Adaptation of Research Methods to Support the Analysis of Inter-Agency Training

Ronald W. Funk
Centre for Operational Research and Analysis
Defence Research and Development Canada
101 Colonel By Drive, Ottawa, Ontario, K1A 0K2, Canada

The analysis efforts in support of various experiments and exercises used to train operators had to adapt to Vancouver 2010 Olympics and Paralympics (V2010) needs. This paper documents how Defence Research and Development Canada (DRDC) observation and survey methods evolved from standard research protocols in order to generate timely and operationally relevant findings for the V2010 Integrated Security Unit (ISU) user community. The goal of this paper is to explain how the circumstances surrounding several experiments and exercises influenced the analysis approaches and subsequent revisions to the methodologies used.

1. Introduction

Defence Research and Development Canada (DRDC) was heavily involved in the delivery of Science and Technology (S&T) to support the Vancouver 2010 Winter Olympics and Paralympics (V2010). That long and complex undertaking ultimately resulted in the successful delivery of a wide range of S&T that helped support V2010 planning and execution needs. The focus was on security and at the Integrated Security Unit (ISU) along with public safety at the Integrated Public Safety (IPS) as well as military support by the Joint Task Force Games (JTFG). The bulk of the activity was managed through the auspices of the DRDC Centre of Security Studies (CSS) using the Major Events Consolidated Security Solutions (MECSS) project.

A major component of the DRDC activity involved provision of analysis support to many training exercises conducted to prepare various staffs for their respective roles and duties within the integrated V2010 Command and Control (C2) architecture. This was not a simple effort because the C2 collective training was conducted in parallel with defining the C2 architecture while building the physical infrastructure. The consequence was that the analysis support had to adapt to an ever evolving set of assumptions and circumstances. This paper describes the author's personal perspective about several of the key training events that he helped to organize and/or analyze. Because of the sensitivity about such matters, this paper does not discuss how the actual exercises events were conducted or talk about the readiness of any units or participants involved. The reader should also keep in mind that there were many similar exercises were supported by DRDC beyond the subset summarized here.

2. Initial Experiments

The start of DRDC exercise analysis began with two separate but concurrent experiments for the ISU and JTFG in November 2007. This occurred at the end of a Concept Development and Experimentation (CD&E) campaign with the goal of helping the V2010 ISU better understood its operating environment. The objective of defining the major processes was to ensure the creation of an effective C2 architecture for the ISU.

The ISU portion of the experiment for the ISU had been in preparation for almost a year while the military portion for JTFG was pulled together in the last two months to support a new military sponsor. As a result, the JTFG portion was run in parallel with the ISU pat so as not to interfere with the ISU detailed events. Data and activity was limited to a one-way controlled transfer from the ISU to JTFG. Each unit used a separate set of scientific staff to pursue distinct (but in many respects complimentary) research objectives. Both experiments had instrumented consoles and with periodic interruptions for operator impressions. Onsite monitoring was done by a large number of scientists following the detailed actions of individual operators. Extensive surveys and interviews were also conducted. Participants provided informed consent by signing waiver forms that explained the uses of the data.

A useful Measure of Performance (MOP) that became apparent after the experiment was the time taken to deliver actionable products to the operators (i.e. product latency). The ISU portion of the report had a high level Quick-Look Report (QLR) released 25 days after the experiment. The JTFG product latency MOP was a Letter Report (LR) of 46 pages passed along within 17 days of the event and the experiment was then briefed to JTFG staff in 45 days. The detailed DRDC reports were published up to a year later.

The experiment analyses involved the usual detailed research procedures and were published in peer reviewed DRDC technical reports. The QLR and LR employed science vernacular and caveats that made the results of limited use to operators except to confirm what they had observed themselves. This classical approach to conducting detailed experiments may have addressed the R&D needs of DRDC but they did little to provide either timely or tangible findings for the operators involved. This was because the accumulated evidence was handled in ways most suitable for subsequent R&D activities and, for the most part, was not delivered in a form that was responsive to or directly actionable by the operational community.

Despite the lack of product applicability, the experiments did provide a tremendous service towards future operational cohesion by bringing together ISU and JTFG staffs to start working on complex security problems. The greatest value of this early training opportunity was to develop a collective shared awareness about the operating environment and make social connections. The insights were subsequently translated into procedures to ensure critical information would flow to where it was needed. As an aside, these experiments also exposed the operator perception that, even though everybody felt the training value was high, they did not like being treated as “lab rats” by the scientists.

The JTFG version of the initial experiment also allowed the JTFG HQ cadre to work together as a team for the very first time while at the same time being exposed to the latest C2 technology that they would eventually use for V2010. This was done at short notice with only limited training at the start of the experiment. JTFG operators initially felt overwhelmed by the situation but within hours they started acting as a team. The JTFG Deputy Commander referred to this experiment as one of the early defining moments for his staff.

JTFG conducted a second experiment in October 2008 using new applications developed from feedback from the first experiment. They felt empowered to see they had helped progress the way C2 is done in Canada and embraced the new capabilities. The experiment also primed them for the exercises that followed soon thereafter.

3. Initial Major Collective Training

Following these initial experiments the ISU and JTFG both conducted detailed planning and built the C2 architecture and physical infrastructure needed to support the anticipated integrated security operations. In November 2008 the ISU and JTFG conducted their initial major collective exercise together to practice with the technology and organizations they planned to use during the Olympics. In the case of the ISU this involved assembling a cadre of police officers from different forces and agencies to work together for the first time through complex regional and theatre level procedures on a brand new network. For the JTFG staffs, this involved borrowing DRDC assets to activate an interim network solution to mimic the permanent infrastructure that was still being assembled.

The first exercise analysis started out as largely a repeat of the previous year's experiments. However, since this was an exercise, the scientists deployed to assist could not employ intrusive methods like periodic interruptions or instrumented consoles. Instead, they were limited to on-site monitoring and a major survey instrument. A major observation drawn from these exercises was a concern about legislation that limits what information can be transferred between public safety and the military. The DRDC analysts got around this conundrum by setting up parallel analysis teams using common methods and survey instruments but not sharing specific results between the teams.

Uncertainty about the analysis requirements existed until shortly before the exercise so the survey was only compiled a week before the exercise started. As a consequence, it was largely based on previous experiment surveys where most questions sought details about how each position operated. The survey was supposed to be administered via Excel on both networks with Visual Basic for Applications to automatically parse the inputs into a single data worksheet and calculate routine statistics (i.e. means, standard deviations, histogram values and correlations). Network arrangements fell through so the survey was converted to paper form with the analysts entering the data into the spreadsheets afterwards. Very few interesting correlations were found because each position had to answer the questions from a unique perspective.

Onsite monitoring was done by a limited number of scientists observing operator activity using clipboards (owing to electronic emission concerns) and list of *a priori* research questions. Unfortunately the ISU portion had very limited time for valid observations because of a need to conduct initial connectivity checks and do introductory operator training. Several of the scientists also expressed a cultural discomfort caused by the compromises imposed on them with having to broadly monitor several operators (without any script to refer to) instead of being allowed to focus on the structured activity of one or two.

The DRDC scientists working with the ISU and JTFG were able to gather a wide range of useful observations despite the shift in focus from experiments and exercises. This occurred despite the scientists being unable to control variables to address *a priori* research questions but instead having to focus on pinning down key decision points and identifying emergent behaviour. The observers in the JTFG Joint Operations Centre (GJOC) also created several innovative ways to quantify interesting observed effects using simple tables of values gathered at regular sample periods and from compiled examples.

The operator perception about the survey instrument ethics statement was quite another matter. The RCMP and CF lawyers raised concerns about legality of the "informed consent" form because it implied *carte blanche* use of the data for any further purposes. In addition, many of the operators (both ISU and JTFG) independently revolted about signing any forms not directly related to their duties during the exercises. At the same time all the operators (including the lawyers) were adamant that they wanted DRDC to use of the survey answers to find ways to improve the operators' working environment. In the end, it was agreed that informed consent had been provided since the consent form was stapled on top of the questionnaire and the act of returning the completed survey was deemed to be consent to use it for operational assessments.

The statistical analysis of the detailed survey also proved to be of limited use because most questions were related to specific responsibilities of each position and thus, every answer had its own specific context and perspective. Trying to aggregate the responses or correlate them was "like chasing clouds". The resulting statistics could only rarely be linked to useful underlying causes or needs.

This DRDC scientific support to the first exercise differed from the previous year's experiments in that the analysis effort was focused on preparing and releasing the LR in time to keep up with operator deadlines. For example, the product latency MOP for draft JTFG LR of 33 pages to JTFG was cut to 4 days in order to facilitate a critical meeting. The official signed version was then released 38 days after the exercise. The use of convoluted scientific phrasing in the previous reports was largely replaced with more direct and specific wording in these and subsequent reports.

The analysis documented many different facets of the exercise. The most important of these were several emergent behaviours[1]. The operators took steps to document and reinforce the positive ones and conducted additional training to correct any potential negative effects.

[1] The specific cases involve operational procedures that must remain classified. One example of a type of emergent behaviour that can be discussed involved military staff monitored ISU events looking for activity that might lead to requests for military support. The military would then start background preparations and notify the ISU of their state of readiness; most often this occurred before the ISU made a formal request.

The analysis of this first major exercise was done with the full knowledge and intent that no subsequent DRDC reports would be written up until after V2010 was completed. The reality was that training preparations were proceeding at a very fast pace, which left no time to do any analysis before the subsequent exercises needed to be supported. The DRDC analysis organizers also had to manage organizational sensitivities about sharing data for other uses and ethical concerns about the extent to which further analysis of the data could be authorized beyond direct operational needs.

4. Second Major Collective Training

The second major exercise of ISU, IPS and JTFG staffs in February 2009 was a week of training activities that was built upon the features of the previous collective exercise. In anticipation of this, some DRDC scientists prepared an experimental design document in the hope of applying an orderly data collection plan and analysis structure. The survey questionnaire was also a repeat of the prior exercise so comparisons between the two events could be made. The major differences were that the ISU networks had become much more robust and the JTFG was using its operational networks.

The actual activities conducted in the second exercises went relatively smoothly but the data collection plan was abandoned. It simply had not adequately accounted for the increased complexity and volatility due to the number of sites and variety of activities that were observed upon. This occurred in part because the ISU had activated a subordinate Command Centre within the building as well as a second site where the ISU and JTFG conducted combined C2 activities. JTFG also ran a separate senior leadership Table Top Exercise (TTX) concurrent with exercise activities.

It was through the dedication of many analysts that all sites and major activities were successfully monitored and then analyzed within a short period. Several new emergent behaviours were observed and subsequently acted upon by the operators. The product latency MOP for the JTFG draft analysis LR of 67 pages was 26 days with a subsequent formal released as a LR at 45 days after the exercise.

5. JTFG Final Collective Training

The final full scale JTFG training session and exercise, using its full set of assigned personnel, occurred in a week long deployment to the operating locations in September 2009. The author had just completed an extended compulsory vacation of three months so the short preparation period available limited what could be planned to support this event. A further complication was the need to summarize key points in time for JTFG hot-wash review by all participating personnel a few hours after the exercise ended.

These constraints limited DRDC analysis support efforts to using two senior scientists to share monitoring duties around the clock at one site. The scientists' familiarity with JTFG operations during previous exercises made it much easier for the scientists to pick out noteworthy events and activities. However, the pace of exercise activity and need for quick deliverables precluded the use of surveys for this exercise.

The analysis initial deliverables MOP started with passing along a copy of the analyst notes to JTFG training staff at end of each day. The other product latency MOPs were a set of briefing slides sent to JTFG within 11 days of the event and a draft LR of 13 pages passed along 19 days after the exercise.

6. ISU Final Collective Training

The final ISU collective training for the ISU command centre staffs was conducted as two similar but separate week long exercises during October and November 2009. Each exercise was focused on training ISU staffs and the national managers in many locations. There were two separate training audiences; the first was the command team coordinating activities within ISU units while the second tested another ISU command team but with main focus on other government agencies needing to test their procedures.

The DRDC analysis support followed many of the previous broad methods but they were updated significantly in the way they were applied. The first major difference was that the analysts monitored specific areas of activities throughout the day but only in terms of noteworthy occurrences rather than details. In the evening they each discussed their key observations with the other scientists and then collected the major points into a summary that the lead scientist immediately passed along to the ISU senior managers. This allowed them to appreciate the issues and correct shortcomings before the exercise ended.

A second major difference was that a general questionnaire was prepared seeking operator feedback on generic effects. Comments were captured using paper surveys with questions organized around 5 point Linkert scale and written remarks. The inputs were manually processed by scientists into simple spreadsheet from which means, standard deviations and histograms were compiled. The results were categorized by location, section and function to determine how the perception about training readiness varied within subsets of the population. The generated results were clear and easy to interpret.

The product latency MOP of analysis results of the first exercise were reported in a quick-look LR of 23 pages only 5 days after the event. There was also a separate special draft LR of 15 pages that took 17 days. No final LRs were prepared because the ISU stood down before the documents could be prepared.

Another major divergence from previous exercises was that the analysts also made a collective qualitative assessment about several capability metrics at the end of each day. The group consensus was used to fill in a colour coded effectiveness matrix along with points to help explain the context of each score. The lead scientist passed the results directly to the ISU lead trainer each evening for his use. There was no intent to report or distribute them anywhere else.

The second exercise repeated the same scenario themes to a new training audience within the ISU. Its focus was on probing and testing the strategic government decision making processes. The survey was repeated so the responses could be compared between the exercises. Several useful insights were obtained about the exercise differences by

locations, sections and functions. The most telling was the overall increased operator confidence due to the less intensive cadence of operations plus a small cadre returning from the first final exercise. Both aspects boded well for the conduct of actual operations during V2010. A draft LR was started but not completed because the lead scientist had to start making final preparations for the actual V2010.

7. JTFG Deployment Verification

This JTFG exercise was conducted in late January after JTFG had deployed and was already conducting operations. This exercise was conducted quite differently from any previous ones because it had to co-exist with regular operations. Its role was to test if a wide range of procedures would behave as expected but to do so it relied upon special procedures to clearly delineate what activities were caused by the exercise scenarios. The exercise itself was run by a JTFG Red Team with the help of several DRDC analysts at the exercise control location monitoring activity to see if any unexpected gaps or bottlenecks occurred in the procedures. The product latency MOP of findings involved a LR of 11 pages delivered 16 days after the exercise.

8. Analysis Lessons Learned

The V2010 analysis support activity started with the classical scientific approach used for the design and conduct of experiments. Those measures started to evolve once it became apparent that they did not address the needs of the client. The shift in research methods took place in the form of measured changes to address the obvious shortcomings rather than any abandonment of standards. The initial changes were to speed up document preparation efforts to achieve a more timely delivery of products. The data collection was also done in less detail so the results could be analyzed quickly to deliver useful analysis products. Despite the compromises required by the circumstances all analysts ensured they applied proper scientific methods to the situation at hand.

Since the goal was the timely delivery of operationally relevant results MECSS put in place procedures for the publication of LRs to ensure information quickly went where it could be used. This included allowing operators informal access to draft versions to enable further planning while the formal wording was polished in the approved product. The product latency MOP measured delays accrued in passing along actionable information. DRDC also accepted that publication of most research reports had to be delayed until after V2010 operations completed so that scientists could remain operationally engaged with the operators.

The range of analysis opportunities over the many exercises exposed a significant number of analysts to working situations that they would otherwise never had a chance to see. In many cases these scientists were able to observe and identify emergent behaviours they would never suspected were possible. This is likely to have a long term positive impact in the way these same analysts research C2 in the future. If nothing else they can at least understand why a lack of time and evolving operator procedures may undo what may have been assumed to be an ironclad experimental design.

There were also many lessons learned from preparing and administering survey questionnaires. The most important of these are that exercise surveys need to focus on generic considerations designed to yield actionable findings. Creating surveys that probe too many details end up forcing every position's answers to be contextually unique with the corresponding loss of statistical related insights.

The ethical dilemma concerning experiments that involve manipulation of individuals does not hold for exercises where injects are imposed for operational reasons. This means that the scientist role must be modified to that of an observer of activity that is independent of what the scientist can control. The operators sensed this shift in the scientists' roles and freely contributed survey comments as professionals instead of feeling they were being personally probed like "lab rats".

9. Conclusions

The effort involved in putting together the analysis support to V2010 was huge. Most analysis objectives were successful but valuable insights were also learned from a few failures encountered along the way. The operators recognized that they had received a level of analysis support that would not otherwise have been possible without DRDC support. Most importantly, it was done in a way that ensured the results were delivered in time to become part of the next training iteration.

For those scientists involved throughout the planning and operations period it was a most gratifying experience to see how the operational staffs evolved. The operators may have been suspicious or confused about the scientific observers at the start but by the end they accepted and relied upon the scientists to provide them a transparent and honest analysis.

The greatest sense of personal accomplishment was when the author saw the ISU and JTFG staffs conduct V2010 operations with a high degree of professionalism and confidence. There is every reason to believe that the knowledge and experience had been built, at least in part, from the DRDC analysis support to the experiments and exercises.

This paper has described why and how research methods of interagency training experiments and exercises need to be adapted to deliver objective analyses when it is most needed. This was a necessary change from classical research methods because the situation did not allow scientists the luxury to control variables to the same extent they would for an experiment done in the laboratory.