

**Proposed Evaluation Plan for the Gulf Oil Spill Training:
A Systematic Comprehensive Training Evaluation Process**

**Sarpy and Associates, LLC
Charlottesville, Virginia**

January 7, 2011

A Proposal Submitted to the National Institute for Environmental Health Sciences

**Sarpy & Associates, LLC Proposed Evaluation Plan for the National Institutes for
Environmental Health Sciences' Gulf Oil Spill Training:
A Systematic Comprehensive Training Evaluation Process**

The health and safety training literature has long recognized the need for comprehensive, systematic evaluations of the effectiveness of training with respect to increasing both safety knowledge and exhibiting safe work behaviors, while reducing outcomes such as accidents, illnesses, and injuries (Ford & Fisher, 1994; Burke and Sarpy, 2003; Burke, Sarpy, Smith-Crowe, et al., 2006; Robson, Stephenson, Schulte, et al., 2010). The Sarpy and Associates, LLC (hereafter Sarpy and Associates) research team has established a rigorous process for objectively evaluating the program effectiveness of various health and safety and emergency preparedness and response training efforts (Sarpy 2007a; Sarpy & Goldstein, 2010; Sarpy & Kaufman, 2005;). The purpose of the present proposal is to describe the Sarpy and Associates evaluation process and its potential contribution to evaluation the National Institute of Environmental Health Sciences Gulf Oil Spill (hereafter GOS) health and safety training including a task list and timeline for the proposed project.

The Sarpy and Associates evaluation process is an integrated programmatic assessment that incorporates elements of both process and impact evaluations as outlined by the NIEHS (NIEHS, 1997; Sarpy, Chauvin, Hites, et al., 2005). Impact evaluations are those that measure *what* effects the program has with respect to its intended goals and objectives. Process evaluations, on the other hand, are evaluations that are designed to measure *how* and *why* a given program achieved these effects. Therefore, impact evaluations tend to involve quantitative/numerical data (e.g., number of participants successfully trained) whereas process evaluations tend to involve qualitative data (e.g., a narrative describing the specific strategies that are used for training and perceptions regarding how and why those strategies are employed). In order to gain a comprehensive representation of *what* practices or strategies are utilized, *why* and *how* those particular strategies have been implemented, as well as the relative effectiveness of the strategies for achieving program goals and objectives, a combination of process and impact measures are collected and analyzed (Berger & Rice, 2000; Burke, Sarpy, Smith-Crowe, et al., 2006; Fitz-Gibbon & Morris, 1987; Goldenhar, LaMontagne, Katz, Heaney, & Landsbergis, 2001; Patton, 1987; Patton, 2008; Prue & Feldman, 1997).

The Sarpy and Associates evaluation process also implements a multi-source evaluation system that is associated with a 360 degree feedback system. This type of evaluation system involves ratings of a particular target, such as the GOS training, from various sources, which may include on-shore and off-shore volunteers, technical specialists, trainers (including the Petroleum Educational Council

members), grantees, and customers (including BP, federal and state governments that had their employees trained; contractors). Obtaining information from these multiple sources, experiences and perspectives imparts a more thorough and accurate analysis of the effectiveness of the training itself. This multi-source evaluation is particularly relevant for evaluating GOS training given the key role each of the stakeholders play in the training programs' success. Further, gathering information from the various stakeholders lends greater perspective and thereby more credibility to the evaluation process, particularly regarding feedback and utilization of the results (Gotsch & Weidner, 1994; NIEHS, 1997; Sarpy & Kaufman, 2005; Sarpy, Vaughn, & Goldstein, 2010; Shannon, Robson, & Guastello, 1999).

To gain a comprehensive depiction of whether the GOS training programs and goals were attained as well as the relative effectiveness of the instructional activities and components for achieving these program goals and objectives, a combination of both quantitative and qualitative data will be collected and analyzed. Further, because data is collected across program stakeholders, a multi-method approach is necessary in which various methods of data collection are incorporated into the evaluation design (Creswell & Plano-Clark, 2007). That is, both qualitative and quantitative data will be gathered using a combination of complementary methods that include questionnaires, focus groups, and structured interviews. In order to integrate this information, content is standardized across survey methods and sources. In this way, direct comparisons can be made across stakeholders and training offered (Sarpy & Kaufman, 2006; NIEHS, 1997; Shannon, Robson, & Guastello, 1999).

The Sarpy and Associates evaluation process will be designed to assess the GOS health and safety training program effectiveness with respect to each of the NIEHS identified worker training program criteria. Additionally, information from the evaluations will be used to identify Best Practices for the GOS training in achieving the NIEHS criteria (Bogan & English, 1996; Sarpy, Kaufman, Smith-Crowe, 2002). Importantly, these best practices typically include appropriate use of adult education techniques and adherence to principles of adult learning for effective direct training and quality assurance. Finally, the GOS training also will be evaluated according to overall performance and impact on their respective trainees, communities, and the field of environmental work.

The Sarpy and Associates evaluation process meets the following specific objectives: (1) to develop and implement an integrated programmatic assessment, which incorporates elements of both process and impact evaluations; (2) to create a multi-source system that provides 360 degree feedback on the relative effectiveness of the GOS training program; (3) to utilize multiple methods in gathering qualitative and quantitative data from program stakeholders; (4) to identify the Best Practices of the GOS training program; and (5) to provide a standardized process and documentation of the GOS training program effectiveness with respect to each of the NIEHS training program criteria, including overall programmatic effectiveness and impact on the trainees, communities, and the field of environmental work. In this

way, the evaluation process will be generated such that it can be applied to other related health and safety training programs in disaster response.

The proposed evaluation process delineates that training will be systematically evaluated according to a four-level framework that examines training outcomes by measuring:

- Reactions to the training program (Level 1);
- Learning during training (Level 2);
- Behavior following training (Level 3); and
- Results of the training program in meeting intended programmatic outcomes (Level 4).

While most training program evaluations utilize only Level 1 criteria (Alliger et al, 1997; Van Buren & Erskine, 2002) the proposed process suggests that all Levels of evaluation measures should be used to assess training program effectiveness (Sarpy, Chauvin, Hites et al., 2005; Sarpy, Chauvin, & Anderson, 2003).

Level 1 measures are designed to assess reactions to the training program, including individuals' thoughts and feelings about the perceived effectiveness of the instructor, training content, format, and delivery method. Typically, Level 1 measures assess learner satisfaction with various facets of the training. This information is considered a critical factor in trainees' successful completion of the GOS training (Alliger et al, 1997). The measures should also assess perceived relevance of the training, which serves as useful indicators of training-related knowledge acquisition and on-the-job performance (Alliger & Janik, 1989). Therefore, results of the Level 1 measures provide evidence of the impact of training on learner attitudes and perceptions.

Level 2 measures are designed to determine the extent to which learners acquired the principles, facts, techniques, and attitudes stated in the competency-based learning objectives. Written and oral tests, including true/false and multiple choice examinations, exercises, and case studies are often utilized to measure training-related knowledge and skill acquisition. Further, hands-on exercises can serve as useful Level 2 measures. Results of the on-line testing and exercises provide evidence of the impact of training on learner knowledge and skill.

Level 3 measures are designed to assess the extent to which the acquired learning transfers to improved job performance. To assess on-the job behavioral changes, Burke, Sarpy, Tesluk, and Smith-Crowe, (2002) created the General Safety-Performance Survey (GSS) using a confirmatory factor analysis on data gathered from 550 hazardous waste workers. The GSS assesses four broad factors of safety performance including: (1) Using personal protective equipment; (2) Engaging in work practices to reduce risk; (3) Communicating health and safety information; and (4) Exercising employee rights and responsibilities. The GSS is designed as a multi-source evaluation akin to a 360-degree feedback system (i.e., information may be gathered from supervisors, peers, and self-assessments). Both qualitative and quantitative data is captured by the GSS to provide evidence of the impact of

training on job performance.

Level 4 evaluations are designed to assess how the training program contributes to the objectives of the organization. Outcomes include cost reductions as well as increases in service, quality, or workforce retention. This type of evaluation provides information on the impact of training on the organization's bottom line. Information obtained from Level 4 evaluations may be used to address issues such as the relative effectiveness of a web-based training program over a traditional classroom training program and the relative effectiveness of the training program in enhancing the preparedness of the various training cohorts. Once results are obtained from the GSS performance-based (Level 3) evaluations, utility analysis models can be used to assess the relative effectiveness of the Gulf Oil Spill training and provide information for related issues such return on investment in workforce development programs (Sarpy, Smith, Burke, & Anderson, under review).

An important contribution of the framework is that the information obtained from the four levels of evaluation may be integrated to provide an accurate and thorough evaluation of the training as a whole (Sarpy, Chauvin, & Anderson, 2003). For example, for a given training course, results from Level 1 (reactions) and Level 2 (learning) evaluations can be combined with those from Level 3 (on-the-job behaviors) to gain a more comprehensive understanding of training outcomes (Level 4). In this way, information can be synthesized to determine the extent to which the trainees perceived the training as useful, learned the relevant information, transferred this learning to improved performance in the clean-up efforts, and, ultimately, whether the training led to intended outcomes of the training. The integration of information from the evaluation measures and inclusion of Levels 3 and 4 criteria provides critical information for ensuring quality assurance and quality control of the training.

Therefore, an important value-added feature for this proposal is that the Sarpy and Associates evaluation process includes assessment of the impact of the program on worker safety-related performance. Both quantitative and qualitative data will be gathered from trainees and other relevant stakeholders to assess the gains in personal and professional growth that can be attributed to participation in the program. Moreover, we propose the collection of contextual information (i.e., factors that are external to the students, but affect the acquisition and transfer of the training-related information). These contextual factors can either enhance or inhibit the knowledge and performance gains of the individual and have an indirect effect on the impact of the training on the trainees' performance subsequent to training (Burke, Chan, Smith, & Sarpy, 2008; Burke & Sarpy, 2003; Burke, Sarpy & Vaslow, 1998; Ford & Fisher, 1994; Peters, O'Connor, & Eulberg 1985). The results of this analysis can be used to better evaluate the impact of the program on the participants post-training knowledge, skills, behaviors and subsequent health and safety related outcomes.

Collectively, the results of the Sarpy and Associates evaluation process can be used to ensure continuous quality improvement of the NIEHS GOS training program as they strive toward program excellence. These type of training evaluation studies are important for making quality improvements in such health and safety training efforts, identifying best practices that can be used by others, and making better use of available resources and strategies (Robson, Stephenson, Schulte, et al., 2010). In short, the findings from training evaluation studies can be used to strengthen worker health and safety programs and guide related policy development (Sarpy, 2007b; Patton, 2008).

Sarpy and Associates is certified as a Small Woman and/or Minority Owned (SWaM) business by the Virginia Department of Minority Business Enterprise. There are several distinct advantages to partnering with Sarpy and Associates. These benefits include, but are not limited to our professional and experienced team, our approach to partnership, our experience in the industry, our familiarity with health and safety and emergency preparedness and response training programs, familiarity with and research projects that include the communities affected by the Gulf Oil Spill, proven expertise reporting to governmental agencies, excellent reputation in the field, and ability to provide broad conclusions and recommendations spanning across the various training events and communities affected by the Gulf Oil Spill.

References

- Berger, P. & Rice, C. (2000). Improving the Study of the Impact of Worker Health and Safety Training. Silver Spring, MD: NIEHS National Clearinghouse for Worker Education and Training Programs.
- Bogan, C.E., & English, M.J. (1996). Benchmarking for best practices. In R.L. Craig (ED.) The ASTD Training and Development Handbook (Fourth Edition) pp. 394-412.
- Burke MJ, Chan-Serafin S, Salvador R, Smith A, Sarpy S. (2008). The role of national culture and organizational climate in safety training effectiveness. European Journal of Work and Organizational Psychology, 17, 133-154.
- Burke, M.J., & Sarpy, S.A., (2003). Improving worker safety and health through interventions. In D.A. Hofmann & L. E. Tetrick (Eds.), Health and Safety in Organizations: A Multilevel Perspective (pp. 56-90). San Francisco, CA: Jossey-Bass Publishers, Inc.
- Burke, M.J., Sarpy, S.A., Smith-Crowe, K., Chan, S., Islam, G., & Salvador, R. (2006). The relative effectiveness of worker safety and health training methods. American Journal of Public Health, 96(2), 315-24.
- Burke, M.J., Sarpy, S.A., & Vaslow, J. (1998). Evaluating transfer to the job of health and safety training. Annals of Behavioral Medicine, 20(S). 218.
- Cascio, W.F. (2000). Costing Human Resources: The Financial Impact of Behavior in Organizations. Cincinnati, OH: Southwestern Publishing.
- Cascio, W.F., & Boudreau, J. (2008). Investing in People: The Financial Impact of Human Resources Initiatives. Upper Saddle River, NJ: FT Press.
- Colligan, M.J. & Cohen, A. (2004). The role of training promoting workplace safety and health. In J. Barling & M.R. Frone (Eds.), The Psychology of Workplace Safety. (pp. 223-248). Washington, DC: American Psychological Association.
- Creswell, J.W., & Plano-Clark, J.W. (2007). Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage Publications.
- Dannenberg, A.L. & Fowler, C.J. (1998). Evaluation of interventions to prevent injuries: an overview. Injury Prevention, 4, 141-147.
- Fitz-Gibbon, C.T., & Morris, L.L. (1987). How to Design a Program Evaluation. Newbury Park, CA: Sage Publications.

- Ford, J.K., & Fisher, S. (1994). The transfer of safety training in work organizations. A systems perspective to continuous learning. Occupational Medicine: State of the Art Reviews, 9, 241-259.
- Goldenhar, L.M., LaMontagne, A.D., Katz, T., Heaney, C., Landsbergis, P. (2001). The intervention research process in occupational safety and health: An overview from the National Occupational Research Agenda intervention effectiveness research team. Journal of Occupational and Environmental Medicine, 43(7), 616-622.
- Gotsch, A.R., & Weidner, B.L. (1994). Strategies for valuating the effectiveness of training programs. Occupational Medicine: State of the Art Reviews (2), 171-188.
- National Institute of Environmental Health Sciences. (1997). Resource Guide for Evaluating Worker Training: A Focus on Safety and Health. Silver Spring, MD: NIEHS National Clearinghouse for Worker Education and Training Programs.
- Patton, M.Q. (1987). How to Use Qualitative Methods in Evaluation. Newbury Park, CA: Sage Publications.
- Patton, M.Q., (2008). Utilization-focused Evaluation (4th Edition). Los Angeles, CA: Sage Publications.
- Peters, L.H., O'Connor, E.J., & Eulberg, P. (1985). Situational constraints: sources, consequences, and future considerations. Research in Human Performance Resource Mangement, 3, 79-114.
- Prue, C.D., & Feldman, R.H. (1997). Evaluating Health and Safety Training Programs. Silver Spring, MD: NIEHS National Clearinghouse for Worker Education and Training Programs.
- Robson, L., Stephenson, C., Schulte, P., Amick, B., Chan, S., Bielecky, A., Wang, A., Heidotting, T., Irvin, E., Eggerth, D., Peters, R., Clarke, J., Cullen, K., Boldt, L., Rotunda, C., Grubb, P. (2010). A Systematic Review of the Effectiveness of Training and Education for the Protection of Workers. Cincinnati, OH: National Institute for Occupational Safety and Health. DHHS (NIOSH) Publication No. 2010-127.
- Sarpy, S.A. (2007a). Using an Integrated Comprehensive Evaluation Process to Enhance the Effectiveness of Community-based Health and Safety Training Programs. Workshop presented at the National Institute of Environmental Health Sciences Trainer's Exchange. Las Vegas, Nevada, March 2007.
- Sarpy, S.A. (2007b). How do we Know What Training Accomplishes: Methods for Evaluating Health and Safety Training Programs. Workshop presented at the

National Institute of Environmental Health Sciences Trainer's Exchange. Las Vegas, Nevada, March 2007.

- Sarpy, S.A., Chauvin, S.W., & Anderson, A.C. (2003). Evaluation of the effectiveness of the South Central Center for Public Health Preparedness training. Public Health Reports, 118(6), 568-571.
- Sarpy, S.A., Chauvin, S.W., Hites, L.S., Santacaterina, L., Capper, S., Cuccia, M., Anderson, A.C., and Petersen, D. (2005). The South Central Center for Public Health Preparedness training system model: A comprehensive approach. Public Health Reports, Supplement 1, 120-126.
- Sarpy, S.A., Vaughn, K. & Goldstein, N.B. (2010). A Multiple Stakeholder Evaluation of Green Construction Training: Assessing the Effectiveness of Green Courses for Minority Workers. Paper presented at the 138th Annual Meeting and Exposition of the American Public Health Association, Denver, Colorado.
- Sarpy, S.A., & Kaufman, J.D. (2006). A Longitudinal Evaluation of the Effectiveness of Minority Worker Health and Safety Training Programs in Eight Communities. Paper presented at the National Occupational Research Agenda Symposium: NORA Makes a Difference of the Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Washington, D.C.
- Sarpy, S.A., & Kaufman, J.D. (2005). Integrated Multiple Stakeholder Evaluation Process: Using Qualitative and Quantitative Data to Evaluate Safety and Health Training Interventions. Paper presented at the 133rd Annual Meeting and Exposition of the American Public Health Association, Philadelphia, Pennsylvania.
- Sarpy, S.A., Kaufman, J.D., & Smith-Crowe, K. (2002). An Integrated Multiple Stakeholder Evaluation of the Brownfield Minority Worker Training and Minority Worker Training Programs for the Center to Protect Workers' Rights. Technical report prepared for the National Institute of Environmental Health Sciences.
- Sarpy, S.A., Smith, A.N., Burke, M.J., & Anderson, A.C. (under review). Evaluating the impact of emergency preparedness and response training: Means for addressing accountability and continuous improvement issues. Public Health Reports (manuscript under review).
- Shannon, H.S., Robson, L.S, & Guastello, S.J. (1999). Methodological criteria for evaluation occupational safety intervention research. Safety Science, 31, 161-179.

The specific tasks, deliverables, and timeline associated with the proposed evaluation project are detailed in Table 1.

Table 1. Tentative Project Outline.

Develop Methodology and Design

Review extant GOS and emergency response evaluations	02/01 – 02/14
Identify relevant stakeholders for inclusion	02/01 – 02/14
Construct sampling frame for study	02/14 – 02/28
Select study participants (relevant stakeholders)	02/14 – 04/30
Recruit study participants (relevant stakeholders)	02/14 – 04/30
Identify evaluation methods for relevant stakeholders	02/14 – 02/28
Identify relevant control groups and cohort groupings (e.g., language, location, extent of training)	02/14 – 02/28

Develop Evaluation Measures

Review extant GOS training materials and evaluation measures	03/01 – 03/15
Construct preliminary evaluation instruments for relevant stakeholders	03/15 – 03/31
Review preliminary evaluation instruments and protocol with subject matter experts	04/00 – 04/15
Revise as needed to create final evaluation instruments in various formats (e.g., on-line questionnaire, structured interview protocol)	04/15 – 04/30

Implement Evaluation Process

Identify and train evaluators according to evaluation protocol	05/01 – 05/15
Assist with focus groups and structured interviews of relevant stakeholders	05/15 – 05/22
Administer questionnaires to relevant stakeholders	05/01 – 05/22
Recalibrate implementation process as needed	05/15 – 05/22
Follow-up administration of measures as needed for non-respondents	05/22 – 05/31

Generate Database and Enter Data

Create database for evaluation measures	06/01 – 06/15
Code data according to relevant evaluation questions	06/15 – 06/30
Enter, verify, and clean database for analysis	07/01 - 07/15

Conduct Statistical Analysis

Conduct content analysis on qualitative data	07/15 – 07/31
Conduct descriptive analysis on quantitative data	08/01 – 08/15
Conduct comparative analyses on quantitative data regarding various training methods and cohorts regarding	08/15 – 08/30

Generate Final Report

Write preliminary report that synthesizes information from the multiple stakeholder evaluation	09/01 – 09/20
Provide preliminary feedback to selected stakeholders	09/20 – 09/31
Write and submit final report in written and electronic formats	10/01 - 10/31