. By using dynamic surveys, it is possible to arrange it such that respondents are no longer merely giving information, but are also receiving information in return for their efforts. Under circumstances where respondents are made aware that they will benefit by participating, they are likely to exhibit increased motivation. If respondents know that the feedback they receive is about them, and based on the data that they provide, then they are likely to supply accurate and thoughtful responses [8].

This paper discusses the architecture and implementation of dynamic web-based surveys with an emphasis on the recently completed *Survey2001* project. *Survey2001* was made available at the National Geographic Web-site for several months starting October 2001 and could be taken in four different languages: English, German, Spanish and Italian. This paper describes the technical details of the implementation of this dynamic web based survey.

2   *Survey2001*: Background

As part of their coverage of the millenium, researchers at the National Geographic Society began collaborating with researchers at Northwestern University and a half dozen other universities to use a brand new research tool, an online survey, to tackle an age-old question, "How does where you live shape who you are?" Launched in October 1999, *Survey2000* was an ambitious experiment in web survey methodology and technology. *Survey2000* asked how often people moved, how strongly they felt about their communities, how extensively they used the Internet, and measured their global and local cultural preferences in terms of food, music, or authors. Magazine and television advertisements were used to reach potential respondents and more than 80,000 respondents in over

175 countries participated in the survey over a ten-week period [10].

Survey2000 was so successful that a follow-up project, *Survey2001,* was funded by the National Science Foundation (NSF). Once again with the cooperation of the National Geographic Society, *Survey2001* studied the impact information technology, particularly the Internet, has had on contemporary society. The survey focussed on impacts in three areas of society: community, culture and conservation. The overarching substantive goal of the survey was to use these three areas to describe how the Internet redefines our sense of geography, particularly the distinction between the global and the local.

The important methodological aim of *Survey2001* was to further explore sampling issues associated with web surveys. Like *Survey2000*, *Survey2001* relied primarily on convenience sampling techniques, with the majority of respondents initiating their survey participation through the National Geographic Magazine's homepage (www.nationalgeographic.com/ngm). However, in this instance the survey instrument was accessed by a number of distinct URLs. These URLs shared the same server and database, but each indicated a separate portal to the survey. Particular Internet sites (for example, the Sierra club) were assigned unique URLs, which they posted on their web sites with an encouragement to their visitors to participate in *Survey2001*. Thus, the characteristics of sample-subsets based on each portal may be compared, providing a unique opportunity to contrast convenience samples recruited from different online locations. Individual respondents were also given the opportunity to send email invitations to others to participate in *Survey2001*. A summary of the participants in *Survey2001* is provided in Table 1. Finally, for comparative purposes a telephone survey based on a subset of *Survey2001* questions was conducted with a randomly selected national sample.

In this paper we focus on the *Survey2001* database design, which allowed for a dynamic online questionnaire in an open Internet environment. The dynamic character of *Survey2001* is introduced in the opening screen, where respondents select to take the survey in one of four languages. Similarly, on the following screen respondents are steered to an adult or a youth version of the survey depending on their age. Next, respondents supply additional demographic information, including current primary residence, marital status and household composition. Subsequent questions asked about race and ethnicity, educational enrollment and attainment; and current employment status. This initial section was very important because many subsequent questions were based on the respondent's demographic data.

Questions about the clients' usage of the Internet and Internet tools were asked in the second section. The third category was composed of questions on environmental issues both global and local to where the respondent lived.

The user was then randomly presented one of four sections: Community, Reading and Politics, Science or Lifestyles. After this randomly selected section, each respondent was given an opportunity to quit the survey or continue with another module, randomly selected from the remaining survey modules. This process continued until the respondent either quit or completed all survey modules. Thus there were dynamic aspects to *Survey2001* within, between and across substantive survey modules. Indeed the inherently dynamic nature of the survey instrument defined the critical constraints in carrying out this survey project. Further, the solution to this problem, as outlined below, was intended to provide the basic database design and presentation software tools for a wide range of web based survey instruments regardless of content or complexity.

3 Problem Statement

None of the online survey authoring tools available at the time could satisfy all of the requirements of *Survey2001*. Specifically, *Survey2001* (a) was multi-lingual, (b) had one version for minors (under eighteen) and another for adults, (c) required moderately complex skip patterns (transitions) from one block of questions to others based on responses to multiple questions, (d) required images to accompany some of the questions for half of the respondents and no images for the other half, (e) had four mandatory and four optional categories of questions, (f) presented optional categories in random order, (g) allowed the respondent to enter comments at any time, (h) recorded what questions the client was viewing when the comment was made, (i) provided visual queues indicating progress made through the survey, and (j) worked with an external data database (of zip codes and the corresponding cities and states) in developing the text for one of the questions.

The decision was made to develop the database according to the needs of the survey, enter the data manually or semi-automatically, and develop presentation software to deploy the survey on the web. The total effort involved eight sociologists, three computer scientists, and six Clemson University and National Geographic personnel and took over six months to enter the data and develop the software to create *Survey2001*.

Section 4 explains how the requirements mentioned above were addressed by a combination of database design and presentation software.

4 Architecture

An overview of software and the database supporting *Survey2001* is shown in Figure 1. The database system used is MySQL (version 3.23.47). The survey is deployed by a collection of servlets, called the *Presentation Manager*, whose HTML output is served to the client by an Apache web server and Tomcat servlet engine (version 3.2.3).

The database tables are divided into five general groups, supporting (a) questions and answers, (b) question blocks, (c) responses and comments, (d) presentation format and layout, and (e) overall survey information. Details on each are provided in Sections 4.1−4.5. Section 4.6 provides a description of *OnQ*, an authoring tool that works with this database and allows a survey developer to more easily create an online survey.

4.1 Questions and Answers

Consider a survey that contains the item shown below.

Example 1

Rate each of the following foods				
	awful	so-so	good	great
hamburgers	o	o	o	o
soft tofu	o	o	o	o
pizza	o	o	o	o
fried chicken	o	o	o	o
bean sprouts	o	o	o	o
which have been suggested for the office picnic lunch.				

Questions and answers have multiple components. In the database, this example is represented by five questions whose `root` fields contain the words *"hamburgers", "soft tofu", "pizza", "fried chicken",* and *"bean sprouts"*. The *answer type* for each question is *"radio button"* and each question has four *answer options*: *"yuck", "so-so", "good",* and *"great"*. Only the first question has a *prefix* and a `suffix` field, which hold the fragments *"Rate each of the following foods"* and *"which have been suggested for the office picnic lunch,"* respectively. Finally, the questions are grouped into a single *radio button table* with answer options aligned in columns.

The same five questions can be presented in different ways. The *answer type* for one or more of the questions may be *drop-down menu*, instead. The questions may also be presented individually rather than in a table. Each question may have a different prefix and suffix, may have up to three media components (still images, audio, or video), and may be laid out in one of several ways. Or all five questions may be completely reorganized as one multiple-response, *check box* question in which the client clicks on all acceptable choices (*"hamburgers", etc.*).

4.2 Question Blocks and Transitions

Questions are grouped into categories (for example, some of the categories in *Survey2001* were *Demography, Internet, Conservation,* and *Culture*). Within a category, questions are grouped into question blocks.

The concept of a question block, and the ability of the survey author to specify that transitions, or transfer of control, conditionally take place from one block to another, provides the dynamism in online surveys. Figure 2 shows an example of four question blocks (QB1, QB4, QB25, and END) and transitions from one to another. In the example, control transfers from QB1 to QB4 if Boolean expression T1 evaluates TRUE. Transitions are arbitrary Boolean expressions whose variables are questions that the client has already seen and answered. For example, transition T1 from QB1 may read:

$$Q5=A10 \text{ and } Q7=A9$$

which is interpreted by the *Presentation Manager* as:

"From block QB1, move to block QB4 if the client answered *'Yes'* (i.e., answer A10) to question Q5 and *'No'* (i.e., answer A9) to question Q7."

If the expression is FALSE, then control will transfer from QB1 to another question block (not shown).

From QB4, control transfers to QB25 if T1 evaluates TRUE, or the survey ends if T2 evaluates TRUE. Otherwise control transfers back to QB1. Note that T1 from QB1 is *not* the same as T1 from QB4 and that the default transition from any node *always* evaluates TRUE. Boolean expressions are evaluated in increasing index order and the default expression is evaluated last.

This simple mechanism is a powerful tool. The author may now view the entire survey as a directed graph representing a deterministic finite state automaton where nodes are questions blocks and transitions are controlled by Boolean expressions. An example is shown in Figure 3. Two distinguished blocks, void of questions and called START and END, mark the start and end of the survey. Initially, START is designated as current; at any other time, the block whose questions the client is answering at the present time is considered *current*. The client's answers determine which Boolean expressions will evaluate TRUE. These, in turn, determine which block is selected next.

The two tables used to support transitioning are shown in Figure 4. Table `question_block` assigns an ID number to each block. Questions contained in the block are listed in field `block_seq`. Fields `rand_paramter` and `rand_type` deal with whether the questions in the block are to be randomized before being presented to the client. Table `block_sequence` provides 4-tuple records of the form:

(`current_block, expression, next_block, seqnum`)

enabling enable the *Presentation Manager* to decide which block of questions to present next. Expressions are evaluated in `seqnum` order and the `next_block` whose `expression` evaluates TRUE first is selected. The field `survey_type_id` allows the survey author to develop different graphs for different survey types. For example, a similar set of questions, possibly blocked and sequenced differently, may be used in a telephone survey. A telephone operator may click through the survey as he or she asks the respondent the questions. Or a survey taker in a public area, such as a shopping mall, may use a handheld digital device with a wireless connection to the

Internet may use an abbreviated version of a survey to assess public opinion.

4.3 Responses and Comments

Client responses are kept in two tables, one for fixed answer choices (Figure 5, Table user_response) and another for open-ended answers (Table user_response text). Both record the client identification (user_id) and the question number (q_id). In both, answers are date- and time-stamped (r_datetime), providing the survey developer with an answer trail for clients who click the back button on their browser and answer questions multiple times. They differ only in the answer field. Table user_response stores an index (a_id) into the table of answer choices. Table user_response_text records the client's text answer.

At any time during the survey, the client may submit a comment (Figure 5, Table comments). Comments are date- and time-stamped. The current question block number is stored in field q_block_id in order to provide to the survey developer the context within which the client made the comment. However, the number of questions within the question block may be large. To narrow the context down, field questions lists precisely those questions which the client has on his or her screen when the comment is submitted.

4.4 Presentation Format

The survey developer can provide the *Presentation Manager* with formatting requests for individual questions and for blocks of questions. For example, say that question block 1 of survey type 1 contains questions 1 through 10, but the developer wants to present to the client first three, then five, and finally two questions on three separate web pages. Entries of (3,1,1,"EP"), (8,1,1,"EP") in Table presentation_format (Figure 6) tells the *Presentation Manager* to end a web page ("EP" stands for "end page") with questions 3 and 8. The final two questions of the block, i.e., questions 9 and 10, will be presented on a separate page because the *Presentation Manager* will end a web page automatically when a question block is exhausted.

If, further, the survey developer wants the five questions on the second page to form a *radio button table*, presentation_format entries of (4,1,1,"RBT1"), (5,1,1,"RBT1"), …, (8,1,1,"RBT1 EP") inform the *Presentation Manager* that questions 4 through 8 are to be presented collectively in a *radio button table*.

The field q_presentation_format also enables the author to select from several question layouts, which determine the relative positions of questions and answers on a page.

4.5 Overall Survey Information

One table, called survey_instance, contains general information about the client. This information includes the language in which the client chose to take the survey, the time the client started, information about the client's computing environment (for example, browser name and version), the URL with which the client accessed the survey, whether the client is a referral by an earlier respondent, and if so, the user id of the referring client.

The URL from which the client linked to the survey is of particular interest to the survey analyst because this may provide insight into potential biases common among clients who frequent that URL. For example, it may be informative to group responses from clients who linked to a survey from a politically moderate website and contrast them to responses from clients who connected from liberal and conservative websites. This may assist the analyst in more accurately assessing true public opinion.

4.6 OnQ: An Authoring Tool

Entering survey information into the tables is difficult without additional software. An authoring tool, called *OnQ* [7], short for Online Questionnaire, is being developed by researchers at Clemson University. The tool provides a graphical user interface that helps a survey developer (a) enter questions and answers, (b) select answer types, (c) select media components, (d) create and sequence question blocks, (e) view question blocks and transitions in graph form, and (f) format questions and question blocks. The graphs in Figures 2 and 3 were automatically generated by *OnQ* from a sample survey.

5 Results

As noted above, an important aim of our efforts to produce a database design for dynamic online surveys is to produce the means to efficiently deploy web surveys without undermining the strengths of web survey technology. Flexibility and complexity in question skip patterns, along with respondent friendly presentation of questions and answers—where the means used to elicit responses may include photo images and sound and video files, as well as simple text—constitute the primary advantages of web surveys. Our database design intends to go beyond existing web survey tools by retaining these advantages while allowing surveys to be developed and deployed with few programming resources.

Our strategy has been to develop a system of survey implementation, where the database design is unaffected by question and answer content. Launching a new survey simply means creating a clone of the general database structure and entering new question, answer and transition rules into the appropriate tables. Similarly changes to the presentation manager are minimal; new graphic files may be referenced to give each survey a distinctive look and feel but the essential operation of the presentation manager is unchanged.

Using this approach developed as part of the *Survey2001*[1] project we have deployed additional web surveys, in each

---

[1] http://johnryan.clemson.edu/ngm/servlet/Page1

instance we have refined the basic tool as part of the process. For example, an online survey project for the Clemson University Office of Access and Equity will track a cohort of incoming freshman and survey them annually as to their attitudes and behavior regarding campus diversity. For the initial survey, *Diversity2001[2]*, a means to invite and register respondents was added to the database. Unlike *Survey2001* participation in this survey was not open to all visitors to particular web sites, but was limited to a sample of invited students who needed to enter their student identification numbers to participate. Moreover these identification numbers had to be retained to track individual students in subsequent years and had to be done in a way that protected the confidentiality of respondents and satisfied university requirements protecting human research participants.

Another web survey, *UnitedWay2002[3]*, a community survey of neighborhood issues and challenges for the United Way of Greenville County South Carolina, marked our first efforts to use the *OnQ* authoring tool to insert questions and answers into our generic database design. At the same time we identified and standardized the graphic elements that make up the basic building blocks used by the presentation manager in the initiation and introductory pages of a survey, as well as for the standard pages that contain survey questions.

The second year Clemson University diversity survey, *Diversity2002*, included a number of complicated skip patterns in a series of questions regarding university housing. Implementing this survey with *OnQ* demonstrated that the authoring tool could effectively enter such a survey pattern into the database, but also showed that changes to the authoring tool are needed to do so effectively. The most complicated skip patterns came as part of a series of questions regarding each student's housing situation. These questions needed to be repeated for each semester with slight modifications to the questions to remind the respondent as to which semester was currently under consideration. Cyclical sequences of questions are not uncommon in survey research (e.g., questions about each job a respondent has ever held, or each consumer product ever tried) and our aim is to improve the next version of *OnQ* so that a set of questions, along with its related answers and skip pattern transition rules may be cloned and inserted into the database. This would then automates the heavy lifting, leaving it up to the survey author to simply edit the basic sequence of questions to indicate each particular cycle of questions.

Finally, *IE2002[4]*, a web survey of industrial engineers regarding issues in undergraduate engineering curriculum

offered a new challenge. This survey included a series of radio button tables that involved rather complicated instructions to the respondents. Incorporating these instructions could have meant significant modifications and customization of the presentation manager were it not for the database design we have employed. One of the question types permitted in the design is called "splash and continue," which we have typically used as a separate screen to mark transitions from one set of questions to the next and to orient respondents to a change in topic. These "questions" are distinctive in that they do not allow for an answer. Rather they serve as clarification or instructions for subsequent questions. Simply by placing a question of this type, one which contained the instructional text for the following set of questions, on the same page as the radio button items we were able to provide the appropriate instructions without customizing the presentation manager.

6 Summary and Future Work

The database described in this paper has served us well over the past one-and-a-half years, and continues to do so. New surveys with new content and completely different skip patterns are now developed quite easily. Each survey, however, suggests new ideas on how to make *OnQ* more powerful and our goal is to continue improving the database, modifying tables as new needs arise. For the immediate future, changes to the database will: (a) allow the author to specify type fonts and sizes of prefixes, roots, and suffixes, (b) enable the author to specify different backgrounds for different categories, (c) add a new question type, the email invitation question, that causes the *Presentation Manager* to send email invitations (to take the survey) to addresses specified by the client, and (d) print paper equivalents (with appropriate skip instructions) of an online survey.

7 Acknowledgements

References

1. Baker, R. P., Bradburn, N. M., and Johnson, R. (1995). "Computer-assisted personal interviewing: An experimental evaluation of data quality and survey costs" *Journal of Official Statistics,* v.11 p 415-34.

2. Couper, M. and W. Nicholls II. (1998) "The History and Development of Computer Assisted Survey Information", *Computer Assisted Survey Information Collection* New York: Wiley.

3. Dillman, D. (1999) *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.

---

[2] http://johnryan.clemson.edu/diversity2002/servlet/Page1

[3] http://johnryan.clemson.edu/UnitedWay/servlet/Page1

[4] http://johnryan.clemson.edu/IE2002/servlet/Page1

4. InfoPoll. http://www.infopoll.com/products/survey
5. Jenkins. C.R., and Dillman, D.A. (1997), "Towards a Theory of Self-administered Questionnaire Design." Lyberg, L, Biemer, P., Collins, M., DeLeeuw, E., Dippo, C., Schwarz. N., and Trewin, D. (eds.), *Survey Measurement and Process Quality*. New York: John Wiley.
6. Nicholls. W.L. II, Baker, R.P., and Martin, J. (1997) The Effect of New Data Collection Technologies on Survey Data Quality," Lyberg, L, Biemer, P., Collins, M., DeLeeuw, E., Dippo, C., Schwarz. N., and Trewin, D. (eds.), *Survey Measurement and Process Quality*. New York: John Wiley.
7. Pargas, R.P., Witte. J.C., Brand, L., Hochrine, C., Staton, M. *OnQ: An Authoring Tool for Dynamic Online Surveys*, Submitted to ITCC2003, Las Vegas, NV.
8. Schmidt, W. C. (1997). Worldwide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments & Computers*, 29, p 274-279.
9. Witte, J.C., "Survey Methods - Large Sample Design & Studying Life Online", *The Internet & American Life*, Philip E.N. Howard and Steven Jones (eds.), [forthcoming] Thousand Oaks, CA: Sage Publications.
10. Witte, J.C., L. Amoroso and P.E.N. Howard, "Method and Representation in Internet-Based Survey Tools: Mobility, Community, and Cultural Identity in Survey2000," *Social Science Computer Review*, Summer 2000 Special Issue, 18(2).
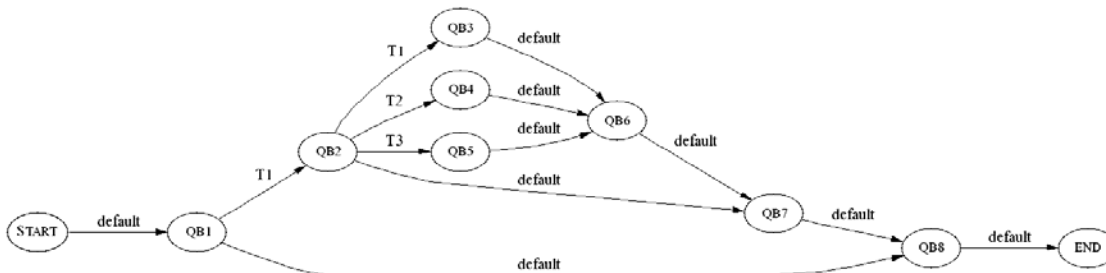
Figure 1. Overview of *Survey2001*: development and deployment.
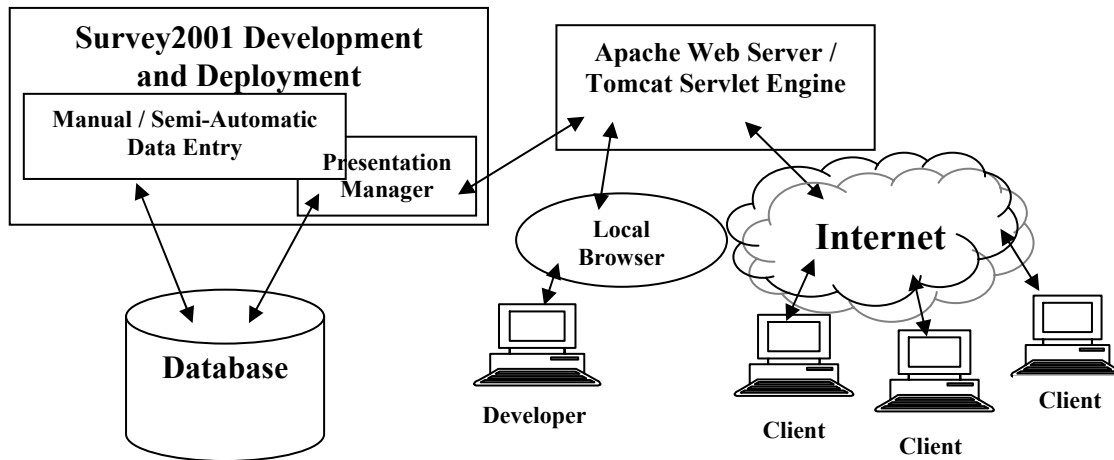
Figure 2. Question blocks and transitions.
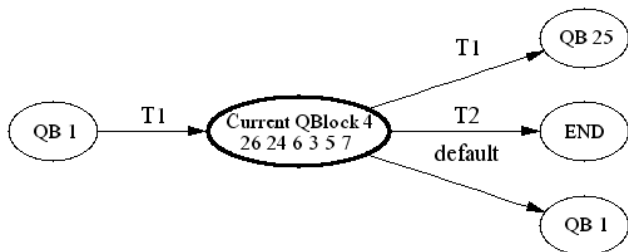
Figure 3. View of an entire survey.

Figure 4. Tables supporting question blocking and transitioning.

```
table question_block;
+---------------+------------+------+-----+---------+
| Field         | Type       | Null | Key | Default |
+---------------+------------+------+-----+---------+
| q_block_id    | bigint(20) |      | PRI | 0       |
| q_block_seq   | text       | YES  |     | NULL    |
| rand_type     | bigint(20) | YES  |     | NULL    |
| rand_parameter| bigint(20) | YES  |     | NULL    |
+---------------+------------+------+-----+---------+


table block_sequence;
+-------------------+------------+------+-----+---------+
| Field             | Type       | Null | Key | Default |
+-------------------+------------+------+-----+---------+
| survey_type_id    | bigint(20) |      |     | 0       |
| current_q_block_id| bigint(20) |      | PRI | 0       |
| boolean_expression| text       | YES  |     | NULL    |
| next_q_block_id   | bigint(20) |      |     | 0       |
| sequence_number   | bigint(20) |      | PRI | 0       |
+-------------------+------------+------+-----+---------+
```

Figure 5. Client response tables.

```
Table user_response;
+------------+------------+------+-----+---------------------+
| Field      | Type       | Null | Key | Default             |
+------------+------------+------+-----+---------------------+
| user_id    | bigint(20) |      | PRI | 0                   |
| q_id       | bigint(20) |      | PRI | 0                   |
| a_id       | bigint(20) |      | PRI | 0                   |
| r_datetime | datetime   |      | PRI | 0000-00-00 00:00:00 |
+------------+------------+------+-----+---------------------+


Table user_response_text;
+----------------+------------+------+-----+---------------------+
| Field          | Type       | Null | Key | Default             |
+----------------+------------+------+-----+---------------------+
| user_id        | bigint(20) |      | PRI | 0                   |
| q_id           | bigint(20) |      | PRI | 0                   |
| r_text         | text       | YES  |     | NULL                |
| r_text_datetime| datetime   |      | PRI | 0000-00-00 00:00:00 |
+----------------+------------+------+-----+---------------------+


Table comments;
+------------+------------+------+-----+---------------------+
| Field      | Type       | Null | Key | Default             |
+------------+------------+------+-----+---------------------+
| user_id    | bigint(20) |      | MUL | 0                   |
| q_block_id | bigint(20) |      |     | 0                   |
| questions  | text       | YES  |     | NULL                |
| comment    | text       | YES  |     | NULL                |
| c_datetime | datetime   |      |     | 0000-00-00 00:00:00 |
+------------+------------+------+-----+---------------------+
```

Figure 6. Question and block formatting information.

```
Table presentation_format;
+----------------------+------------+------+-----+---------+
| Field                | Type       | Null | Key | Default |
+----------------------+------------+------+-----+---------+
| q_id                 | bigint(20) |      | PRI | 0       |
| q_block_id           | bigint(20) |      | PRI | 0       |
| survey_type_id       | bigint(20) |      | PRI | 0       |
| q_presentation_format| text       | YES  |     | NULL    |
+----------------------+------------+------+-----+---------+
```

Table 1. Origin and language of *Survey2001* respondents

	Total	NGS site	Other site	Email referral
Surveys: initiated	23,192	14,064	8,569	559
Survey language				
English	75.0%	77.7%	69.7%	86.8%
German	6.9%	6.8%	7.1%	5.4%
Italian	7.2%	6.3%	8.8%	3.0%
Spanish	11.0%	9.2%	14.1%	4.8%
Surveys: demographics complete adults	12,361	7,583	4,470	408
Surveys: complete	7,767	4,831	2,669	267
Survey language				
English	85.1%	85.1%	84.8%	88.8%
German	6.6%	7.3%	5.5%	5.6%
Italian	3.5%	3.6%	3.6%	3.2%
Spanish	4.7%	4.0%	6.1%	3.4%