Automated Surveys: Lowering the respondent's burden

Richard Kist MSc., Igor Franken MSc.
National Aerospace Laboratory (NLR)
Amsterdam, The Netherlands
Richard.Kist@nlr.nl, Igor.Franken@nlr.nl

ABSTRACT

Measuring effectiveness and the operational utility of new techniques, technologies and training through feedback of the Warfighter is critical. Paper questionnaires and interviews have merits. However, offering a relevant, compact questionnaire after each mission to test and exercise participants with different roles and backgrounds can be made more efficient with an automated survey. This paper discusses the challenges faced in survey data collection in an operationally realistic environment, and lessons learned during years of survey data collection. During the Bold Quest cycle, a survey effort was set up using such an automated tool (Questionnaire for Utility Evaluation and Survey Tool, QUEST). The concept of employment for automated surveys described here addresses many of the concerns associated with web-based surveying techniques (Sills & Song, 2002). Use of clear and concise questions and other measures lowering the burden for the respondents yield the best results from a survey effort. An automated tool ideally should work locally on any laptop or handheld device, as well as in networked conditions. It should be tailored to military environments, offering questionnaires for all phases (pre/post exercise, daily post mission). By tooling the questions to reduce burden, analyzing the language used, and taking steps to assure the relevance of the questions to each participants, more efficient data collection can occur – the kind of data collection which provides 100,000s of survey responses helping to determine the effectiveness of new military developments.

ABOUT THE AUTHORS

Richard Kist received his Master of Science degree in Aerospace Engineering at the Technical University of Delft. He started his career in an IT company. In 1999 he joined the National Aerospace Laboratory (NLR) in Amsterdam in the department of Defense Systems, supporting the Royal Netherlands Air Force. From 2007 Richard was involved together with colleague Igor Franken in programs involving combat identification and Digitally Aided Close Air Support (DACAS). He was lead analyst of the NLD analyst team at exercise Bold Quest in 2007, 2009, 2011 and 2013. Richard and Igor developed the survey tool QUEST for use in this series, and played an important role in the successful subjective data collection during Bold Quest events.

Igor Franken received his Master of Science degree in Electrical Engineering at the Technical University of Delft. Shortly after graduation Igor's professional career started in 2006 at the National Aerospace Laboratory (NLR) in Amsterdam in the department of Defense Systems. From 2007 Igor was involved together with colleague Richard Kist in programs involving combat identification and DACAS. He was an active member of the Dutch team at exercise Bold Quest in 2007, 2009, 2011 and 2013, focusing on aircrew. Igor and Richard developed the survey tool QUEST for use in this series, and played an important role in the successful subjective data collection during Bold Quest events.

Automated Surveys: Lowering the respondent's burden

Richard Kist MSc., Igor Franken MSc.
National Aerospace Laboratory (NLR)
Amsterdam, The Netherlands
Richard.Kist@nlr.nl, Igor.Franken@nlr.nl

INTRODUCTION

Military operations research can involve introduction of new technology and tactics, techniques and procedures, or improvement of existing ones. For new technologies, techniques and procedures, it is important to assess the utility of these at an early stage. Part of this analysis is evaluation of technical data from systems, location information and data exchanges. This is objective data, which captures how well the systems perform from a technical point of view. However, to determine the true utility in military operations, it is also critical to pair this data with feedback data from the personnel operating the new technologies, techniques or training methods.

During the operational demonstration and test cycle Bold Quest (BQ), the participating researchers realised that there was a need for a more advanced type of surveying than paper questionnaires to handle the anticipated amount of data from a large number of respondents. This survey effort needed to be able to provide the questionnaires before, during and after the exercise, tailored to different types of participants. On top of that, a wide range of questions provided by analysts from different nations needed to be correlated and presented only to the appropriate respondents.

A team from the National Aerospace Laboratory (NLR) coordinated this effort and explored ways to combine all requirements and at the same time take into account the burden of the questionnaires on the participants while maximizing the data outcome. A tool called Questionnaire for Utility Evaluation and Survey Tool (QUEST) was developed to implement their findings. The focus of this paper is on how the automated survey capability concept significantly reduced the burden on respondents while at the same time increasing the quality of the subjective data collected.

QUESTIONNAIRE DEVELOPMENT

Crucial for success of a survey is a good set of efficient questions. This means the questionnaires should not burden the respondents more than necessary, especially given the long mission days found in most military exercises and demonstrations. This means no question should be asked that is not relevant, and all questions need to be clear, concise, unambiguous, and applicable to the participants. Duplicate questions or lines of questions should be avoided.

Clear Questions

From a data collection point of view, the questions should be clear to avoid getting incorrect answers because of misunderstandings. The questions should not contain multiple questions within one question. Especially if a statement requiring a multiple choice Likert-scale response (where the answer is between "fully agree" to "fully disagree") is used, two different components within the statement may result in different agreement levels for the respondent. For example, the statement

"The new technology enables a faster kill chain and lowers risk on fratricide."

consists of two components not per definition linked.

The use of double negatives can unnecessarily overcomplicate the intention of the question. For example, the statement

"None of the pilots in the convoy was unable to identify the moving ground targets."

needs to be read at least two times to be understood.

Complex and long sentences with uncommonly used words, although perfectly correct, will cause an increased burden to most participants, especially non-native English speakers. Colloquial instead of formal language in questions can be more effective.

Question Type

Part of a question's clarity is in the answers. For example, the question

"What did you primarily use to complete this mission?"

could result in a various number of answers when asked as an open question, like

"My eyes", "My intuition", "Targeting pod", "Sniper pod" or "Low level pop-ups"

each of which in itself is correct, but only a few of these answers are the ones you are looking for.

In this case – apart from more specifically formulating the question – providing a set of possible answers the respondent can select from more precisely indicates which range of answers you are looking for and enables concise analysis after the event.

Apart from improving the data set you are collecting, the burden on the respondents will be less, because the time it takes to check a box is shorter than the time to write down the answer in plain text. Analysis on the feedback data found that, during Bold Quest events, the check box type of question took respondents 5-10 seconds to answer while free text took as many as 40 seconds.

Question Timing

Certain questions can be repeated per mission, per event in a mission or per technology used. The type of questions determines at which level it should be asked and thus how many times the respondents will be confronted with the same question. For example, the question

"Does the targeting pod give additional information on hostile ground tracks?"

could be asked after every mission the pilot was in during an exercise or even after each event in each mission, but the answers are unlikely to differ from one event to the next. The pilot will be asked the same question again and again, forcing him to provide the same answer each time. The result will be an irritated respondent and repetitive data; asking the question only after a certain number of missions or at the end of the exercise will be less of a burden to the pilot and may result in a more elaborate response while the pilot combines the experiences from all missions into one answer.

Other questions will require a compromise between the number of potentially different data points and the burden of the respondents. The choice can be dictated by the goals of the survey or exercise and the total burden of all questionnaires for the respondent. Be careful not to create over-simplification of data points by transferring questions from post mission surveys to end of exercise, to be asked one time only. For instance, the effectiveness of the technology can be determined by mission environment and interaction with other systems. If so, transferring questions on how effective the technology was from post mission to end of exercise (just asking once) could result in a significant loss of detail in the data.

Question Relevance

Asking an operator a question on a system he or she is not carrying or isn't even aware of exists will also cause an unnecessary burden on respondents. There is even the risk of respondents getting rebellious and entering bogus answers which pollute the results.

It is crucial that the systems or missions the individual respondents are involved in are identified and the questions are selected accordingly. A login script with user assigned login numbers can help to achieve this. If for some reason this information is not known in advance or if the system is not used during every mission, a question like

"Did you use the Firestorm system during this mission?"

as a conditional question prior to system specific questions will minimize the burden on the respondents to only one "irrelevant" question.

It is crucial to use clear, accurate and unambiguous trigger questions. In one of the BQ surveys in 2007 one such trigger question was not unambiguous. The question regarding having been trained on a system "prior to BQ" was used as a trigger for an entire one-time questionnaire. The analysts expected training on the system would be given in advance of the event. Since training was given **at** BQ (although prior to the missions), many respondents answered "No" discharging them from further questions on training. This resulted in a very incomplete data set entered by too few respondents.

Avoid Overlap

When the survey needs to be paired with a one-time direct respondent interview, based on feedback in the questionnaires, it is important that this is done a single time by one analyst, not sequentially by several different analysts on different occasions. If any type of survey is selected, whether paper, automated or web-based, it is important to make this the single type of survey the respondents will be confronted with during the test or exercise. Especially if part of the data sought after in the different surveys is identical. Experience has proven that overlap between the two resulted in an aversion to finish either of the questionnaires.

Even within a single survey, questions that are the same or appear similar can cause respondents irritation, and reluctance to fill out the questions accurately. If these similar questions are required to measure a construct, explanation of the reason behind the repetitiveness may alleviate this irritation.

Stating Importance

It is advisable to encourage the respondents to fill out the questionnaires by highlighting the importance of the results and the effect it may have on their future operations by giving feedback to improve equipment that is operationally useful. Promotion of the importance of survey data collection to seniors of the respondents and analysts of all participating nations assists in getting this message to the respondents.

Planning

The planning should include time to allow respondents to finish the questionnaires. Suitable locations easily accessible for the respondents are important as well. Survey proctoring by dedicated analysts who have knowledge of the automated survey is also recommended.

Questionnaire Optimization and Burden Analysis

Two time consuming phases in the preparation of the survey are critical for success:

1. Checking all questions for clarity, unambiguousness and singular statements, a correct and full selection of answers, and checking if logic is correct. Subject Matter Experts (SMEs) can be of great help as they can usually better place themselves in the respondents' situation. This applies to making the questions understandable for the respondents and knowing whether they will have the information to be able to answer the question after a mission.
2. Calculating the burden of the questionnaire per user type and phase of the event and acting on excessive burden.

Burden analysis takes into account the number and type of questions and the expected time per question, resulting in a total time per respondent. Whether this burden is acceptable depends on the respondent's background, available time and the mission impact on that same day. Overstepping the acceptable time will only result in degradation of data points by respondents who want to finish the surveys as quickly as possible.

Experience during BQ showed that questionnaires given daily post mission should be aimed at 10 minutes at best, up to 20 minutes at worst. On the first day responses may take longer than predicted but daily repeated questionnaires will lower the response time as respondents familiarize with questions and will be more aware during the exercise which events and facts are the focus of the questions. In general, about 60 to 120 multiple choice or 15 to 30 free text questions are practical for a daily debrief, given the historical timing information gathered on these question types. If a respondent experienced few events of interest (EOIs), the shortest time would apply, while with a large number of EOIs the total time should be within the maximum of 20 minutes. The latter can be achieved by limiting the number of question lines linked to EOIs.

AUTOMATED SURVEYS

In operational environments where preconditions change rapidly, automated surveys are indispensable as we will see later on. Automated surveys adapt to those changing preconditions according to for instance exercise phase, respondent or mission outcome resulting in respondent tailored questionnaires of minimal length.

Dependent Questions

Depending on the answer of a question the follow-up question may be different or may not even be necessary at all. For example, the question

"The DaCAS system I used was an addition to my mission success."
(Fully Agree, Agree, Neither Agree nor Disagree, Disagree, Fully Disagree)

should result in a different follow-up question for different answers. For example, when selecting *Fully Agree* or *Agree*, the follow-up question could be

"Which mission aspect did the DaCAS system enhance?"

or in case of one of both *Disagree* answers

"What should be added to the DaCAS system in order for it to be an addition to your mission success?"

That single first question may trigger a series of questions directly, or following more triggering answers, trying to unveil the basis of the system's mission success improvement or failure to be an addition to mission success depending on answer to the first question. This relieves the respondent disagreeing with the first statement from the burden of going through irrelevant questions on how the system added to mission success and provides the analysts with a clean and thorough data set.

Mission, Event or Phase

On repeated missions or during multiple events (i.e. attacks) in one mission some questions apply that generate repeated, almost statistical, data values. For example, the question

"How many nearby friends were displayed on your Blue Force Tracking system?"

can be compared between missions or events to get an idea of the maturity of the system over the course of the exercise or during the day. Although technical data may also be available, easily collected feedback in the survey from technology operators can assist in the technology analysis and assessment.

Respondents Role

During an exercise, various missions can be identified in which respondents play their own role and have a different perspective on the scenario taking place according to that role. It is no surprise that the same set of questions should not be presented to each respondent. However, similar questions that deal with the same data element aimed at different types of respondents should be formulated the same to ensure the perspective of these types can be genuinely compared.

Technologies

In a technology demonstration, a variety of new and under development technologies are subject to testing in a realistic operational environment. Many of them are at a low maturity stage and their developers seize every possible opportunity to get the operators' feedback needed to eventually field their system.

Many of the questionnaires used during BQ have a specific technology they apply to. Keeping track of which aircraft was equipped with which kit and which kit was used during a certain mission gave the analysts the opportunity to further tailor the questionnaire selection to the respondent indicating which mission he had just performed. The flexibility to even adjust the set of systems used on a day to day basis in the questionnaires enabled analysts to adapt the survey to the dynamic environment the concept development arena is known for.

Protecting Privacy

By password protecting the user login and by using numbers instead of names, privacy of personal information is guaranteed and anonymity maintained, as is ethically sound. For some nations, anonymity is required by law in surveys involving personal data used in research involving demographics. Respondents will be inclined to answer more frankly or enter more criticism when they know it is anonymous.

Identifying the respondents in advance can further reduce the number of questions asked. If the nationality, role, respondent type or even demographics are combined in a single user login name or number, the respondent can easily identify himself by logging on to the automated survey. This allows the survey tool to provide the appropriate questionnaires to the respondent. It also allows statistical data analysis based on type, nations, service and any part of demographics.

With this approach, questionnaires filled out by the same respondent on different instances in time can be easily linked without any effort by the respondent other than entering his user login. He will be able to look at his previous questionnaires and even finish or complement them when necessary.

Expectation Management

Besides reducing the questionnaire length and reducing the time it takes to fill out all questions, knowing what percentage is done and how long remains will manage a respondent's expectation. Knowing you are close to finishing the survey is a good motive to press on for just a few more minutes. Even knowing there is still 20 minutes of questionnaire left has a better effect on a respondent than not knowing if it will take a few seconds or several minutes.

Alternative Methods to Stand-alone Automated Survey Tools

To collect feedback from participants of evaluations and training, several methods can be used, such as interviews, paper and pencil questionnaires and electronic questionnaires (Leary, 1995). In the first method an analyst will speak with the participant directly (Leary, 1995). This gives excellent depth to the data collected as the interviewer can address unclear points and ask follow-up questions. However, the method also has several disadvantages. For example, it is time consuming and requires either a low number of participants or a significant number of analysts, analysts may influence the quality of the data, etc. (Gregory, 2007). If daily interviews are required due to different missions and events, this practical limitation is even more significant. The second method is a paper-based questionnaire (or in a simple text editor environment) allowing uncomplicated questionnaires with few decision points in the line of questioning (Rosnow & Rosenthal, 2002). It usually involves a conversion of the hand-written data to a database after the event. The third method is an electronic questionnaire, based on a database and using an interface such as an internet browser or specific software (Ryan, Corry, Attewell, & Smithson, 2002).

When a low number of participants is involved, both paper questionnaires and interviews may suffice to collect this feedback. However, as soon as more participants are involved for a longer time, or at more locations, these methods are impractical. Paper feedback results in a significant amount of work to transfer to a database or system where it can be analyzed. If 80,000 data entries (as collected at some BQ events) on paper need to be inserted in a database, this is a labor-intensive job. Although actual times for these transfers are not known in detail, if a low estimate of on average 10 seconds per answer is required to transfer free text answers, and using a historical (BQ) 30% free text answers, the total time to transfer those is 4000 minutes or 67 hours. This does not include the 70% checkbox type answers, which may be transferred in an automated way. Interviews may be less traceable for analysts not present and also more difficult to share as raw data. Incorrect transfer of results or losing data is a risk for both methods.

The electronic questionnaire is often used for surveys where each respondent goes through the questionnaire only once in a linear way, using for instance a web server to provide the questions and collect the data. However, in military operations research, usually both pre-exercise demographics and background questionnaire and a post-exercise questionnaire are required, as well as post-mission questionnaires. Furthermore, during military exercises the participants can not usually log on to the internet or even a local area network during the exercise. Sometimes the collected data is also classified, ruling out use of the internet.

Use of an automated tool may be combined with interviews. The data in the tool assists in finding the right person to interview and events or subjects that warrant further investigation by interview. This increases depth and understanding of the collected survey data and makes the interview and the questionnaire more effective.

Return on Investment of a Survey Effort Using an Automated Tool

Based on the experience at BQ, where all paper surveys, automated surveys, and interview methods were used in the past, rough estimates in total time and in man-hours (total for all analysts involved) are possible. The figures are indicative, not exact, and depend on how it the data collection is approached, size, collection ambition and analytical goals. Man hours may not always match time lines as more analysts may contribute for a task (for instance analysts doing interviews during the exercise). On the other hand, a task may not be worked on full time (such as improvement of questionnaires where feedback of other analysts and SMEs is often required).

In Table 1, these figures are listed for a BQ type event, 200 participants, 80,000 collected survey data points, several questionnaires, user types and locations. Strong differences occur in transferring the results to a database and total time lines of interviews compared to surveys. The man hours at the exercise to support the surveys and the interviews are significantly different. Finally, the resulting dataset, given as percentage from an ideal 100% (all pre exercise, post mission and post exercise surveys completed), is different as well. The lower percentage for paper is based on having less opportunity to repeat questionnaires for different events of interest in a mission. Also, in general less complexity in triggered parts of questionnaires is possible, as well as less different questionnaire versions for different phases in the exercise.

Table 1: Time lines and man-hour estimates of tasks for survey/interview BQ-type event (man-hours all analysts)

Task	Paper Survey	Electronic Survey	Interview
Creating Questionnaires, including SME feedback	8 weeks/300 hours	8 weeks/300 hours	4 weeks/150 hours
Collecting Questionnaires removing duplication, improving questions	3 weeks/40 hours	3 weeks/40 hours	0 weeks/0 hours
Transferring questionnaires into paper/tool, test	3 days /25 hours	1 week/45 hours	0 weeks/0 hours
Analysts' time at exercise assisting data collection	2 weeks/55 hours	2 weeks/55 hours	2 weeks/300 hours
Time to transfer results into reporting or database format	2 weeks/70 hours	1 day/5 hours	1 week/40 hours
Resulting dataset, % of possible total	40-80%	80-95%	15-30%
Total time/total hours	15 weeks/ 490 hours	14 weeks/455 hours	7 weeks/490 hours

Compared to other methods, total man-hour costs of use of an automated tool is significantly lower regarding extracting and using the data. Preparation and proctoring is comparable to paper- or web-based version as respondents need assistance in all cases. Interviews are different in man-hour costs; preparation is expected to be shorter than preparing surveys. However, since every interview requires at least one analyst and involves only a few respondents, cost of the interviews will be higher if a significant number of participants and events need to be covered. When interviews are used for depth of data only, the data in the tool may assist in cutting the interview hours.

If the tool itself needs to be developed entirely or further developed, this amounts to a significant investment. If the tool already exists and is set up in a flexible way, this is not required. Entering the prepared questionnaires into it and testing will take around a week for a Bold Quest sized event. These costs are comparable to those required for paper and web-based questionnaires.

The preparation of the questionnaires, as is the case with paper and web-based surveys, takes a large number of man-hours and total time. However, this phase determines the overall data collection success, as the respondents' burden and clarity of data sought strongly influences data quality and quantity. As this feedback data can influence development and procurement of future technologies at an early stage, the benefit of these costs can be significant.

RESULT MANAGEMENT

A data set containing thousands of answered questions and data entries is somewhat tedious to handle during daily or post-exercise analyses. An automated survey is able to maintain a well-organized structure that decreases the analyses burden enabling rapid after mission debriefs and quick reports.

Setup and Synchronization

Synchronizing data means making sure that all relevant data that has been entered in the survey up to that point is available to the respondent returning to the survey or to the analyst with a minimal time delay after data entry. In case of the respondent this means that he is able to revisit earlier entries and continue where he left off with his questionnaires. For the analyst it means that he is able to get a complete overview of all data entered by all respondents or a specific subset of respondents that he is interested in.

There are basically two approaches in keeping the data set synchronized. In a controlled environment where all data entry points are network connected via a LAN or the internet if classification is not an issue. With these types of connectivity, a server-based application with continuous synchronization of the data set on the server is preferable. Data managers do not have to worry about synchronization, while all data is stored in a central location and readily available to all analysts.

However, in most operational exercises this networked solution is not possible, and stand-alone automated surveys are required. The reason for this stand-alone survey setup can be a lack of internet connectivity, or classification issues. A local network may not be present at all or may not cover all locations where respondents return from missions. Daily questionnaires should be conducted in the field as respondents do not return to base for hours, or even days after the mission.

The challenge in using a stand-alone solution is that systems need to be synchronized on a frequent (usually daily) basis. This way respondents and analysts work with the same full data set in any location and on any machine. The automated survey needs to facilitate this synchronization with export/import functionality.

During events where respondents fill out their questionnaires daily, the data set should also be synchronized daily. Otherwise respondents returning the next day on a different survey machine will not find their previous survey data entries frustrating the respondent. Trust in the automated survey will be sharply reduced as data appears to be lost.

Data Output

Key in any automated survey is the immediate availability of the data for the analyst to do assessments. The ultimate goal of the survey collection is not to gather and store a large data set, but to use the collected data for research, system development or system acceptance.

Reports with for instance statistics of number of positive responses versus total number of responses will increase the usefulness of the automated survey with respect to quick assessment of the data and analysis for a final report. Filtering is useful for analyzing different responses for different groups or situations. An overview of all free text answers to a certain question can help to support the statistics or assist in early discovery of problems respondents may have in understanding or answering a question.

A sample of direct statistics of the responses to closed questions can be seen in Figure 1. An extract from a full report of statistics and free text answers can be seen in Figure 2. Export of the response data to a common format like Excel allows analysts to use more advanced analysis tools or use this exported spreadsheet directly. As was mentioned earlier in the process of questionnaire generation extra attention should be paid to rewriting free text answered questions to multiple choice type questions. This does not only result in more clarity like explained in the earlier example, but also enables statistical analyses. Substituting the scale of *Fully Disagree* to *Fully Agree* to a Likert scale of one to six transforms the measurement level of the data to an ordinal level. This opens up the possibility to calculate an average among all respondents of a certain group and translating that back into a general feeling of that group (Meyers, Gamst, & Guarino, 2006). An average of 5.5 would mean that the group's opinion generally lies between *Agree* and *Fully Agree*.

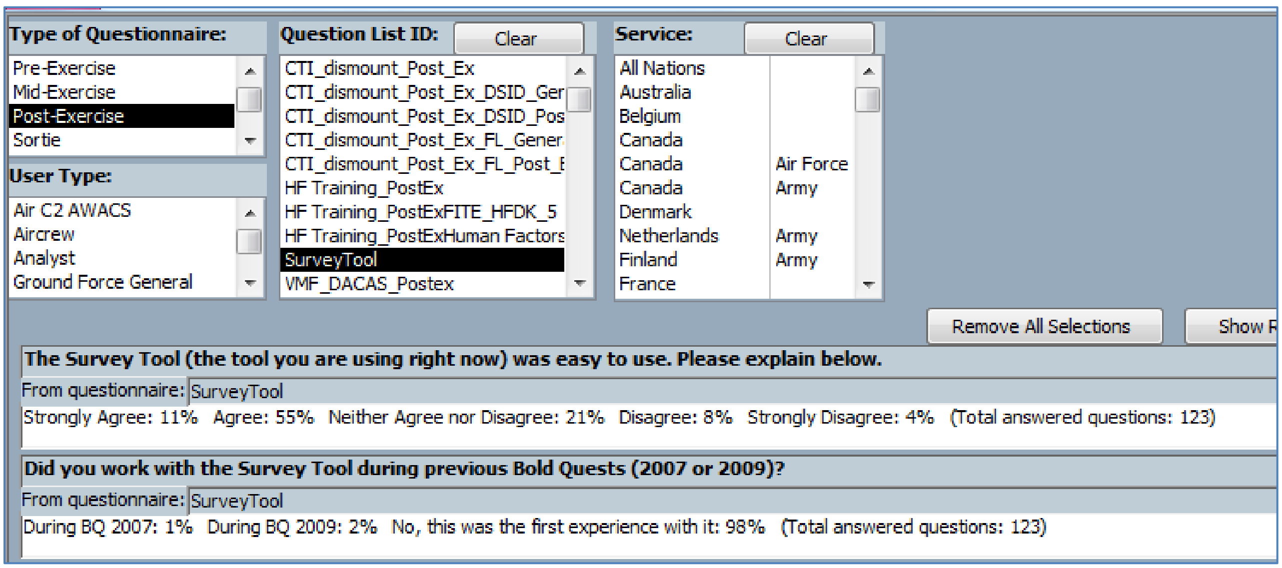
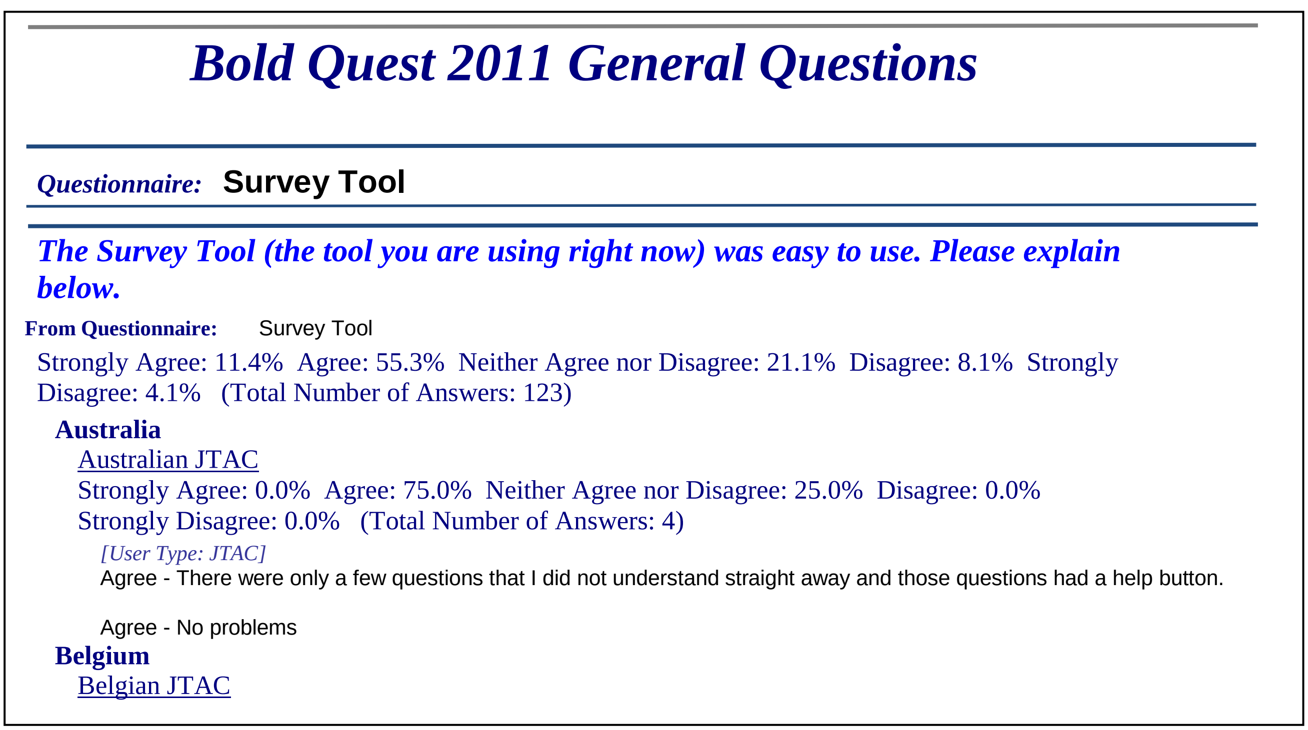


Figure 1: QUEST statistics form for analysts



Bold Quest 2011 General Questions

Questionnaire: **Survey Tool**

The Survey Tool (the tool you are using right now) was easy to use. Please explain below.

From Questionnaire: Survey Tool

Strongly Agree: 11.4% Agree: 55.3% Neither Agree nor Disagree: 21.1% Disagree: 8.1% Strongly Disagree: 4.1% (Total Number of Answers: 123)

Australia

Australian JTAC

Strongly Agree: 0.0% Agree: 75.0% Neither Agree nor Disagree: 25.0% Disagree: 0.0% Strongly Disagree: 0.0% (Total Number of Answers: 4)

[User Type: JTAC]

Agree - There were only a few questions that I did not understand straight away and those questions had a help button.

Agree - No problems

Belgium

Belgian JTAC

Figure **2: QUEST automated report of statistics and free texts**

LESSONS LEARNED

Lessons learned during Bold Quest are applicable to most survey opportunities. They can be summarized as follows:

- Instructions and a manual for survey developers is crucial. A template in MS Excel assists in submission of the questions and answers.
- Preparation of the survey takes up to 4 weeks for large numbers of questionnaires, which should be planned for. Milestones for analysts submitting questionnaires and SMEs assisting in questionnaire improvement are determined to allow survey development for this period. The deadline for draft survey submission should be more than a month prior to the exercise, depending on complexity and numbers of questions.
- Respondent community feedback is an effective means to improve future questionnaires. Determination of question priority can help survey development and burden alleviation while still collecting the most important data.
- Respondent community feedback (using SMEs) is crucial to improve the face validity of questionnaires.
- End of exercise surveys are an effective means of alleviating the daily burden. However, survey fatigue or leaving the exercise early may lead to less or no responses to the end of exercise questionnaire. Therefore, the end of exercise survey needs to be offered from the very beginning of the exercise, communicating to the respondents when this needs to be addressed.
- Ensure the use of terms is clear to the respondent. If for instance event of interest is mentioned, the respondent may interpret this differently to an analyst. Especially if these are triggering questionnaires, the loss of data can be significant.
- During BQ events, high numbers of questions were transferred to the post exercise questionnaire. It was found that planning for respondents to finish this was crucial. At some events it was indeed left to post exercise (after last mission), to find that many people already left. At one occasion the last mission day was hit by bad weather, so all missions were cancelled. Very few post exercises were responded to at the air base because of that. Too early may interfere with the quality of the answers as more experience with the technology will occur during the last missions.
- At some BQ events up to 60 laptops were used, distributed over up to six locations. A network with a MS SQL Server as database was also used. Synchronizing 60 laptops with each other is time consuming, first drawing data from all, and next repeating this to push all data from central database to all instances. A number of instances were on analysts' or respondents' laptops as well, which technically worked well. Keeping track of response rate is more difficult with such a distributed effort. At remote locations, effort and a dedicated survey proctor is required to achieve the excellent response rates usually achieved at the BQ event's main location.
- Survey burden analysis is crucial to collect data from all respondent groups. Using historic response times to types of questions and estimating the average amount of triggered sub-questionnaires assists in more specific survey response time management for all groups.
- During BQ, the analysts' effort in improving the questionnaires and lowering burden to 10-20 minutes showed to be a long process. However, as seen during some event where budget was limited and the full process of improvement was not possible, the response rate was low. After dynamically redesigning part of the questionnaires during the event, getting close to 20 minutes of response time, this rate improved. As stated earlier, getting respondents to return after a bad experience is hard. Although most data is still valid, many questionnaires were either not finished or never responded to at that early phase.
- During BQ events, both active promotion and more informal offering of this survey occurred. Especially when senior leadership of the organizers and nations was also involved, the data set was close to 100%, while in other cases response rate started with 20-30% only, going up slightly to the end. If the momentum is lost at the start, getting up to speed is difficult. Cooperate with those planning your exercise to ensure the survey data collection by the respondents is planned for in their schedules. Inform the respondents as well as their superiors of the reason for surveying, as well as the importance of this data collection.

CONCLUSION

The previous sections describe the findings during past survey experiences with respect to efficient automated surveys with a minimized burden on the respondents and the analysts. This section summarizes the features a good automated survey should incorporate as listed in Table 2.

Table 2: Key Features of the Perfect Survey during exercises

	Feature
1	Clearly formulated questions with concise answers
2	Respondents spend minimal amount of time on questionnaires
3	Only relevant questions with follow-up questions when applicable
4	Questions can be asked once, mission-based, or event-based
5	Stand-alone and networked setup of survey tool, avoiding internet connection requirement as usually not allowed/available
6	Questionnaires change flexibly with exercise phases
7	Survey tool reports dynamically grouped results, like timing and respondent type
8	Survey tool provides output formats compatible with other software tools
9	Survey tooling should be 100% stable, negligible faults in software

It is these features that make an automated survey very adaptive and flexible to every situation. At the same time, as respondents are only subject to the absolute minimum amount of questions needed for data collection, the respondent's burden is minimized and data quality maximized.

The stability of the technical solution is a crucial requirement for using any tool. During all Bold Quests, the setup used achieved this, resulting in no lost data and increased satisfaction of respondents in use of the tool, as was shown in user feedback within the questionnaires. This use of surveys, paired with with interviews, proved to be effective to collect warfighter feedback data in a complex exercise with up to 200 respondents in over 10 roles and from 10 nations and services. The results assisted in thorough analysis and assessment of technologies, allowing early adjustments of operational shortcomings.

ACKNOWLEDGEMENTS

Thanks to Emilie Reitz for her support of the survey effort during Bold Quest events. As a strong advocate of the survey effort and QUEST, she made the data collection during BQ such a success. Also thanks to Anneloes Maij (NLR) for assistance on the statistical and psychological aspects of surveys.

REFERENCES

De Vocht, A. (2003). *Basishandboek SPSS 11 voor Windows.* Utrecht, UT, The Netherlands: Bijleveld Press.

Gregory, R. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston, MA, USA: Allyn and Bacon.

Leary, M. R. (1995). *Introduction to behavioral research methods.* Parcific Grove, CA, USA: Brooks/Cole.

Merriam-Webster dictionary. (sd). Opgeroepen op 5 28, 2015, van http://www.merriam-webster.com/dictionary/objective

Meyers, L., Gamst, G., & Guarino, A. (2006). *Applied multivariate research, Design and interpretation.* Thousand Oaks, California, USA: Sage Publications, Inc.

Rosnow, R., & Rosenthal, R. (2002). *Beginning Behavioral Research* (Vol. 4). Upper Saddle River, New Jersey, USA: Pearson Education Inc.

Ryan, J., Corry, J., Attewell, R., & Smithson, M. (2002). A comparison of an electronic version of the SF-36 general health questionnaire to the standard paper version. *Quality of life research, 11*(1), 19-26.

Sills, S.J. & Song, C. (2002) Innovations in Survey Research: An Application of Web-Based Surveys *Social Science Computer Review* 2002; 20; 22