CONDUCTING A SOCIAL POLICY EXPERIMENT IN AN INTERNATIONAL SETTING: RESEARCH DESIGN AND IMPLEMENTATION ISSUES

Linda Young, Knowlton W. Johnson, Denise Bryant, Community Systems Research Institute, Inc.
Linda Young, CSRI, 1300 Fourth, Suite 300, Louisville, KY 40208

INTRODUCTION AND BACKGROUND

For decades social scientists have grappled with research design and implementation issues inherent to conducting cross-cultural research (e.g., Adler, 1982, Brislan, 1973, 1986; Bulmer, 1998; Deutcher, 1973; Fontes, 1998; Hantrais & Mangen, 1996; Peil, 1993). Recently Western methodologies have been increasingly exported and adapted for use in developing countries. Social scientists have begun to examine Western assumptions that underlie the application of these methodologies (Alers, 1993; Bulmer & Warwick 1983, 1993; Fontes, 1998; Kuechler, 1998; Mitchell, 1993; Newby, Amin, Diamond & Naved, 1998; Warwick, 1993).

As social policies and programs cross international borders, researchers will confront challenges in designing, implementing and evaluating social programs. Fontes (1998) suggests that whenever research investigators are from a different culture or country than that of the research subjects, they should give serious consideration to important lessons learned from cross-cultural, and comparative research.

The social policy experimental design can pose unique challenges for social scientists even in their native countries. Berk, Brouch, Chambers, Rossi and Witte define social policy experiment as "a theoretically based effort to introduce social change in a way that permits discerning the net effect of the change on social outcomes" (1985, p.407).

This paper examines inherent challenges for U.S.-based researchers in conducting a social policy experiment in another country, specifically Peru. We also present lessons learned during the implementation process that have implications for similar studies. Using the social policy experiment in Peru as a case study, the paper presents relevant research design and implementation experiences that focus on four central issues: (1) establishing partnerships with key stakeholders in the host country, (2) developing and translating research instruments, (3) implementing a true experimental design with repeated measures, and (4) ethical considerations for conducting transnational research.

The social policy experiment under discussion was initiated by the U.S. State Department Bureau of International Narcotics and Law Enforcement (INL) to test the effects of two Therapeutic Communities (TC) training courses that target staff and organizations providing substance abuse services. INL frequently receives requests from other countries for training on the therapeutic communities substance abuse treatment model. Peru was interested in implementing alternative TC training models under experimental conditions, building on its past efforts to reduce drug abuse, and develop infrastructure and expertise in substance abuse treatment. INL supported the initiative and its careful evaluation.

INL selected Daytop International to conduct their Therapeutic Communities (TC) training curriculum in Peru. Daytop's TC treatment model incorporates five dimensions, behavior, emotional, spiritual, intellectual, and vocational aspects of substance abuse treatment (Biase, 1985). The training provides didactic instructions and exercises to practice what is being learned. The TC training course consists of a core curriculum in which trainees are taught how to use TC concepts and tools during six, 1-week training modules. The extended TC training course includes the six core curriculum training modules plus two additional 1-week modules that focus on methods for managing organizational change.

Methodology

The evaluation of the TC training courses includes both process and outcome evaluations. Seventy-four organizations that provide substance abuse treatment or early intervention services in Peru participated in the TC training and evaluation. A randomized block design with repeated measures was used, and the organizations were assigned to three experimental conditions: Groups A and B received the six weeks of training on the core TC curriculum modules; Group A also received the two additional training modules on managing organizational change; and Group C served as the control group and received delayed core curriculum training approximately 16 months later. A total of 234 staff representing the 74 organizations participated in the experiment. Twelve separate data collection instruments were developed and translated into Spanish. Outcome data were collected via in-person interviews and a short self-administered questionnaire at three points in time: (1) before the core training curriculum began (Wave 1), (2) immediately after all of the core training modules were completed (Wave 2), and (3) approximately 15 months after the wave one data were collected (Wave 3). Process data were collected via a self-administered questionnaire completed by participants on the last day of each of the eight training modules. The evaluation will be concluded in 1999.

ESTABLISHING PARTNERSHIPS
Creating Collaborative Relationships

The most consistent finding of previous studies of research use and planned change is that the key actors of a social experiment must be involved at every stage of the research process (Johnson, 1980; Johnson, 1990; Johnson, et. al., 1996). Cross-cultural researchers have also emphasized the importance of collaboration with host-country scholars and experts to provide a critical review of the appropriateness of the proposed methodology and to enhance the scientific rigor of the project (e.g., Bamberger, 1999; Bulmer & Warwick, 1993; Fontes, 1998; Munroe & Munroe, 1986; Newby, et al., 1998; Peil, 1993; Schooler, et al., 1998).

Involvement of key players in the host country is important for obtaining legitimation and endorsement and for gaining an adequate understanding of the social, historical, and political context of the experiment. Buckley (1998, p.227) warns that Western social scientists may be perceived as "ethnomethodologists" when designing and implementing studies in another country. She advises researchers to question their assumptions and to avoid imposing views on design and implementation issues that do not take into account different cultural perspectives. In this regard, the TC Training Experiment was designed as a collaborative partnership between the evaluators, the trainers, and key representatives of the TC community and data collection firm in Peru.

Putting Together the Evaluation Team

The partnership consisted of the trainers, the evaluation team, and an on-site Peruvian project coordinator and assistant. The evaluation team selected a research firm in Lima with experienced staff in substance abuse research to assist in project planning and to manage and implement the data collection design.

The conceptual framework was developed in collaboration with the Daytop International training team using a logic modeling approach to design the intervention and the evaluation methods. The Peruvians provided input in developing the hypotheses and research questions, the research design and interpretation of findings. They reviewed the instrument translations, assisted us in developing data collection strategies, and helped communicate with key officials in the Peruvian government and substance abuse treatment community.

Our Peruvian partners helped us identify data elements that could be consistently collected. With their guidance, we modified some of our original evaluation protocols and adopted more realistic procedures for collecting organizational data. We were advised that some key variables (i.e., number of admissions and discharges for each organization) were not available. As a result, we trained TC directors to record and maintain client information so that organizational data could be consistently collected. We were also advised that a portion of the TC organizations did not have telephones and in-person interviews were recommended instead of telephone interviews. We followed their advice.

Understanding the Political, Historical and Economic Context of the Experiment

Cross-cultural and transnational researchers emphasize the importance of understanding the political, historical and social environment of the local community within the host country where research is to be conducted (e.g., Buckley, 1998; Bulmer, 1998; Bulmer & Warwick, 1983, 1993; Fontes, 1998; Newby et al., 1998; Peil, 1983, 1993). Researchers must be aware of the political implications of the project to host country constituents and understand local sensitivities and potential conflicts that can threaten the project or place research subjects at risk (e.g., Buckley, 1998; Fontes, 1998; Newby et al., 1998; Peil, 1983, 1993).

During initial planning meetings held in Peru, the project team learned important information about the history, social, economic and political realities of substance abuse in Peru that had significant implications for the project. Residential and non-residential treatment and early intervention organizations in Peru are clearly much more diverse than those in the U.S. Our Peruvian partners advised us that the conditions and practices at some organizations increasingly concerned the Peruvian government. Recent incidents of violence had provoked the Ministry of Health to move toward regulating residential and non-residential treatment organizations and establishing a process of formal certification. Less than one-fourth of the 100 or so treatment or early intervention organizations were certified by the Peruvian Ministry of Health. The TC training was perceived as an important step in the direction of regulation and formal certification. Some organizational directors were concerned that the training might ultimately affect their ability to operate a treatment organization. Within the substance abuse treatment community, there was disagreement about who should attend the training and whether organizations that were suspected of human rights violations should be invited.

Some directors of organizations assigned to the control group feared that the delayed training design would place them at a political disadvantage; a few threatened not to participate in the TC Training Experiment. Other directors were concerned about how the evaluation results would be used and whether the evaluation data would be kept confidential.

INSTRUMENT DEVELOPMENT

Specific protocols for developing culturally sensitive instruments that are conceptually and linguistically equivalent are widely documented (e.g., Brislan & Thorndike, 1973; Brislan, 1986; Bulmer & Warwick, 1993; Deutcher, 1973; Herrera et al., 1993; Jowell, 1998).

The measures required for our outcome evaluation corresponded to specific objectives for mastery of the principles, practices, competencies and behavior management models presented in the training. Consequently, most of the questionnaire items were generated exclusively for this study and were not initially intended for cross-cultural comparison. Issues related to conceptual and linguistic equivalence were encountered primarily during the translation process.

Translation of Instruments

The development and translation of data collection instruments requires a rigorous process of examining the concepts, language and data collection methods to ensure measurement equivalence across the culture of the investigator and that of the research subjects (Bulmer & Warwick, 1983, 1993). When an English language questionnaire developed for English-speakers is literally translated into another language, the results can be disastrous (e.g., Brislan, 1986; Herrera et. al., 1993; Jobert, 1996, and Jowell, 1998). Words can have many meanings and nuances, and translators must fully understand the intended meaning of words and phrases in order to select the most accurate translation.

Brislan (1986) and others also caution against using American colloquialisms in research instruments. For example, during the development of the Spanish version of the instruments, the translators were confused by the phrase *going through the motions.*

Some questions used terms referring to TC behavior management concepts such as *haircuts* and *pull-ups* that could not be literally translated. For these questions, a parenthetical explanation of the meaning was included in the wording of the item.

Researchers sometimes lack demographic knowledge and understanding about the target population in the host country (e.g., Jobert, 1996; Fontes, 1998). For example, during the TC Training Experiment, we learned that the standard categories typically used for questions about income and about racial or ethnic background were inappropriate. We were advised by our Peruvian evaluation team members that the typical respondent for our study could be expected to earn less than $6,000 per year. The Peruvian team members advised us to use the categories, "White, Mestizo, Indigenous Indian, Asian, Black and Multi-racial" rather than the standard race and ethnicity categories typically used in the U.S.

Conceptual and linguistic equivalence in data collection instruments can be achieved through a process of negotiation between the researchers, host country collaborators and professional, native speaker bilingual translators. The following translation process was used for the Peruvian study.

(1) The English versions of the instruments were pre-tested on 15 individuals representing a cross section of staff members in various types of substance abuse treatment facilities in Louisville, KY. Problematic questions, ambiguous or awkward terms or concepts were revealed through this process.

(2) The English versions of the instruments were then translated into Spanish by a native Peruvian professional translator.

(3) The Spanish versions of the instruments were reviewed by Peruvian bi-lingual members of the evaluation team who were familiar with how the training concepts were operationalized into measures. They identified words and phrases for which there was disagreement on the translation and consulted with the evaluation team until an agreement was reached.

(4) The translated instruments were then pre-tested with members of four institutions that were not part of the evaluation.

(5) A Spanish bi-lingual member of the American team reviewed the final Spanish version of all instruments and back-translated them into English.

We learned to carefully examine conceptual equivalence issues during the measurement development stage of the project. For example, we should have questioned whether concepts used in measures of TC knowledge and behavior management practices such as *privilege systems, individuality, role modeling,* and *self-help,* mean the same in the Peruvian culture as in the United States. We are confident that for the most part, our translation process addressed most of the conceptual equivalence problems and resulted in Spanish instruments equivalent to our English instruments. However, we are not certain that all of the concepts used in our measurement items capture all of the cultural nuances that may be unique to Peruvians.

IMPLEMENTING AN EXPERIMENTAL DESIGN

When research is implemented in an international setting, the underlying assumptions and associated method and protocols of research design must be examined to a greater extent (Bamberg, 1999; Peil, 1983, 1993). The evaluation design required randomized assignment of the 74 organizations to three experimental conditions. The target population were to be identified by the organizations' directors who were instructed to apply specific selection criteria. The following are some of the difficulties the evaluation team encountered and the lessons we learned in implementing the TC Training Experiment in Peru.

Random Assignment of Organizations

A system for classifying organizations on several different dimensions was developed through a collaborative process that included representatives from key organizations in Peru. The classification system grouped the organizations according to three main dimensions: (a) the degree of formal authorization by the Peruvian government's Ministry of Health, (b) the functional definition of the target population served by the institution, i.e., street kids, prison, outreach, etc. and

(c) the perceived level and adequacy of the administrative structure of the organization with regard to staffing, treatment modality, facility, skill development opportunities, and the perceived risk for residents with respect to perceived potential for violation of human rights. The classification system allowed us to stratify the organizations in a way that created three fairly equivalent groups.

Each of the three groups included one organization with multiple facilities. These organizations were allowed to send more than three participants to the training. Specific decision rules based on the administrative and programmatic structure of the multiple-site organizations were applied to obtain in consistent representation from the multiple-site facilities while holding the three groups to a comparable size.

Characteristics of the Target Organizations

Substance abuse treatment and early intervention organizations in Peru are diverse. Some organizations are very similar to those in the U.S.; and others are loosely structured facilities located in private residences or in more primitive accommodations and in remote locations in the Amazon jungle. New facilities open and others close down fairly frequently. No medical insurance covers the cost of treatment; many facilities depend on financial support from families of the residents, funds generated by resident activities (hand made crafts, bakery, etc.), and some have to resort to daily street solicitation to provide food for their clients or residents.

The first step in identifying the target organizations for the TC training was to compile a list from various sources. A team of Peruvian collaborators identified the residential and non-residential treatment or early intervention organizations in Lima and the surrounding provinces. An association representing several of the organizations provided a list of its membership. Information from another study designed to identify and classify Peru's treatment organizations according to the types of facilities, their infrastructure and operating conditions was used. The Peruvian Ministry of Health provided information they had collected on organizations operating in Peru. The size and functional type of organization was determined by a questionnaire sent to all of the organizations on the list.

Characteristics of the TC Staff

Staff of Peruvian organizations may be professionally trained, or they may be recovering addicts or former clients of the organization. This wide variation in organizations had significant implications for implementing the repeated measures design of the evaluation. The directors were given a specific set of criteria to use in selecting training participants: specifically, the director of each facility and two other professionals or administrators were invited to attend. However, the evaluation team learned during Wave 2 data collection

that the criteria had not always been followed. Some organizations rely heavily on volunteers and part time staff, and some even ask residents or clients to assume some operational responsibilities, though these residents/clients were are still technically in treatment. Some of these residents/clients had actually been selected to attend the training. The inclusion of residents or clients meant that the training team for some organizations included participants who were not actually part of the intended target population.

Several of these resident/client participants left the organization between Wave 1 and Wave 2 because they had completed or dropped out of their personal drug treatment programs. As a result, the evaluation team decided that it was necessary to attempt to obtain follow-up data from all Wave 1 participants regardless of whether they were still working in the organization. Additional items were added to the Wave 3 instrument that allowed us to determine which of the training participants were actually residents at Waves 1 and 2.

Selecting the mode of data collection

Telephone coverage in developing countries is much lower than in the U.S., even in business establishments (Peil, 1983, 1993). However, our Peruvian partners cautioned us that the mail delivery system in Peru was unreliable and that the mail questionnaires would take longer and produce a much lower response rate. As a result, in-person interviewing was the only way of reliably collecting data.

Locating Respondents

Newby et al., (1998) discussed some of the problems associated with locating study respondents in a developing country. Typical Western protocols commonly used may be problematical, if not impossible. For the TC Training Experiment, no contact information was obtained from study participants at Wave 1 because we assumed that the most of the selected training participants would remain through the duration of the study. However a sizeable number of the training participants left their organizations. One problem was that the selected residents or clients who attended the training, had left their organization by the time the Wave 2 data collection occurred (approximately seven months after Wave 1). Another problem was that several of the "unavailable" participants had relapsed or left their organization under unfavorable circumstances. In these cases, the director and other staff were not interested in providing contact information.

In the absence of the participant contact information interviewers resorted to every possible means of finding participants who had left the organization since the Wave 1 interview. They used contact information provided by the project assistant, other staff, residents or clients. To avoid further attrition, we obtained follow-up contact information for each participant during

904

an interim on-site visit by the data collection team. As a result, the attrition rate was much lower at Wave 3.

The evaluation team learned several important lessons of interest to other researchers considering a repeated measures data collection design in a developing country such as Peru. The following are some of these lessons:

(1) Standard telephone tracking procedures are not always feasible. Telephone directories are not readily available, and phone numbers may not be available in a database format for on-line name searches.

(2) The mail service may not forward mail nor is "address correction requested" always an option. Tracing respondents through the mail delivery system may not be very productive.

(3) Millions of people live in "settlements" or squatter communities in developing countries. The streets or lanes in these settlements are primitive and do not appear on maps nor do residences have a house number. Time-consuming and labor-intensive personal networking with neighbors or other residents is the only way to locate someone living in these settlements.

ETHICAL CONSIDERATIONS

Bamberg (1999), Warwick (1983, 1993) and others strongly encourage researchers to examine the ethical implications of their research in other countries, such as who will have access to the data and study results and how the results will be used. For the TC Training Experiment, the evaluation team carefully established protocols to ensure the confidentiality of data. The master list that linked organizations and participants to their ID's was securely maintained under lock and key by the evaluation team. Participants were informed that participation in the evaluation was voluntary and that they could refuse to answer any or all of the questions.

Since some of the organizational directors were concerned that their participation could affect licensure or other governmental compliance regulations, the evaluation team intends to use the results only to assess the effectiveness of the training and participation. Information or questionnaire responses will not be provided to anyone outside the evaluation team. The names of participating organizations will not be disclosed to government officials, and results will be reported only in aggregate form. In addition, results will be shared with our Peruvian partners and collaborators to assist them in developing strategies and initiatives for the prevention and treatment of substance abuse in Peru.

One interesting ethical issue emerged during the final stages of data collection. In the U.S. people are surveyed so frequently that the use of "respondent fees" has almost become an expectation. In Peru, however, social surveys are not so pervasive, and our Peruvian data collection team was adamantly opposed to the idea of paying respondents to participate in the evaluation. First, they were very concerned about the safety of the field interviewers having to carry around cash to pay respondents. But they also were concerned that paying the participants for providing data would set a precedent making it more difficult for other research projects. Ultimately, we deferred to their judgement to give participants an inscribed calendar as a small token of appreciation.

In an experiment with a control group, another ethical issue that researchers often struggle with is the use of placebos or the withholding of a desperately needed intervention. Training on the TC treatment model was perceived by all the members of the substance abuse treatment community in Peru as a serious need. Fortunately, INL recognized the political and ethical implications of withholding TC training from the TC Training Experiment control group and agreed to provide funding for the control group to receive the TC training curriculum after the evaluation data collection was completed.

CONCLUSION

In summary, this Peruvian case study provides evidence that the successful implementation of a social policy experiment conducted in a foreign country is possible but is dependent on specific key ingredients. One of the most important of these is to create partnerships and collaborate with host country interest groups to raise the intellectual and scientific level of the project, and to build cultural bridges for communication, mediating and resolving problems. The TC Training Experiment shows that researchers will face the same kinds of challenges they encounter when implementing a social policy experiment in their native country but that these challenges are compounded by cultural and language differences when the experimental setting is in another country.

References

Alers, J., (1993). Reliability of Survey Techniques in Highland Peru. In Bulmer, M. & Warwick, D. P. (Eds.), *Social Research in Developing Countries – Surveys and Censuses in the Third World.* London: UCL Press, p. 183.

Bamberger, M. (1999). Ethical Issues in Conducting Evaluation in International Settings. New Directions for Evaluation, 82, Summer, pp. 89-97.

Berk, R. A., Boruch, R. F., Chambers, D. L., Rossi, P. H., & Witte, A. D. (1985). Social Policy Experimentation. *Evaluation Review, 9* (4), 387-429

Biase, D.V. (1985). *A multi-year therapeutic community treatment research project.* New York: Daytop Village, Inc.

Brislan, R. Lonner, W., & Thorndike, R. (1973). *Cross-Cultural Research Methods.* New York: John Wiley & Sons, pp. 32-58.

Brislan, R., (1986). The Wording and Translation of Research Instruments. In Lonner, W. & Berry, J. (Eds.), *Field Methods in Cross-Cultural Research*, Beverly Hills: Sage Publications, pp.-137 – 164.

Buckley, C. (1998). Ideology, Methodology and Context – Social Science Surveys in the Russian Federation. *American Behavioral Scientist*, 42, 223-236.

Bulmer, M. (1998). The Problem of Exporting Social Survey Research. *American Behavioral Scientist*, 42, 151 – 167.

Bulmer, M., & Warwick, D., (1983/1993). Data Collection. In Bulmer, M. & Warwick, D.P. (Eds.), *Social Research in Developing Countries – Surveys and Censuses in the Third World*, London: UCL Press. pp. 145 - 160

Deutcher, I. (1973). Asking Questions: Linguistic comparability. In D. Warwick & S. Osherson (Eds.), *Comparative research methods*, Englewood Cliffs, N.J.: Prentice Hall. pp. 163-186).

Fontes, L.A. (1998). Ethics in Family Violence Research: Cross-Cultural Issues. *Family Relations*, 47, 53-61.

Geisinger, K. (1994). Cross-Cultural Normative Assessment: Translation and Adaptation Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*. 6, 304-312.

Hantrais, L., & Mangen, S., (1996). Method and Management of Cross-National Social Research. In Hantrais, L. and Mangen., S. (Eds.), *Cross-National Research Methods in the Social Sciences*, London: Pinter.

Herrera, R., DelCampo, & R., Ames, M. (1993). A Serial Approach for Translating Family Science Instrumentation. *Family Relations*, 42, 357-360.

Jobert, A. (1996) Comparing Education, Training and Employment in Germany, The United Kingdom, and Italy. In Hantrais, L. and Mangen., S. (Eds.), *Cross-National Research Methods in the Social Sciences*, London: Pinter.

Johnson, K. W. (1980). Stimulating evaluation use by integrating academia and practice. *Knowledge: Creation, Diffusion, Utilization*, 2, 237-262.

Johnson, K. W. (1990). Impact of a university-government partnership in criminal justice: A linkage model for the 1990's. *Journal of Criminal Justice Education*, 1, 167-193.

Johnson, K., Hexter, K., Garrison, D., & Sweet, D. (1996). Evaluation of outreach programs: The Ohio Urban University Program evaluation system. *Metropolitan Universities*, 6, 109-127

Jowell, R. (1998). How Comparative is Comparative Research? *American Behavioral Scientist*, 42, 168-177.

Kuechler, M. (1998). The Survey Method: An Indispensable Tool for Social Science Research Everywhere? *American Behavioral Scientist*, 42, 168-177.

Mitchell, R., (1993). Survey Materials Collected in the Developing Countries: Sampling, Measurement, and Interviewing Obstacles to Intranational and International Comparisons. In Bulmer, M. & Warwick, D.P. (Eds.), *Social Research in Developing Countries – Surveys and Censuses in the Third World*, London: UCL Press. pp. 219-239.

Munroe, R. & Munroe, R., (1986). Field Work in Cross-Cultural Psychology. In Lonner, W. & Berry, J. (Eds.), *Field Methods in Cross-Cultural Research*, Beverly Hills: Sage Publications. p. 115.

Newby, M., Amin, S., Diamond, I., & Naved, R. (1998). Survey Experience among Women in Bangladesh. *American Behavioral Scientist*, 42, 253-275

Patton, M. Q. (1996). *Utilization focused evaluation* (3rd ed.). Newbury Park, CA: Sage

Peil, M. (1983/1993). Situational Variables. In Bulmer, M. & Warwick, D.P. (Eds.), *Social Research in Developing Countries – Surveys and Censuses in the Third World*, London: UCL Press.

Schooler, C., Diakite, C., Vogel, J., Mounkoro, P., & Caplan, L., (1998). Conducting a Complex Sociological Survey in Rural Mali. *American Behavioral Scientist*, 42, 276-284.

Warwick. D. P. & Osherson, S. (1973). Comparative Analysis in the social Sciences. In D. P. Warwick & S. Osherson (Eds.) *Compartive Research Methods*, Prentice Hall, N.J: Eglewood Cliffs.

Warwick, D., (1983/1993). The Politics and Ethics of Field Research. In Bulmer, M. & Warwick, D.P. (Eds.), *Social Research in Developing Countries – Surveys and Censuses in the Third World*, London: UCL Press.