

Use of a Web-Based Sample Replacement Software Application for Data Collectors in the Field

Andrey Vinokurov¹, Daniel Geller¹, Pedro J. Saavedra¹, and Tamara L. Martin²
 ORC Macro¹, U.S. Department of State²

ABSTRACT

ORC Macro International recently conducted the evaluation of the International Visitor Leadership Program (IVLP) for the U.S. Department of State. This long-running program identifies, recruits, and selects emerging foreign leaders and promising professionals in a wide variety of fields and funds short-term visits for them to the United States, where they are exposed to a broad program of professional training, career development, American cultural experiences, and interpersonal contacts with Americans. The data collection was subcontracted to the Institute for Comparative Social Research (CESSI based on the Russian acronym), a major research company in Russia with a field network of over 1000 professional interviewers in all parts of Russia and other countries of the Commonwealth of Independent States (CIS). The data collection involved difficult to locate and recruit alumni of the IVLP from four study countries: Russia, Ukraine, Kazakhstan, and Georgia. These countries have transportation, communication, and economic impediments to data collection process, which was further complicated by the need to adhere to the sampling procedures. The sample had to be representative with respect to the year of program participation, and therefore ORC Macro could not simply provide the list of potential respondents to CESSI and let them collect any available or easily accessible respondents. ORC Macro had to oversee the data collection process to ensure that respondents were interviewed in a certain order according to the sampling procedures. To facilitate the supervision of the data collection, ORC Macro developed a web-based sampling system accessible to the data collection coordinator in Russia. This sampling program provided the data collection coordinator in Moscow with a simple tool to select the appropriate respondents and their replacements, without having to be very knowledgeable about the sampling procedures or having to contact ORC Macro regarding each replacement. In addition, it allowed ORC Macro to continuously monitor the data collection progress and resolve any emerging issues without having to worry about the communication delays. The web-based sampling program was easy to develop and provided a reliable and simple tool for selecting respondents to be interviewed. It solved numerous logistical problems associated with collecting data simultaneously in multiple countries, while preserving the integrity of the sampling design.

Overview of the Program Evaluation

ORC Macro International recently conducted the evaluation of the International Visitor Leadership Program (IVLP) for the U.S. Department of State. This long-running program identifies, recruits, and selects emerging foreign leaders and promising professionals in a wide variety of fields and funds short-term visits for them to the United States, where they are exposed to a broad program of professional training, career development, American cultural experiences, and interpersonal contacts with ordinary Americans. These one-to-three week visits are intended to both improve the professional practices and career positions of participants, as well as allow opportunities for them to learn first-hand how democratic institutions and processes, such as free-market economies, rule of law, and other values of Western democracy and American society are manifested in professional and daily life.

The purpose of the IVLP outcome assessment was to conduct an evaluation study with alumni from four countries of the Commonwealth of Independent States (CIS) (Russia, Ukraine, Kazakhstan, and Georgia) who had participated in IVLP visits to the United States between 1996 and 2001. The program evaluation's overall objectives were: 1) to determine the immediate and long-term outcomes of the IVLP experiences on the participants, their institutions, and their home countries; 2) to assess the levels of participation in and value of the IVLP for alumni and their affiliated organizations; and 3) to document the alumni's demographic and professional characteristics. Specifically, the project involved assessing alumni's views of the program; measuring its impact on their views of America, its democracy, government institutions, culture, society, and people; determining the effects of the program on participants' professional development, and examining how program alumni disseminate acquired professional and cultural information to people and institutions in their native country.

Issues in Overseeing Data Collection in Other Countries

The fact that the IVLP employs so many different modalities (lectures, governmental visits, peer relationships, classes, just to name a few), combines them in many novel ways, and implements them with

visitors representing widely varying disciplines, cultures, languages, religions, economies, and political systems presents conceptual and logistic challenges in terms of designing and implementing data collection methods. Furthermore, the study included countries that are transitioning from formerly state-controlled societies and are currently enduring many resulting dislocations, which meant working with a difficult to locate and, perhaps, reluctant respondent population. The data collection was subcontracted to the Institute for Comparative Social Research (CESSI based on the Russian acronym), a major research company in Russia with a field network of over 1,000 professional interviewers in all parts of Russia and other countries of CIS. Based in Moscow, CESSI has offices in St. Petersburg (Russia), Astana (Kazakhstan), Kiev (Ukraine) and over 100 regional representatives in all countries of the CIS. CESSI also has an established cooperative relationship with the Institute for Polling and Marketing in Tbilisi, Georgia.

In addition to the issues of collecting data from difficult to locate and recruit respondents from four countries that have numerous transportation, communication, and economic problems impeding the data collection process, the logistics of the data collection was further complicated by the fact that it had to follow strict sampling procedures to ensure representativeness of the final sample. The data was collected via the face-to-face, questionnaire-driven, structured interviews with 90 percent of the sample, as well as, the open-ended, in-depth interviews with the remaining 10 percent of the sample. The overall sample was selected from the lists of IVLP visitors who participated in the program from 1996 through 2001. The State Department obtained the lists from its exchange visitors' database. After examining the database for completeness, ORC Macro staff determined that the only available criterion for sampling was IVLP program year, sampled separately for each country. From the overall list of 3,308 potential respondents ORC Macro had to collect data from 827 people. Due to the anticipated difficulties in locating and recruiting respondents we quadrupled the number of cases included in the final sample. Thus, we had three replacement samples to supplement the respondents who could not be located or refused to be interviewed (See Table 1). If a respondent could not be contacted or refused to participate, the respondent from the second sample of the same country and with the same year of the IVLP visit had to be used as a replacement. If the respondent from the second sample was not interviewed, the assigned respondent from the third sample had to be substituted. If that did not work, the fourth sample had to be used. Finally, if no more replacements were available, then replacements available from the completed interview could be used,

provided that the years of the program participation were not more than three years apart. This provided a rigid structure for the sampling order of the data collection process.

Table 1. Replacement Samples by Country

COUNTRY	SAMPLE				TOTAL
	1	2	3	4	
Russia	514	514	514	514	2056
Ukraine	150	150	150	150	600
Georgia	75	75	75	75	300
Kazakhstan	88	88	88	88	352
TOTAL	827	827	827	827	3308

Given the need to obtain a representative sample with respect to the year of program participation, ORC Macro could not simply provide the list of potential respondents to CESSI and let them interview any available respondents. Instead we had to oversee the data collection process to ensure that sampling and replacement parameters were followed. This created a myriad of logistical problems. First, replacements would not be necessarily located in the same city or region and given the transportation difficulties it required highly coordinated efforts in assigning interviewers. Second, interviewers would have to be knowledgeable about the sampling selection procedures and would have to strictly follow the assignment protocol. Finally, communication delays due to the time zones precluded us from providing continuous and timely feedback and guidance regarding the sampling selection order. As a result, we decided to develop a web-based sampling system that could be accessed by CESSI's data collection coordinator in Moscow. This sampling program provided the data collection coordinator in Moscow with a simple tool of selecting the appropriate respondents and their replacements, without having to be very knowledgeable about the sampling procedures or having to contact ORC Macro regarding each replacement. In addition, it allowed us to continuously monitor the data collection progress and resolve any emerging issues at the time convenient for us.

Description of the Web-Based Sampling Program

We've developed two Web interfaces: one to be used by the CESSI data collection coordinator in Moscow and the second to be used by the staff at the ORC Macro office in the United States. The Web page for the data collection coordinator had minimal functionality and displayed the number of cases equal to the sample size for each country without displaying all available replacements (See Exhibit 1 attached at the end of the paper). This was done to provide the parsimonious and easy to use sampling web interface.

Conversely, the Web page for the ORC Macro staff displayed all of the cases from all four samples and had full functionality for manually adjusting the interview status of the respondents by clicking on their names (See Exhibit 2 attached at the end of the paper). The Web page for the ORC Macro staff was developed to monitor the data collection and to resolve any mishaps related to the use of the sampling program or any issues emerging during the data collection process.

The first page of the data collection coordinator (operator) website is presented in Exhibit 1 at the end of the paper. The sampling Web page indicates on the top of the page the sample country, which can be changed by using the pull down menu. Once the desired country for the sample is selected, the Web page displays the initial sample for the country. The initial sample consists of the total number of respondents to be interviewed. However, there are three replacement samples that are used when a respondent in the initial sample cannot be located or refuses to participate in the study. These replacement samples are not displayed, but are linked for each respondent via the algorithm of the sampling program.

The first three columns of the sampling table are the action buttons to be used by the data collection coordinator to change the status of a respondent (See Table 2 for examples of coding a respondent’s status):

C -- Completed interview: activated when an interview with the respondent is completed.

H -- Hard refusal: clicked if the respondent specifically indicated that he/she refuses to participate in the study.

S -- Soft refusal: activated if the respondent could not be located or he/she could not participate in the study right away, but might be available for an interview later.

Table 2. Examples of Coding Respondents’ Status

STATUS	EXAMPLES
Hard Refusal	Never participated in the IVLP Respondent deceased Refused to participate in the study
Soft Refusal	<u>Bad contact information</u> – busy phone number, no answer, answering machine, non working/disconnected phone number, or no such person/not named respondent <u>Break off</u> – uncompleted interview <u>Gatekeeper Refusal</u> – other people in the household refuse to get the respondent on the phone <u>Initial refusal</u> – respondent is busy, asked to callback, or not available during study period <u>No contact information</u> in the database, but might be updated later

The next column of the sampling table (Exhibit 1 at the end of the paper), after the action buttons, is the SLOT number. This column stays constant and indicates the sampling slot to be filled. The total number of sampling slots on the Web page is equal to the total number of cases for the given country’s sample. All four samples are linked to the slot numbers and the algorithm of the sampling program substitutes a respondent once the particular respondent could not be located (S) or refuses to participate (H). Thus, after either the S or H action button is clicked, the Web page refreshes and the program selects and displays the name and the ID number of another respondent from the four replacement samples to be substituted into the slot. Conversely, if the respondent in the slot has been interviewed and the C button is pressed, then the page refreshes and indicates the status FILLED instead of the name of the respondent. In this case, all replacements that are still available for the filled slot are being released by the algorithm to be used as substitutes for other slots, if needed. By the end of the project, all of the slots on the Web page display “FILLED”.

The sampling table also displays the ID numbers, NAME of the respondents (blacked out to protect confidentiality), TYPE of the interview to be done with the respondent (structured or in-depth), STATUS, ROW and SAMPLE of the respondent.

ID – composed of the Country (in descending order by size; Russia=1, Ukraine=2, Kazakhstan=3, Georgia=4), Sample (1, 2, 3, or 4), and Row (from 001 to the maximum number of cases to be interviewed for the country).

SAMPLE – indicates the sample number that the particular respondent is selected from.

ROW – is the respondent’s initial slot number. It indicates whether the replacement for the slot is selected from the slot’s original three replacements (in this case row number equals to the slot number) or the replacement is selected from another sampling slot that has been filled and has available replacements to share (in this case row number is not equal to the slot number).

STATUS – the operator’s Web page displayed only two statuses of the respondents: completed interview and available case to be interviewed; while the administrator’s Web page displayed four additional statuses: primary replacement (to be selected from the slot’s original three replacements), secondary replacement (to be selected from another sampling slot that has been filled), soft refusal, and hard refusal.

The web-based sampling program also included a contingency—one that we never needed to use—for a situation when there would not be any available replacements for a given SLOT number. In this case,

the word WAITING would have been appeared instead of the name of the respondent. This indicates that the program is waiting for a completed interview, so that available replacements for the completed interview could be released and used for other slots. If WAITING appears for any of the slots, it can be filled only after another interview is completed. There are two steps to obtain the replacement. First, the operator must click on the C button for the completed case, which releases its replacements. Second, the operator has to locate the WAITING case and click on either the H or S button (it does not matter which) so that the program can substitute a released replacement for the WAITING case. Substituting released replacements for the WAITING case was done manually to allow control over which WAITING case to be replaced (if there are more than one), as well as to avoid having to search through the sampling table to find the SLOT that was replaced. However, this issue was never encountered by the operator, because there were enough replacements to be substituted.

The Web page also provided an undo button on the top of the page that reloaded the previous version of the Web page to cancel an action done by mistake. However, the undo button only replaced one previous action. If more than one mistake was made, the status of the case had to be changed manually using the administrator's Web page accessible only to ORC Macro staff (See Exhibit 3). Having undo for only one action prevented the operator from compounding mishaps, as well as allowing ORC Macro staff a greater control over any issues that might emerge during the data collection process.

Description of the Procedures for Operating the Sampling Web page

- 1) The Web page was used only by CESSI's data collection coordinator in Moscow, who had the username and password. It was **not** designed to be used by the interviewers. The Web page was used only for the selection of the replacements. All of the contact information was located in a separate Excel database, sorted by IDs. As a result, the data collection coordinator had to rely on two databases: an Excel spreadsheet to maintain and update contact information and the sampling Web page for sample replacement management.
- 2) Once the data collection was on the way, the data collection coordinator obtained continuous feedback from the field interviewers on the status of their efforts to contact respondents (interviewers provided the updated status of the respondents by using the Names and ID numbers). The updated status was coded by the

data collection coordinator into the three categories of Completed, Hard refusals, or Soft refusals.

- 4) Using the ID number and the Name provided by the interviewer, the data collection coordinator located the respondent on the operator's Web page and clicked on the appropriate action button to update the status of the case.
- 5) After the status of the respondent was updated, the Web page displayed either the word FILLED (if the status of the respondent was "completed") or displayed the name and the ID number of the next available replacement.
- 6) After obtaining the name and ID of the next available replacement, the data collection coordinator provided the interviewer with the ID and name of the replacement respondent to be used. This still left the data collection coordinator with logistic issues, as the replacement may not have been living in the same city or region, so it might have been necessary to assign the replacement to another interviewer for completion.

Suggestions for Improvement

The web-based sampling program was easy to develop and provided a reliable and simple tool for selecting respondents to be interviewed. It solved numerous logistical problems associated with collecting data simultaneously in various locations while preserving the integrity of the sampling design. The sampling program was especially instrumental for the IVLP project as it required data collection in multiple countries with little possibility of direct supervision due to communication delays. ORC Macro received a very good feedback from the data collection coordinator who lamented only that this program was not available for her other projects. Furthermore, the programmer, who actually created the program, did not report having difficulties in following the sampling specifications provided by the statistician and was able to complete the work without having to go through numerous revisions. As a result, the program saved ORC Macro budgetary funds that would have been spent on supervising the data collectors' adherence to the sampling design.

In spite of the positive performance and feedback, the web-based sampling program could be further improved and fine-tuned. First, the display design of the Web page could be improved to provide a more user-friendly interface. For example, the status could be changed to the pull-down menu that would not be limited to three status options, but would include all possible response dispositions outlined in Table 2. Furthermore, the interface could be improved by integrating the contact information into the web-based sampling program. This would allow the data

collection coordinator to sample the appropriate respondent, as well as to have the contact information for the respondent readily available, thus removing the need to match the selected respondent to his/her contact information in a separate database. Another major improvement involves adjusting the sampling algorithm to accommodate more complex sampling designs. Although the sample for our study was drawn only with respect to the country and year of program participation, the web-based sampling program could be altered for use with other projects that require numerous sampling parameters. Currently ORC Macro is examining the applicability of the web-based sampling program for other projects.

Exhibit 1. First Page of the Sampling Web-Based System for the Use by the Operator

UNDO											
SAMPLES for KAZAKHSTAN											
Completed	Hard Refusal	Soft Refusal	S L O T	ID	FIRST NAME	LAST NAME	TYPE	STATUS	R O W	S A M P L E	
C	H	S	1	32002	FILLED		Structured	completed	2	2	
C	H	S	2	31002	FILLED		Structured	completed	2	1	
C	H	S	3	32003	FILLED		Structured	completed	3	2	
C	H	S	4	32004	FILLED		Structured	completed	4	2	
C	H	S	5	31005	FILLED		In Depth	completed	5	1	
C	H	S	6	32006	FILLED		Structured	completed	6	2	
C	H	S	7	31007	FILLED		Structured	completed	7	1	
C	H	S	8	33010	FILLED		Structured	completed	10	3	
C	H	S	9	33009	FILLED		Structured	completed	9	3	
C	H	S	10	31010	FILLED		Structured	completed	10	1	
C	H	S	11	32011			Structured	available	11	2	
C	H	S	12	34013			Structured	available	13	4	
C	H	S	13	32013			Structured	available	13	2	
C	H	S	14	32014			Structured	available	14	2	
C	H	S	15	31015			In Depth	available	15	1	
C	H	S	16	31016			Structured	available	16	1	
C	H	S	17	31017			Structured	available	17	1	
C	H	S	18	31018			Structured	available	18	1	
C	H	S	19	31019			Structured	available	19	1	
C	H	S	20	31020			Structured	available	20	1	
C	H	S	21	31021			Structured	available	21	1	
C	H	S	22	31022			Structured	available	22	1	
C	H	S	23	32023			Structured	available	23	2	
C	H	S	24	32024			Structured	available	24	2	
C	H	S	25	31025			In Depth	available	25	1	
C	H	S	26	31026			Structured	available	26	1	

Note 1: The exhibit presents only the first 26 cases of the sampling Web page of 88

Note 2: To ensure confidentiality respondents' names were blacked out

Exhibit 2. First Page of the Sampling Web-Based System for the Use by the Administrator

<u>UNDO</u>										
SAMPLES for KAZAKHSTAN										
Completed	Hard Refusal	Soft Refusal	SLOT	ID	FIRST NAME	LAST NAME	TYPE	STATUS	ROW	SAMPLE
C	H	S	1	34001			Structured	secondary replacement	1	4
C	H	S	1	31001			Structured	soft refusal	1	1
C	H	S	1	32001			Structured	secondary replacement	1	2
C	H	S	1	33001			Structured	secondary replacement	1	3
C	H	S	1	32002	FILLED		Structured	completed	2	2
C	H	S	2	33002			Structured	secondary replacement	2	3
C	H	S	2	34002			Structured	secondary replacement	2	4
C	H	S	2	31002	FILLED		Structured	completed	2	1
C	H	S	3	31003			Structured	soft refusal	3	1
C	H	S	3	32003	FILLED		Structured	completed	3	2
C	H	S	3	33003			Structured	secondary replacement	3	3
C	H	S	3	34003			Structured	secondary replacement	3	4
C	H	S	4	31004			Structured	soft refusal	4	1
C	H	S	4	32004	FILLED		Structured	completed	4	2
C	H	S	4	33004			Structured	secondary replacement	4	3
C	H	S	4	34004			Structured	secondary replacement	4	4
C	H	S	5	31005	FILLED		In Depth	completed	5	1
C	H	S	5	32005			In Depth	secondary replacement	5	2
C	H	S	5	33005			In Depth	secondary replacement	5	3
C	H	S	5	34005			In Depth	secondary replacement	5	4
C	H	S	6	33006			Structured	secondary replacement	6	3
C	H	S	6	31006			Structured	hard refusal	6	1
C	H	S	6	32006	FILLED		Structured	completed	6	2
C	H	S	6	34006			Structured	secondary replacement	6	4
C	H	S	7	31007	FILLED		Structured	completed	7	1
C	H	S	7	32007			Structured	secondary replacement	7	2
C	H	S	7	33007			Structured	secondary replacement	7	3
C	H	S	7	34007			Structured	secondary replacement	7	4

Note 1: The exhibit presents only the first 26 cases of the sampling Web page of 352

Note 2: To ensure confidentiality respondents' names were blacked out