An Online Strategy for Evaluating Educational and Social Intervention Programs

Hansen, William B.; Reese, Eric; Wyrick, Cheryl H.; Jackson-Newsom, Julia; Bryant, Kelvin S.; and Dyreng, Douglas I.

Abstract— *Evaluation Lizard is a web application designed to provide decision making support and operational capacity for completing evaluations of educational and social interventions. Users complete essential evaluation design tasks online. These include defining treatment and comparison groups, defining treatment and survey events, and creating surveys. The application allows users to administer surveys online or to prepare specialized scannable paper-and-pencil surveys. The application allows users to manage an evaluation team (research assistants and data collectors) and the client organization (staff administering the intervention and participants). The application allows users to download their data and retrieve basic analysis reports.*

Index Terms—*evaluation, online tools, survey*

1. INTRODUCTION

THIS paper addresses the development of a web-based application for supporting the completing the evaluation of social interventions. The web application described, Evaluation Lizard, was developed specifically for use in the evaluation of drug abuse prevention programs, but can be used to assist in the evaluation of a broad array of social interventions.

2. PROBLEM STATEMENT

There has been a discernable shift in public policy throughout the United States and Europe toward the evaluation of programming designed to affect the health and well-being of their citizens. For example, in Europe, the European Monitoring Centre for Drugs and Drug Addiction

[1] has set forth principles that promote the adoption and evaluation of research-based drug abuse prevention efforts in schools. As noted in a handbook for evaluation produced by EMCDDA [2],

"In the last few years, a growing number of prevention activities have been carried out in all the member states of the European Union (EU). Most of the projects, however, haven't been effectively evaluated, and so there is an urgent need for improving the knowledge about the process of 'prevention evaluation' and exchanging experiences and results."

Similarly, the US Department of Education Safe and Drug-free Schools and Communities program now includes four principles of effectiveness [3]. These principles state that a grant recipient shall:

(1) base its program on a thorough assessment of the objective data about the drug and violence problems in the schools and communities served; (2) establish a set of measurable goals and objectives and design its activities to meet those goals and objectives; (3) design and implement its activities based on research or evaluation that provides evidence that the strategies used prevent or reduce drug use, violence, and disruptive behavior; and (4) evaluate its program periodically to assess its progress toward achieving its goals and objectives and use its evaluation results to refine, improve, and strengthen its program and to refine its goals and objectives as appropriate.

The emphasis that emerges from these principles is one that reinforces a value on research-based programs and evaluation. Two of the four principles call for data collection – needs assessment and local evaluation.

3. THE LIMITS OF EXISTING APPROACHES TO EVALUATION

Requirements to complete local evaluations place a new burden on schools and communities who are not experienced with either needs assessment or evaluation methods. This is especially true when non-academic programs, such as drug or violence prevention programs,

are introduced. Many school districts and community groups feel overwhelmed with these new requirements. In short, they are being required to complete an activity for which they are unprepared and for which they have few options or tools. Hiring an evaluation specialist to design and complete a needs assessment or an evaluation is often costly, typically beyond the budget of most small projects. Even when hired, the standardization of measures and methods have not been established. Such conditions often lead to sub-standard evaluation projects. These conditions make cross-project comparisons nearly impossible to complete.

4. THE SOLUTION PROVIDED BY EVALUATION LIZARD

Evaluation Lizard was initiated to be a low-cost, easy-to-use alternative to hiring a professional evaluator. The rationale for this approach rests of several research achievements in the field of prevention. These include:

(1) The development of standardized measures. As with any field of science, methods for assessing substance use behaviors and correlates of behaviors (e.g., attitudes and beliefs) were initially idiosyncratic and rarely shared by individuals and research groups investigating drug use. However, the field has increasingly moved towards standardized measures. In the United States, the Core Measures Initiative of the Center for Substance Abuse Prevention (CSAP) [4] has resulted in the compilation and publication of many relevant measures. Similarly, in the European Union, the EMCDDA has created an Evaluation Instruments Bank [5] that includes survey items in many languages.

(2) The development of a list of recognized programs. The ultimate result of this funding on prevention research in the United States has been the creation of a National Registry of Effective Programs and Practices (NREPP) [6]. The goal of NREPP is to identify drug abuse prevention and treatment programs and mental health promotion programs with proven effectiveness that can be disseminated broadly to schools and community service organizations. During the past several years, over 900 programs have been nominated for inclusion in the list of proven programs. As of January 1, 2004, 153 had been judged to have evidence of effectiveness and were included in the registry. North American program developers continue to search for effective strategies with the goal of reducing drug use.

(3) A by-product of research has been the systematization of social science evaluation designs [7]. Social scientists pride themselves on their ability to create evaluation designs that allow studies to interpret findings with validity – that is, to be able to demonstrate reductions in drug use that can scientifically be attributed to the effect of the program and not the effect of extraneous factors. The field has generally progressed to the point where there are now a handful of recognized evaluation designs. Thus, the structure of evaluations has also become simplified.

The cumulative effect of this progress allows evaluation designs to be specify criteria that can be used to evaluate known programs using standardized measures in local settings. The goal of Evaluation Lizard was to create an online system that allowed the simplified evaluation of local implementations of research-based drug abuse prevention programs.

This system was designed for use by local evaluators who are expected to continue to conduct evaluations according to acceptable practice, including protecting the rights of subjects as well as interpreting findings. The best use of this application currently is to assist evaluators conduct local evaluations of previously proven programs.

5. THE SOLUTION PROVIDED BY EVALUATION LIZARD

Designed to support schools and community groups who wish to complete outcome evaluations of drug prevention programs, Evaluation Lizard [8] facilitates the the following evaluation tasks: (1) create an evaluation project, (2) describe and define an intervention that will be evaluated, (3) define an evaluation design consisting of evaluation conditions, measurement events, and treatment events, (4) select outcomes measures including demographics, measures of drug use, and measures of mediating variables targeted by the intervention, (5) specify informed consent procedures, (6) collect data either online or using paper-and-pencil surveys, (7) create a database of survey data, and (8) complete basic descriptive analysis of the data. Our vision of this web application is that it will not only serve the short-term needs for evaluation, but that it will also a searchable library of measures, a repository of data from research and evaluation projects, and an exceptionally large database that can be mined for information.

Organized around primary functions associated with evaluating social interventions, the website's core competencies include: (1) managing projects, (2) managing people associated with projects, and (3) completing analysis and getting reports. The library of measures is a feature associated with the Evaluation Lizard that allows researchers and evaluators to find survey research measures from a large bank of measures. A help feature is also included to guide users through the website. Each of these functions is described in the section that follows.

Evaluation Lizard

Log In/Out

Manage Projects

Manage People

Analysis and Reports

Library of Measures

Help

5.1. Managing Projects

The first step in Evaluation Lizard is giving a new project an appropriate name and providing descriptive information useful to the Project Coordinator and others who are participating in the project. . Crucial to subsequent definitions is for the Project Coordinator to specify the type of organization in which the evaluation will be conducted. Evaluation Lizard selects educational organizations as the default organization type. Other organization types currently available include ecclesiastical and community.

To define a project, three elements must be specified: (1) the program being evaluated, (2) an evaluation design to be applied to the project, and (3) the measures that will be used in surveys that are administered to participants in the project.

5.2. Prevention Programs

A meta-analysis of NREPP prevention programs was recently completed by members of our team [9] that has allowed us to systematically identify important characteristics of research-based prevention programs. This meta-analysis evaluated the program content of 48 programs that specifically addressed substance abuse prevention and that had published program implementation manuals. Program content refers to the specific kinds of information presented in the program. Most often the program content of educational approaches focus on change personal management skills (such as decision making and goal setting), social skills (such as assertiveness and resisting peer pressure), or disposition (such as attitudes towards drugs,

commitment to avoid drug use, and developing positive social norms that promote the non-use of drugs). Thus, the specific program content of each program is profiled. Once users select a program for inclusion in their project, the profile of that program is identified.

5.3. Measurement Library

We have assembled a library of measures that currently includes measures from four domains, demographics (gender, age, and ethnicity), behaviors (use of alcohol, tobacco, marijuana, inhalants, cocaine, heroin, etc.), mediators (measures that assess psychological changes targeted by program content such as skills, attitudes, and beliefs), and moderators (measures that assess participation in a program). Measures come from a compendium of survey research items developed by researchers in the field of drug prevention. Items are grouped into scales that are hierarchically categorized for ease of use by users. Thus, within each domain, items are grouped into classes, categories, and scales. For example, a measure that assesses students' beliefs about the consequences of drinking alcohol will be categorized with similar items that allow the evaluator to find items of interest.

5.4. Survey Design

Evaluation Lizard uses program content to define the appropriate components of surveys. Thus, we use NREPP programs' profiles to define what items should be included in surveys. Surveys for programs that teach students about decision making include items to assess their decision making ability and ability to control impulsivity. Surveys for programs that promote the development of personal commitments to avoid drug use assess students' intentions to use substances. Programs that focus on alcohol use include survey items for students to self-report about their alcohol consumption and drunkenness. We have created standardized surveys for each of the substance abuse prevention programs on the NREPP list. Thus, for any of the 48 programs for which we have been able to document program content [9], each has a content-appropriate, ready-to-use survey available in Evaluation Lizard.

5.5. Evaluation Design

Evaluation Lizard includes an application that allows evaluators great flexibility in terms of specifying evaluation designs. However, to simplify its use, a set of prescribed templates are available for the evaluator to choose from that automatically define basic treatment group and measurement options. These prescribed designs include a needs assessment, a pretest-posttest

treatment group only design, a posttest only treatment and control group design, a pretest-posttest treatment and control group design, a pretest-posttest multiple group design, and a lagged-cohort treatment and control group design.

Design Template

Template	Design
○ Assessment	Group 1 O_1 $[O_2 O_3 O_4 O_5 \ldots]$
○ Pretest-Posttest Single Group	Group 1 $O_1 X_1 O_2 [O_3 O_4 O_5 \ldots]$
○ Posttest Only Control Group	Group 1 $X_1 O_1 [O_2 O_3 O_4 \ldots]$ Group 2 $O_1 [O_2 O_3 O_4 \ldots]$
● Pretest-Posttest Control Group	Group 1 $O_1 X_1 O_2 [O_3 O_4 O_5 \ldots]$ Group 2 $O_1 \quad O_2 [O_3 O_4 O_5 \ldots]$
○ Pretest-Posttest Multiple Comparison Group	Group 1 $O_1 X_1 O_2 [O_3 O_4 O_5 \ldots]$ Group 2 $O_1 X_2 O_2 [O_3 O_4 O_5 \ldots]$ Group 3 $O_1 X_3 O_2 [O_3 O_4 O_5 \ldots]$
○ Lagged Cohort Comparison	Group 1 $O_1 X_1 O_2 [O_3 O_4 O_5 \ldots]$ Group 2 $O_2 \quad O_3 [O_4 O_5 O_6 \ldots]$

Users can modify the number of elements in their design. They can modify the number of treatment and control conditions as well as the number of testing events. They can also order treatment and testing events. For instance, Evaluation Lizard will allow the user to create a design in which there was a pretest, a treatment intervention, a posttest, a second treatment intervention, and additional posttests. Evaluation Lizard tracks all events in a design.

| Back | Continue | | Modify Design... | Reset Design |

Design Review

Miscellaneous Project Parameters	
Condition Assignment Level	Section
Condition Randomization	Manual
Subjects' Informed Consent Method	None

Condition	Event 1	Event 2	Event 3
Group 1	Pretest 1	Treatment 1	Posttest 1
Group 2	Pretest 1	{No Treatment}	Posttest 1

When specifying the design, users also define three important features of the design. First, users define the level at which condition is assigned. Users define this using the organization type specified. Users define whether and how units will be randomized. There may either be no randomization, randomization controlled and reported by the user, or randomization automated by Evaluation Lizard. Finally, users define informed consent procedures to be followed. Options for the latter include none (e.g., no informed consent required), passive (e.g., subjects are assumed to be included unless specifically excluded) or active (e.g., subjects are excluded until permission to participate has been granted).

As a final feature, the time frame for completion of each event can be specified. Time frame is useful for limiting access to online surveys should the evaluator wish to specify this. Time frame is also important for defining when the database will be made publicly available outside the project for analysis. (We plan to embargo data for five years following the completion of a project. After this date, elements of the database stripped of identifiers will become available for public use.)

5.6. Managing People

Evaluations involve people. Evaluation Lizard facilitates the management of people from three different groups. The first group is the project staff who comprise the evaluation team.

The Project Coordinator for each project can give privileges to other members of the evaluation team. Privileges include: modifying personal information and privileges of other evaluation team members, modifying information about the client organization, modifying the project (e.g., selecting programs to be evaluated, modifying evaluation designs, selecting measures to be included in surveys), assigning individuals or groups to experimental conditions, requesting data analysis, having access to raw data files, adding and modifying participant information, viewing data collection status reports, and initiating the printing of paper and pencil surveys.

The second group is the client organization – the organization where the evaluation is taking place, such as a school system. The client organization is described in hierarchical terms that include a geo-coding of people's roles. Thus, for an evaluation conducted with clients who are in a public school educational setting, people are assigned roles for the nation, region, state, county, city, district, school, teacher, and section.

The third group of people is the evaluation participants – subjects who complete surveys and receive interventions. One of the features added to the application that is important in conducting evaluations in schools is the ability to associate participants with classes. In longitudinal evaluations, students may change which classes and which schools they are enrolled in. Evaluation Lizard provides a tool for associating students (or other types of

participants) with the group in which the survey will be administered.

Evaluation Lizard includes online tools for managing all three groups. Evaluation team members and members of the client organization can be assigned privileges crucial to completing evaluation tasks. Tools for importing or entering participant information and assigning participants to conditions are provided.

5.6. Survey Administration

Evaluation Lizard has two strategies for administering surveys. The original intent was to administer surveys online. Online surveys are not new [10] and Evaluation Lizard follows commonly employed strategies. Students use a web browser to navigate to a URL provided to survey administrators. Students login with a login id and password, both of which are provided to the administrator Students then complete the survey. Data are tied to student login information.

The second strategy for administering surveys was developed in response to requests from schools. Scannable surveys are printed. Initial surveys are created using a word processing system and then printed to Adobe Acrobat® PDF files. These surveys are further processed so that coded information is individualized for every page of every survey, allowing its source to be traced. Thus, survey pages have bar coded identifiers (but no human readable text) printed at the bottom of each page to identify the survey once it is scanned. Printed surveys have tear off cover sheets that have participant information to ensure it gets to the correct person. The local team is responsible for survey administration and interpretation of findings.

5.7. Data Analysis and Reports

Evaluation Lizard allows evaluators to complete basic analysis tasks common to evaluation. Three tasks may be completed by evaluators: (1) generate survey status reports, (2) complete basic analyses, and (3) download raw data for local analysis.

Evaluation Lizard tracks survey participants across events and can provide evaluators with dynamic information about survey completion. This report allows administrators to identify missing subjects, which is particularly valuable for projects that wish to be diligent in getting as much participation as they planned.

Evaluation Lizard automatically generates descriptive analyses for all items included on the survey. The application generates frequency tables. All events completed to date are broken down by event (e.g., pretest, posttest, etc.) and condition (e.g., treatment versus control). Tables include frequencies for each response as well as percentages of individuals within-condition who responded. For behavior items and for mediating variable scales, Evaluation Lizard also produces bar charts that compare conditions at each testing event as in the sample below. Reports are printed as Adobe Acrobat® PDF files.

Evaluators can download the raw data from their projects for further analysis. For example, evaluators will likely want to complete inferential statistical analyses on their own. When evaluators download data from Evaluation Lizard, data are stripped of positive identifiers. Thus, the data cannot be traced back to the individuals who completed the surveys. Subjects are instead linked by an ID number generated by Evaluation Lizard. This allows confidentiality to be maintained.

6. ANALYSIS OF THE CAPABILITY OF THE EVALUATION LIZARD APPLICATION

6.1. Functionality of the Application

Evaluation Lizard is a complex relational database-driven application, consisting of numerous condition-driven relationships. Originally, it was hoped that the application would be usable by untrained evaluators. However, because of its complexity, the use of many elements of the application continues to be reserved for use by trained evaluators. During the initial phases of its formal launch, a quality control specialist documented non-functioning and error-producing elements of the application. These errors have gradually been cleared and, as of this writing, few known programming errors exist. The greater need seen with the application in its current state is increased simplification of tasks and the adoption of additional pre-coding wherever possible. For example, the geo-coding of states has been pre-loaded so that the user can select in which states in the US the evaluation will occur. Region and nation then follow automatically. Additional pre-coding for other areas of the world, and pre-coding of counties and cities would significantly aid the speed with which client organizations can be entered. The system has not yet been tested to verify compliance with Section 508 of 36 CFR Part 1194.

6.2. Accumulation of Survey Items

As of January, 2005, there were 2,481 survey items in the measurement library. Most items do not stand alone, but are part of a scale, of which there were 436. The library contains 814 questions that assess behavior (such as alcohol consumption), 64 demographic items, and 1,603 mediating variable items which form 95 different scales. The measurement library is now continually expanding with new measures being

added as they are developed, discovered, or required by users.

6.3. Use of the Application

During the first year of its operation as a commercial service in 2004, 7,540 surveys were administered using Evaluation Lizard. In 2004, 25 organizations ordered surveys. In the first nine months of 2005, 10,803 surveys have been administered through Evaluation Lizard, roughly doubling the rate of use from the first year. To date in 2005, 55 agencies have purchased surveys. Every agency that purchased surveys during 2004 returned to purchase surveys in 2005, indicating a satisfaction with the service. The system has been used primarily to evaluate a select number of SAMHSA drug prevention programs: e.g., All Stars, Project Northland, and Creating Lasting Family Connections.

6.4. Cost

Surveys are billed at a rate of $2 per printed survey and $.50 for online surveys for standard surveys. This includes all costs except shipping and specialized report writing. Thus, survey design, evaluation design, printing or the creation of online surveys, data reduction (scanning for printed surveys), database creation, and the creation of basic analyses and reports are all provided within this fee. This is a significant savings compared to the costs of a typical evaluation consultant who will often have costs exceeding $10 per administered survey.

To date, few customers have opted for online surveys. Availability of computers in schools, bandwidth, and control when administering surveys are thought to account for customers' preference for paper-and-pencil versions. This situation is expected to be improved in the future.

7. CONCLUSION

Evaluation Lizard is a web application that can make the local evaluation of drug prevention programs affordable and easy to complete. The system takes advantage of the state of the art in prevention evaluation practice, eliminating many barriers schools and community groups might otherwise encounter. Novice evaluators with little experience can produce effective evaluations using Evaluation Lizard. In addition, there is enough flexibility for seasoned researchers to build custom evaluations.

Evaluation Lizard has been a collaborative effort between social scientists – who have extensive training and practical experience in evaluation design – and a team with expertise in information technology. This marriage has been unusual and revealing in that complex tasks typically done by hand by social scientists have been translated into algorithm- and database-driven procedures design by computer programmers. The experience of producing this application has revealed the degree to which human-based decisions are often complex and idiosyncratic. Finding strategies to simplify and systematize decisions about how to conduct evaluation designs required extensive effort and a willingness to adapt approaches by both teams.

Evaluation Lizard, while overcoming many procedural problems, is not a simple guaranteed solution for the difficult and complex evaluation process. In the future this tool (or others like it) need to be developed to allow users to do such things as warehouse data collected using other methods of data collection and to mine data in various ways. For example, most local evaluation as actually implemented by teams often lack the rigor called for in order to have high degrees of internal and statistical validity [7]. New methods for creating synthetic control groups and for combining multiple evaluations through meta-analytic or other techniques need to be developed. In many respects, it is only after a large repository of data exists that such developments will be possible. Thus, even though Evaluation Lizard cannot yet address many of the challenges that accompany evaluating social interventions, this web application provides an opportunity to make progress towards addressing these issues.

REFERENCES

[1] G. Burkhart, M. Nilson, and D. Ballotta. "Drug prevention in EU schools: Information and reporting systems are crucial," Drugs in Focus. September-October 2002, http://www.emcdda.eu.int/index.cfm?fuseaction=public.AttachmentDownload&nNodeID=526.

[2] C. Kröger, H. Winter, and R. Shaw. "Guidelines for the Evaluation of Drug Prevention Intervention: A Manual for Programme Planners and Evaluators," Institut fur Therapieforschung: Munich, Germany. 2000, http://pl-www.emcdda.eu.int/?fuseaction=public.AttachmentDownload&nNodeID=1753.

[3] US Department of Education. "Safe and Drug-Free Schools Program; Notice." *Federal Register.* 63(104), pp. 29902-29906, June 1, 1998.

[4] CSAP, "Core Measures Initiative Phase I Recommendations," January, 2003, http://alt.samhsa.gov/grants/2004/downloads/CSAP_Core_Measures.pdf.

[5] EMCDDA, "Evaluation Instruments Bank," 2005, http://eib.emcdda.eu.int/.

[6] http://modelprograms.samhsa.gov.

[7] W. R. Shadish, L. C. Leviton, and T. Cook. *Foundations of Program Evaluation.* Thousand Oaks, CA: Sage Publications Inc, 1990.

[8] http://lizard.tanglewood.net/default.asp.

[9] W. B. Hansen, J. H. Derzon, L. Dusenbury, D. Bishop, K. Campbell, and A. Alford, "Analysis of Magnitude of Effects of Substance Abuse Prevention Programs Included in the National Registry of Effective Programs Through 2003: a Core Components Analysis," Report submitted to the National Center for the Advancement of Prevention. Tanglewood Research, Greensboro, NC, USA, June 2004.

[10] S. McCoy, and P. V. Marks. "Using electronic surveys to collect data: Experiences from the field," 2001. Seventh Americas Conference on Information Systems. 2001, pp. 1502-1505.