# Thursday, November 2, 10:20 A.M. to 11:50 A.M.

**SESSIONS 219: Presidential Strand Panel** **Room:** Lanai

Sponsored by the Topical Interest Group on Evaluation Use

**The American Cancer Society's Collaborative Evaluation Fellows Project: A Nationwide Strategy for Increasing Evaluation Capacity**

Chair: Donald Compton, American Cancer Society

Panelists: Donald Compton, American Cancer Society
Hallie Preskill, University of New Mexico
Paul Mattessich, Wilder Research Center
Leona Richardson, Macro International
Deborah G Bonnett, DBonnet Associates

Discussant: Doris Redfield, Appalachian Educational Laboratory

The American Cancer Society (ACS) is in its fourth year of an initial five-year effort to build local evaluation capacity nationwide. This strategy, called CEFP, involves a nationwide set of 20 collaborations among ACS, local schools of public health (and other graduate schools), faculty and students to design and implement necessary, useful, low cost, practical and timely evaluations for use in local program improvement. CEFP includes each of the five elements of evaluation capacity considered as necessary. The project's first three years included 85 studies, with an independent evaluation and a case study. The case study showed how the reports were used for program improvement, policy decisions, and allocation of funds. Basic to CEFP is the Utilization-Focused (UFE) model which as used in ACS, enhances evaluation capacity by involving local ACS staff in studies, and supplements this participation with training materials and events. This session will discuss how the CEFP incorporates the basic five elements of evaluation capacity, and the application of CEFP to other settings. Suggestions for the diffusion of this model throughout an organization will be made. CEFP materials will illustrate how the UFE strategy was implemented nationwide, and will showcase the value of several local studies.

**SESSION 220: MultiPaper** **Room:** Akaka

Sponsored by the Topical Interest Group on Program Theory and Theory-driven Evaluation

**Theory-driven Evaluation Findings**

Chair: E William Strang, Westat

Presenters: *A Multi-phase Evaluation of Best Buddies High School Programs: Uniting Program Theory with Psychological Research*, Barbara J Lehman, Claremont Graduate University

Best Buddies High Schools programs are student-led clubs designed to promote interactions between general education students and students with developmental disabilities. A conceptual model identifies several program outcomes, including increased intergroup friendships, improved social skills among students with developmental disabilities, and decreased stigmatization of students with developmental disabilities. This conceptual model is tested using a pretest-posttest control group design which examines program success through three phases. Phase 1 utilizes a survey of approximately 800 general education students to assess stigmatization and contact with students with developmental disabilities. Phase 2 examines individual student participant outcomes, including interpersonal skills, peer victimization, and loneliness. Phase 3 examines program implementation, and analyzes program "dosage" effects on the outcomes in phases 1 and 2. This evaluation makes use of established developmental and social psychological theory to develop a framework for the study of student attitude change and social development in high schools implementing the Best Buddies program.

*Understanding Program Impact by Illuminating Program Theory: Evaluation of the Families United To Prevent Teen Pregnancy Program*, Kathryn A Bowen, Bowen's Evaluation and Consulting Services

The evaluation of the Families United to Prevent Teen Pregnancy (FUPTP) Program is the subject of this paper. The program was implemented in a rural, poor Pennsylvania community among children (9-13 years) who were particularly at risk for early pregnancy. The program took a developmental approach that looked at the total human being and recognized the importance of family and community in helping to build protective factors for youth while decreasing risk factors.Using the six elements of program theory (problem definition, mediating processes, expected output, exogenous factors, critical inputs and implementation issues), this paper describes the FUPTP program, explains why, how, and under what conditions the program effects occurred, predicts the outcomes of the program, and specifies what needs to be done in order to bring about the desired program effects. Finally, the characteristics of the children enrolled, the children's families, program staff and setting are assessed in order to make linkages between mediating and/or outcome variables, the influence of program dose and the relationships among outcomes.

*Measuring the Impacts of Service Projects on Communities*, E William Strang, Westat

Measuring impacts of service projects on communities has always been a challenge. This challenge is compounded when national programs have different objectives across local projects but program officials want impacts measured using a common metric. For two years Westat has been conducting an evaluation of the impacts of AmeriCorps *National Civilian Community Corps projects on communities. The design is based on the premises that projects can be compared based on their logic models, with a metric resting on the extent expected outcomes are achieved, and that the reasoned judgments of project sponsors are valid indicators of impact. At AEA '99, we presented on the design of the study and its rationale. For AEA 2000 we will report on whether the design works.

**SESSION 221: Panel** **Room:** Koko

Sponsored by the Topical Interest Group on Evaluation of Services for Special Needs Populations

**Evaluation of Virginia's State Improvement Plan (SIP) for Special Education**

Chair: Sharon deFur, The College of William and Mary

Virginia's State Improvement Plan for Special Education (SIP) is a statewide strategic initiative intended to improve academic literacy and long-term outcomes for students with disabilities. The three strategic directions of the SIP are:
* to facilitate, in cooperation with local school divisions, an increase in the graduation rate of students with disabilities and school completion rate of students with disabilities in the context of higher academic expectations,
* to improve the performance of children and youth with disabilities by enhancing the knowledge, skills, abilities, and performance of all personnel who work with children and youth with disabilities, and
* to improve meaningful parent and student involvement with special education services.
The panel will present the strategic planning for the SIP, out of which the evaluation design grew; the evaluation design; the performance indicators, which are part of the strategic evaluation; and the futures scanning process, a unique aspect of the evaluation design.

Panelists: *Strategic Planning and Evaluation Design*, Sharon deFur, The College of William and Mary

SIP strategic planning began in March, 1998 with a stakeholder and partner meeting that included almost 100 representatives of stakeholder and partner groups. The evaluation plan grew out of the strategic planning process and focuses on issues identified through this vital function. The evaluation design of Virginia's SIP includes the following components:
* a strategic evaluation focusing on performances indicators and their relationship to the strategic directions
* the effectiveness of implementation evaluation, which examines the process and outcomes of SIP implementation
* management and monitoring
* project technical assistance
* scanning
The evaluation uses a learning orientation to guide its implementation and focuses on what factors in the project's implementation influence emerging and final results, and in what ways; and, what are the project's unintended effects?

*Performance Indicators*, Jeffrey M Schuyler, The College of William and Mary

Twenty-nine performance indicators were identified by stakeholders to monitor progress toward the five-year goals established to address each SIP strategic direction. The project management team will examine performance indicator data annually to determine progress toward it's five-year goals and strategic directions. The performance indicators and five-year goals provide operational definitions by which to gauge success of the project and are the basis for the project's strategic evaluation. The presentation will include indicator descriptions and data.

*Futures Scanning*, Lissa Power-deFur, Virginia Department of Education

The SIP uses scanning to gather information from Virginia and the nation about trends and future events that might affect special education in Virginia. Scanning is a proactive tool for program development and revision, and seeks to answers such as:
* From the perspective of stakeholders, what are the emerging trends and issues that could influence, either positively or negatively, the success of the SIP and the special education process in general?
* From the perspective of stakeholders throughout Virginia, what is working or not working with respect to SIP initiatives and the special education process in general?

**SESSION 222: MultiPaper** **Room:** Waimea

Sponsored by the Topical Interest Group on Needs Assessment

**Strategic Planning and Needs Assessment**

Chair: Doug Leigh, Pepperdine University

Presenters: *An Internal Strategic Planning Institute: A Mechanism for Building Organizational Capacity within a Geographically Dispersed Consortium*, James W Altschuld, Morgan V Lewis & Floyd McKinney, The Ohio State University

In 1999, the United States Office of Vocational and Adult Education funded a consortium of institutions dedicated to conducting research, professional development, and dissemination in regard to Career and Technical Education (CTE). It consists of 5 university sites and 2 associate partners spread throughout the country. To develop a sense of purpose across the consortium as well to maintain its national leadership role in the field of Career and Technical Education, an internal strategic planning institute was proposed. In this presentation an overview of the following topics will be provided: the nature of the consortium; the design of the total strategic planning effort; electronic data collection; data obtained from a unique needs sensing strategy employed by the consortium; the structure of the Strategic Planning Institute; results from the Institute; and problems encountered. It is felt that the above approach will generalize to other similar consortia.

*Increasing Needs Assessment Capacity in a Community-university Partnership*, Mary Sue Hamann, Carol MacKinnon-Lewis & Doug Easterling, University of North Carolina at Greensboro

The mission of the Center for the Study of Social Issues is to enhance individuals' and families' lives through interdisciplinary research, evaluation, and education. We carry out our mission by facilitating applied research and evaluation across university departments and with numerous community agencies and organizations. We are currently conducting needs assessments or providing technical assistance for three community projects: a coalition of homeless providers, an Early Head Start project, and a Family Resource Center. We are also currently planning a community-wide needs assessment of area nonprofits and funders to determine the evaluation and program planning needs to which we might respond. We will ask these agencies to describe and prioritize their needs for each of the following: strategic planning, program design and planning, formative evaluation, summative evaluation, participatory evaluation, outcomes based management, ongoing needs assessment of local social issues,and accessibility to a pool of evaluation contractors. We will present not only a brief review of our needs assessments findings, but also address whether our technical assistance improved the needs assessment efforts of local groups. Moreover, we will present the results of the community-wide needs assessment and describe our efforts to respond to the documented interest in improving local evaluation capacity.

Discussant: Susan Crawford, University of North Carolina at Greensboro

**SESSION 223: Think Tank** **Room:** Niihau

Sponsored by the Topical Interest Group on Pre-K - 12 Educational Evaluation

**High Stakes Testing in Schools: Drafting an AEA Position**

Chair: Sandra Mathison, State University of New York at Albany
Facilitators: James R Sanders, Western Michigan University
Linda Mabry, Indiana University
Charles L Thomas, George Mason University

Most US states have adopted high stakes testing, that is, tests used to make high stakes decisions about students, teachers, and schools including retention, graduation, tracking, merit pay, school take-overs, and budgeting. While on the one hand there is great enthusiasm for such tests among policy makers, corporations, and state/district school administrators, parents, teachers, and some researchers are raising serious questions about high stakes testing. There are currently strong opposing forces debating the uses of high stakes testing in schools. This session will provide a forum (through brief framing presentations by evaluators with expertise and interest in student assessment, presentation of other professional association positions on high stakes testing, and discussion by participants) which will lay the groundwork for drafting a written statement regarding the use of high stakes tests in K-12 schooling. This statement will subsequently be presented to AEA with a request for endorsement.

**SESSION 224: Panel** **Room:** Hilo

Sponsored by the Topical Interest Group on Non-profit and Foundations Evaluation

**The Consuelo Foundation's Use of Evaluation To Guide and Strengthen Its Hawaii and Philippine Community Development and Housing Programs**

Chair: Terry George, Consuelo Zobel Alger Foundation

The Evaluation Center at Western Michigan University has worked closely with the Consuelo Foundation since 1994, providing evaluation services and building evaluation capacity within the foundation. Unique aspects of the work that will be discussed include grounding of the evaluations in explicit foundation values and professional evaluation standards, longitudinal and interactive evaluations focused on continuous improvement, adaptation of the CIPP Model for integration in the foundation's development process, use of multiple evaluation methods, and ongoing engagement of stakeholders. This session will offer a foundation's perspective on evaluation and building evaluation capacity among foundation personnel; describe an evaluation framework used to evaluate foundation projects in Hawaii and the Philippines; and report on an evaluation of a community development project in the Philippines, a self-help housing project in Hawaii, and metaevaluations of children and youth programs in the Philippines. Members of the audience will be engaged in a discussion of the foundation's evaluation approach.

Panelists: *A Foundation President's Perspective on Evaluation*, Patti Lyons, Consuelo Zobel Alger Foundation

The Consuelo Foundation is devoted to serving the needs of neglected and abused children and women and other poor persons in Hawaii and the Philippines. The Foundation was established with a charitable grant from Consuelo Zobel Alger in 1988. The Foundation is grounded in values of a healthy community, including safe, drug-free, and violence-free living environments and a spirit of giving back to the broader community. It develops and implements programs such as self-help housing, services for street children, and livelihood projects for women. Soon after the Foundation was established, its Chief Executive Officer and President, Patti Lyons, recognized that programs should be guided by a sound program of external evaluation. She commissioned The Evaluation Center to conduct ongoing evaluations of some of the Foundation's core programs. Ms. Lyons will describe the role of evaluation in the Foundation, reflect on evaluation experiences, and project steps needed to develop and sustain an evaluation capacity.

*Evaluation Framework for Consuelo Foundation Evaluations*, Daniel L Stufflebeam, Western Michigan University

The overall evaluation approach was grounded in the Foundation's mission and values and professional standards for evaluations and was longitudinal. Context, input, process, impact, effectiveness, and viability evaluations were conducted. These were applied to help the Foundation move successfully and accountably through its developmental stages, including selection of priority needs, design of a project, successfully implementing the project, reaching the targeted beneficiaries, effectively addressing the beneficiaries' assessed and targeted needs, and institutionalizing and -- as appropriate -- transporting the project's successful products and services. The evaluations employed multiple evaluation techniques and included regular interactions with staff and other stakeholders.

*Evaluation of the Consuelo Foundation's Community Development Programs in the Philippines*, Jerry Horn, Western Michigan University

In collaboration with city government, the Counselo Foundation engaged in a community development project that focused on improving the quality of life for some of the poorest of the poor in San Carlos City, Negros, Philippines. The project targeted squatters -- who had previously lived in make-shift houses or shanties on private and public lands, railroad right-of-ways, cemeteries, and other properties that they did not own -- and families of the lowest paid city government employees. The Evaluation Center was contracted to conduct an ongoing evaluation of the project to determine its impact on residents' quality of life, report findings to Foundation personnel, and develop a capacity among team members and other stakeholder groups for using evaluation as a tool for decision making and program improvement. This paper describes the approaches and the key elements of the evaluation of a dynamic program for which there were multiple decision makers and a variety of stakeholders.

*Evaluation of the Consuelo Foundation's Self-help Housing Project in Hawaii*, Arlen R Gullickson, Western Michigan University

The Evaluation Center has been conducting an 8-year longitudinal evaluation of the Consuelo Foundation's self-help housing program for Hawaii's working poor. The evaluation has used multiple methods, including program profiles, environmental analysis, builder interviews, case studies, goal-free evaluation, and feedback workshops. The Consuelo Foundation systematically insists on and takes steps to assure quality in all aspects of the self-help housing and community development work. It has a unique and exemplary explication and application of values for its programs, providing an explicit value base for the evaluation work. This paper describes the evaluation methods and results denoting the significance of the longitudinal and interactive characteristics of this program and the Center's working relationship with the Foundation.

*Metaevaluation of Consuelo Foundation's Children and Youth Programs in the Philippines*, Lori A Wingate, Western Michigan University

The Consuelo Foundation solicited the help of The Evaluation Center to provide a substantive analysis of Philippine Children and Youth Foundation (a subsidiary of the Consuelo Foundation) evaluation reports. In response, metaevaluations were conducted by students in a graduate evaluation seminar at Western Michigan University taught by Dan Stufflebeam and Lori Wingate. The students used a metaevaluation checklist developed by Dan Stufflebeam to judge the quality of evaluations against the Program Evaluation Standards. This paper summarizes the main issues surfaced by this metaevaluation and summarizes the Consuelo Foundation's reaction to the metaevaluations. It also distills the main lessons learned vis a vis improving the theory and methods of metaevaluation.

Discussant: Terry George, Consuelo Zobel Alger Foundation

**SESSION 225: MultiPaper** **Room:** Puna

Sponsored by the Topical Interest Group on Business and Industry

**Evaluating Learning Programs in Business and Industry**

Chair: Maryann Durland, Rockman ET AL

Presenters: *Maximizing Value: Identifying Drivers of Effective Corporate Training in a Variety of Corporate Settings*, Darryl L Jinkerson, Arthur Andersen; Terry L Shoemaker, Impact Consulting Inc and MA Mrzlak, McDonalds Corporation

Last year, American Businesses invested nearly $66 billion dollars on corporate training according to Training Magazine. This amount reflects an increase of 25% over the last four years and an increase of $29 billion over the last eight years. Maximizing the value of these tremendous investments is partially the charter of corporate training evaluation efforts. Helping to ensure that training is effective and efficient is critical to delivering a quality training product and is a role championed by evaluators. Training's reputation and ability to deliver desired performance improvement effectively, efficiently and economically is often predicated on evaluation efforts pin-pointing specific areas for improvement and identifying key initiatives to be followed to deliver the desired end product. This paper will examine the key drivers of effective training as evidenced in three very distinct training organizations: Professional Services, Services Sector (Fast Food) and the Entertainment Industry. The core component of this research will be that a common instrument will be used in all three settings to collect participant reactions to training. This will allow direct comparisons across industries and will allow the identification of key driver trends within and across industries. By identifying the key drivers, specific areas for targeted improvements can be made to maximize the impact rather than subjectively selecting improvement efforts in the hopes that they will be effective. By employing a key driver's approach, the cost benefit of precious resources in terms of dollars and people can be maximized. Design efforts can be focused on select high impact initiatives that are highly correlated with overall training quality. Attendees will learn what are common key drivers are for corporate training in various industries. Discussion will include specific decisions made based on this information and the implications for the decisions. Issues of external validity and other issues relevant to applications to other environments and organizations will also be discussed.

*An Evaluation of Goal-specific Mentoring in a Manufacturing Setting*, Steven C Jones, Utah State University

A case study is provided of an evaluation that was conducted at a manufacturing facility to determine the effectiveness of a new training program. The company had increased its workforce by more than two-thirds within three years. Many of these new-hires lacked the necessary skills to machine extremely difficult parts. In response to this problem, a team was instigated to develop a mentor-training program that met the unique learning needs of each participant. Information is provided regarding the training model, the evaluation design, and the lessons learned.

*Training Evaluation from the Participant's Perspective*: *Doing Level 1 Evaluation Right*, Greg V Michalski, Action Evaluation Research Institute

The evaluation of training and development programs in the private sector has traditionally emphasized measurements of participant satisfaction using what has come to be labeled as 'level 1 evaluation'. Although several authors have critiqued the overuse of level 1 evaluation (Alliger & Janak, 1989; Dixon, 1987; Sleezer, Hough, & Gradous, 1998) its use remains deeply entrenched. According to Bassi, Benson, and Cheney (1996) 'Only Level 1 evaluations are common, conducted for 94 percent of all companies' (p. 38). This paper takes a fresh look at the potential of level 1 evaluation methods as a source of valid quantitative and qualitative data to assess the quality and impact of training programs. The paper focuses on the development, deployment, analysis, and use of level 1 evaluation as a viable means of training evaluation. Referencing fundamental concepts from several arenas including program evaluation, survey construction and data analysis a framework of good practice is developed and discussed as a guide to good level 1 evaluation practice.

**SESSION 226: Panel** **Room:** Kohala

Sponsored by the Topical Interest Group on Collaborative, Participatory & Empowerment Evaluation

**South Carolina First Steps to School Readiness: A Results-based Accountability Model of Community Coalitions**

Chair: Pamela S Imm, University of South Carolina

In July, 1999, the South Carolina Legislature passed enabling legislation to fund First Steps, a statewide school readiness initiative designed to improve children's readiness for first grade. With almost 20% of South Carolina's children testing "not ready" for first grade in 1998, the newly-elected Governor was successful in convincing the legislature to allocate $20 million dollars to fund this education initiative. First Steps is designed to be a community-based coalition process to involve all 46 counties in South Carolina. The legislation requires that each of the 46 community coalitions work in two phases: a planning phase where agencies/organizations collaborate to conduct needs and resource assessments and develop a plan, and an implementation phase where counties compete for state money to fund their implementation plan. The legislation requires that First Steps be a results-focused initiative with the emphasis on accountability at the state and state levels. The need to demonstrate accountability is a priority issue for all involved including the Governor, the legislature, and the First Steps state and local offices. This panel will include a variety of presenters to discuss how South Carolina is addressing issues of accountability in the grantmaking processes, the implementation strategies, and in evaluation methods.

Panelists: *Results-oriented Grantmaking/Grant-implementation*, Pamela S Imm & Abraham H Wandersman, University of South Carolina

The First Steps legislation requires that counties apply and receive a planning grant prior to requesting funds for an implementation grant. The application for the planning grant is based on a model of results-based accountability, known as (ROGG), results-oriented grantmaking/grant-implementation. This system is designed around 10 accountability questions that contain the elements of effective programming. For the planning grant application, grantees answer the first several ROGG questions, and in the implementation grant application, the remaining questions are addressed. During this presentation, the accountability questions/model will be presented as well as empowerment evaluation methods for assisting the coalitions to successfully complete the applications. By building an accountability model into the applications, it is expected that the coalitions will increase their likelihood of achieving meaningful results in their programming.

*Building Coalition Capacities To Improve County-level School Readiness Initiatives*, Matthew Chinman & Joy Kaufman, Yale University

In the context of the First Steps statewide initiative, capacity is being defined as the ability to implement the elements of the accountability model: conduct needs/resources assessments, formulate logical and realistic goals and objectives, choose the best programs, ensure the fit of the program, plan the program well, implement the program with quality, conduct useful process and outcome evaluations, utilize continuous quality improvement strategies, and plan for the continuation of funding given successful outcomes. The goal of the ROGG assessment is to empirically demonstrate the link between this type of capacity and the quality of grants funded. The evaluation design for assessing capacity building includes several different methods. Some of these are: a documentation system for the statewide technical assistance staff, the impacts of the training sessions on county program operations, a comprehensive assessment of board functioning, and the quality of the grant applications. The following table summarizes the approach to evaluate the degree to which counties utilize and understand the accountability model and the quality of their grant applications. Hierarchical multiple regression analyses will be conducted to determine the effects of the various predictors on the quality of the grant proposals.

*Building a Foundation of Knowledge: How Statewide Initiatives Are Evaluated*, Georgann Stemper & April Ace, University of South Carolina

This presentation will focus on reporting the common methods of 12 states for evaluating, both internally and externally, the process and results of statewide initiatives aimed at children and their families. These states include: North Carolina, South Carolina, Florida, California, Iowa, West Virginia, Ohio, Hawaii, Oregon, Vermont, Rhode Island, and Missouri. This project was conducted in order to develop a foundation of knowledge about common evaluation practices in order to share this with other initiatives and to compare it to the evaluation methods used in the South Carolina First Steps statewide initiative. Specific data gathered include models of evaluation in use, organizations in charge of evaluation, collaboration between state offices and evaluators, and details about the implementation of the evaluation models. As this is a "work in progress," the presentation will stimulate a discussion about the evaluation of complex community initiatives.

**SESSION 227: MultiPaper** **Room:** Kona

Sponsored by the Topical Interest Group on Extension Education Evaluation

**Developing and Strengthening Evaluation Capacity**

Chair: Michael E Newman, Mississippi State University

Presenters: *Indicators of Dependency in Participatory Extension Education*, Nancy Grudens-Schuck & Tasha M Hargrove, Iowa State University

Extension and non governmental organization staff who use participatory education seek to decrease dependency of people involved in their programs. This goal is consistent with the extension mission that values engagement and capacity-building. Using examples from participatory agriculture and forestry programs in Canada, Indonesia, and the

US, this paper discusses the nature of indicators of dependency that facilitators use to gauge progress of participatory programs. In two programs, facilitators believed that evaluation processes failed to measure gains in independence, initiative, honesty, networking, collaboration, and other informal indicators of successful capacity-building. The paper proposes fine-tuning of evaluation criteria for participatory and capacity-building programs such that dependency indicators become more common, explicit and public. Development of criteria for understanding dependency must also include views of participants. The paper contributes to our understanding of extension education that seeks to build people's capacities for self-management in the area of agriculture and natural resource management.

*A Framework To Build Evaluation Capacity of Extension Professionals*, Rama Radhakrishna, Clemson University

Cooperative Extension, like all other public agencies, has seen an increased governmental emphasis on program performance and accountability. A frequently asked question is, what happened as result of your extension program? Furthermore, several studies have identified the lack of time and resources, limited expertise in evaluation methodology and in developing surveys, as factors inhibiting agents in conducting systematic evaluation of their programs. To address these issues at the county and state levels, a three-prong evaluation framework has been designed to assess the effectiveness and impact of extension programs. In addition, a two-year program to build evaluation capacity of county staff and establishment of an Evaluation Clearing House to systematically evaluate extension programs is being implemented.

*Developing Evaluation Capacity among Penn State Cooperative Extension's 4-H/Youth and Family Living Extension Educators*, Tena L St Pierre & D Lynee Kaltreider, Pennsylvania State University

This paper will describe strategies university research faculty are employing to enhance evaluation capacity of 4-H/youth and family living extension educators within Penn State's Cooperative Extension Service. The land-grant university's outreach model has been described as the most influential innovation in higher education for its ability to apply and disseminate research-based knowledge on a large scale. However, 4-H/youth and family living extension educators have limited expertise and experience in systematically evaluating multi-year community programs. Strategies reported are within the context of a five-year research project entailing collaborations among Penn State researchers, extension educators, and middle/high schools in 10 Pennsylvania communities to implement and evaluate a multi-year substance abuse prevention program.

The paper will inform evaluation practice by providing strategies employed to: (1) develop collaborative teams of university researchers, extension educators, and schools; (2) implement and evaluate the project by extension educators; and (3)increase evaluation skills of extension educators though education and application.

**SESSION 228: Panel** **Room:** Ewa

Sponsored by the Topical Interest Group on Quantitative Methods: Theory and Design

**Evaluation Methods: Lessons from Meta-analysis**

Chair: Mark W Lipsey, Vanderbilt University

This panel will draw on a considerable body of cumulated meta-analysis work to examine various methodological issues pertinent to the design and interpretation of evaluation. By focusing on cross-study comparisons within different treatment domains, meta-analysis allows investigation of the relationship between evaluation findings and the methods and procedures used to generate those findings. Such inquiry illuminates the role of method in shaping the findings and ways to improve those methods.

Panelists: *Measurement: The Achilles Heel of Outcome Evaluation*, Mark W Lipsey, Vanderbilt University

An impact evaluation is only as good as its outcome measures. If those measures are unreliable, unresponsive, or irrelevant, they will not capture meaningful program effects. A large meta-analysis of delinquency interventions with multiple operationalizations of multiple outcome constructs measured at different times provides an instructive database for exploring outcome measurement issues. The results will be discussed in terms of the relative importance of the different facets of outcome measurement relating to construct identification, sensitivity to change, psychometric properties, and timing of measurement.

*Noise, Noise, Everywhere: Reflections on Method Variance in Evaluation Research*, David B Wilson, University of Maryland

Experience in conducting meta-analyses and summarizing collections of meta-analyses has shown that evaluation studies are fraught with nuisance variance. Despite this, real treatment effects do emerge from collections of studies. This paper explores the sources of variability in observed treatment effects attributable to method using large scale meta-analytic databases. A counter intuitive finding is that often high quality comparative studies produce larger mean effects than low quality comparative studies within a treatment domain. This finding appears to remain after controlling for other observed study difference. It is argued that this stems from better control over nuisance variance, such as careful attention to measurement and greater standardization of the treatment intervention. Implications for evaluation research are explored.

*Contributions of Meta-analysis to Evaluation Research: The Case of School Violence Programs*, Sandra J Wilson, Vanderbilt University

A meta-analysis of school violence prevention and treatment programs is used to illustrate a mean-change approach to synthesis and to identify and explain some common threats to internal validity in quasi-experimental research. Most meta-analyses exclude studies from eligibility that do not employ control groups. Excluding studies that examine a single treatment group measured before and after an intervention, or studies that compare multiple treatment groups to each other, may not reflect the available program evaluation literature. This is especially important with field-based interventions where it can be expensive and difficult to obtain suitable comparison groups. This analysis uses a standardized mean-change measure, computed for each sample within a study, in which the effect size of interest is the change from pretest to posttest rather than the traditional treatment versus control group difference. This approach allows the meta-analyst to synthesize a wider range of literature on the effectiveness of school violence programs. In addition, since change is indexed separately for treatment and comparison groups, several common threats to validity, including regression toward the mean and maturation can be studied directly.

*More Likely Than Not: Estimating Potential Program Impact from Retrospective Data*, James H Derzon, George Washington University and David B Wilson, University of Maryland

Accurately managing expectations of preventive interventions can be as important to the survival of intervention efforts as demonstrating the impact of the intervention itself. In interventions designed to ameliorate the potential impact of risk factors on outcomes, the underlying relationship between the targeted risk factor and the outcome approximates the upper bound at which an intervention can be expected to be effective. Thus, accurately estimating the potential of a risk-targeted intervention to improve outcomes should be a central concern of both formative and summative evaluators. When retrospective case-comparison sampling designs are used to identify risk factors, as when delinquent youth are compared with non-delinquents, the estimated relationship is upwardly biased. Using a series of Monte Carlo trials we illustrate the magnitude and nature of this overestimation and provide a formula to correct this inherent bias. Overstating the magnitude of a predictive relationship biases theory development and may lead to overly optimistic estimates of the anticipated impact of social interventions.

**SESSION 229: Panel** **Room:** Honolulu

Sponsored by the Topical Interest Group on Alcohol, Drug Abuse, and Mental Health

**Evaluation Findings from the SAMHSA Homelessness Prevention Cooperative Agreement**

Chair: Kendon J Conrad, University of Illinois at Chicago

Although there has been a considerable investment in housing, treatment, and support services over the past decade, homelessness continues to be a risk for individuals with serious mental illnesses and particularly those with co-occurring alcohol and other drug disorders. In 1996, the Center for Mental Health Services and the Center for Substance Abuse Treatment of the Substance Abuse and Mental Health Services Administration launched a two-phased, three-year initiative to document and evaluate the effectiveness of homelessness prevention interventions that focus on persons with psychiatric and/or substance use disorders who are formerly homeless or at-risk for homelessness and who are engaged with the mental health and/or substance abuse treatment system(s). This panel presents the background, logic models, goals, and findings of four of the eight projects in the Collaborative Program to Prevent Homelessness.

Panelists: *Can Money Management Improve Financial Stability in Mental Illness? Combining Regression Discontinuity and Non-equivalent Control Designs*, Kendon J Conrad & Michael Matters, University of Illinois at Chicago; and Daniel Luchins & Patricia Hanrahan, University of Chicago

This study evaluated whether a community-based representative payee (RP) program coordinated with case management could improve financial stability in a population with severe mental illness. The study design compared a treatment group receiving RP to two comparison groups: a regression discontinuity (RD) control group that scored below a need cutoff and was not assigned to RP and a non-equivalent control group at a different agency that scored at or above the need cutoff but had no RP program available to it (total study n=186). The regression discontinuity design showed such a strong positive effect that the treatment group scores went from being worse than the RD controls at baseline to being better at the six-month posttest. The positive effect was cross-validated by a significant positive effect in the non-equivalent control group design. Implications of the findings and reflections on the usefulness of the unusual program evaluation design are discussed.

*Measuring Salient Aspects of Service Provision in Preventing Homelessness among Individuals with Serious Mental Illness*, Colleen Clark, University of South Florida and Colleen Rich, Louis de la Parte Florida Mental Health Institute

The Boley Homelessness Prevention Project examines facets of services and service provision as part of a national evaluation effort to determine which homelessness prevention interventions are effective with people who have a severe mental illness and may also have a substance abuse disorder. A comprehensive process evaluation and measurement of intermediate processes are used to better understand the results of a quasi-experimental design with non-equivalent control groups.Preliminary results will be presented of the changes over time in residential stability and other outcomes for the Boley participants compared to two other programs that take different approaches to prevention services. The relationships among organizational climate, working alliances and the match between services needed and received will be presented to describe the saliency of relationships among staff and residents, flexibility of service provision, and a supportive organization as intermediate processes by which outcome effects are achieved.

*The New York Housing Study: A Randomized Experiment*, Sam Tsemberis, Sara Asmussen & Linda Moran, Pathways to Housing; and Beth Shinn, New York University

The New York Housing Study in NYC is a randomized experiment, comparing the effectiveness of two models of housing for homeless individuals with major mental illnesses, many with substance use problems (91%), and history of criminal justice involvement (74%). The study compares a consumer preference supported model where clients are offered immediate access to independent permanent housing with treatment services available if clients choose them with the continuum of care model where treatment and sobriety are prerequisites for placement in congregate housing with treatment required. The supported housing program is operated by Pathways to Housing and the continuum is represented by numerous housing providers serving the same population in NYC (treatment as usual). Upon recruitment, individuals were randomly assigned to study condition. There are 225 individuals (23% female, 77% male) in the study coming from streets/parks/subways (51%), psychiatric centers (35%), shelters (8%), other (2%); having diagnoses of major psychosis (51%), bi-polar (16%), major depression (16%) with 51% of these having an additional Axis II diagnosis. The 12 month follow-up rate is 89%. Results will compare the one year outcomes for the two groups on measures that include residential stability, housing retention, psychiatric symptoms, substance use, utilization of services and social networks, and other and relevant variables. The discussion will focus on the strengths and weakness of each program model, the feasibility and cost of the two models, as well as their underlying clinical and value assumptions.

*Homelessness Prevention Therapeutic Community*, JoAnn Y Sacks and Stanley Sacks, National Development and Research Institutes Inc

This study compares a Homelessness Prevention TC (for mothers and their children) with a standard TC. The mothers reported high levels of substance abuse, depression, anxiety, parenting stress, engagement in illegal activities, history of arrest and incarceration, HIV risk behaviors, homelessness, physical, sexual or emotional abuse, and have exposed their children to trauma and abuse. Both groups showed significant decreases in substance abuse, psychological symptoms, and illegal activities 12 months after treatment entry, but only mothers in the Homelessness Prevention TC showed significant improvements on measures of housing stability and positive supports in the community. This group also showed significantly greater improvements on the number of children for whom they were responsible and obtaining economic support. These results provide preliminary empirical support for integrated mother-child programs and extend the use of TC methods found effective in reducing substance abuse to the area of homelessness prevention.

Discussant: Frances L Randolph, Center for Mental Health Services

**SESSION 230: Roundtables** **Room:** Kahuku

Sponsored by the Topical Interest Groups on Health Evaluation and Alcohol, Drug Abuse and Mental Health

**Evaluation of Health Programs**

*(This session includes two 45-minute rotations of roundtables. The Host will ask the tables to rotate at 11:05.)*

Host: Tom Cook, Vanderbilt University

**Roundtable A (First Rotation): Evaluating Health Education Events and Exhibits**

Presenters: *Building Evaluation Capacity at the Local Level to Establish Measures for Assessing the Impact of a National Health Education Program*, Debra J Holden, North Carolina State University and Rena L Large, National Education Association Health Information Network

The process for building evaluation capacity at the local level through training of data coordinators in a national breast and cervical cancer education project will be discussed. The Health Information Network (HIN), a not-for-profit organization operating within the National Education Association (NEA), is implementing this pilot project in two states. Through funding from the Centers for Disease Control and Prevention (CDC), NEA HIN has trained state affiliates and local teams to conduct educational events to address these diseases. A comprehensive evaluation plan was developed to collect process and outcome measures at the national, state, and local levels. Data collection from the local communities was key to measuring impact on health behaviors. To facilitate this data collection, a strategy was developed for utilizing local data coordinators. Review of this process for training and data collection will be discussed, as well as lessons learned and strategies for use in similar programs.

*Healthworks! Kids Museum: Connecting Community through Evaluation*, Dennis W Rudy, Indiana University; Deborah Drendall, Health Works Kids Museum; and Reg Wagle, Memorial Health Foundation

Healthworks! is an interactive children's museum designed to enhance student learning, parent and teacher collaborations, and the overall health of the community. The health model used was designed as a prototype for other health organizations and communities to replicate complete with learning histories and artifacts. The data provides evidence for both formative and summative aspects of the program evaluation. A mixed-methods approach has been utilized with a five-year plan outlining multiple outcome measures related to the knowledge, behaviors and attitudes necessary to build healthy communities nation-wide.

**Roundtable B (First Rotation): Working with Stakeholders in Health Programs Serving Underserved Populations**

Presenters: *Tactful Strategic Tips When Evaluating Social Development Projects in Latin America*, Rosa Maria Nuñez-Urquiza, Instituto Nacional de Salud

A review of actual case studies of health related projects evaluated in Latin American Countries.a) How assess project's political environment and how to weigh its interference with the project's development. b) The importance of partnership with local leaders in evaluating the project: the dilemma between making evaluations more feasible and avoiding evaluation bias. c) Assessing the extension and penetration of the project in different communities before selecting communities for impact evaluation. d) The social impact in LDC from impact evaluations done by "neutral" foreign experts. e) Developed a methods to be able to "harverst" all the good outcomes including the unexpected ones.

*Health Disparities: Evaluation Strategies with Communities of Color*, Ana A Lopez–De fede & Jeannine Salone, University of South Carolina

Evaluation efforts with rural women of color seeking prenatal care services and a statewide health needs assessment with limited English speaking Latino/as will be used to illustrate the impact of each of these factors on the evaluation process. Renewed concerns for health inequalities regarding the poor has begun to produce important findings. Evaluators focusing on health disparities within communities of color have struggled with the best way to categorize members this population in order to examine health disparities. Some of the factors used to examine health disparities by evaluators include economic status, ethnic and racial affiliation, geographical location, occupation (concerning social class and economic connotations) and health condition (as a function of health status, service use and financing). This paper will present an overview of each of these factors and the strategies necessary to ensure culturally competent evaluations with rural communities of color.

**Roundtable C (First Rotation): Inter-organizational Relationships and Health Evaluation**

Presenters: *Partnering across the Continuum: Developing an Evaluation Strategy that Satisfies Multiple Agencies*, Jennifer W Campbell, Albert Einstein College of Medicine; Harold Keister, Department of Public Welfare; and Jim Burd & Cheryl A Martin, Pennsylvania Department of Aging

Developing effective multi-organizational partnerships is crucial to addressing contemporary health care issues. However, those very multi-organizational partnerships present endless challenges when developing an evaluation strategy that will effectively meet each stakeholder's requirements. This presentation will discuss a Robert Wood Johnson Foundation funded initiative to develop a coordinated system of health care for residents of SSI-funded personal care homes. The project involved a large health care organization, an urban Area Agency on Aging, the State Department of Aging and multiple components of the State Department of Welfare. Each organization involved, including the funder, had different assumptions about the project, and different conceptualizations of the final evaluation product. In addition, the changing political context of community-based initiatives continually challenged the effort. This paper will discuss the process, the barriers and a series of best-practice recommendations for nurturing the development of inter-agency evaluative efforts.

*The Delivery of Health Care to the Elderly: Factors that Impact the Success of Interorganizational Linkages between Area Agencies on Aging and Managed Care Organizations*, Joyce M Iutcovich, Keystone University Research Corporation

Gerontologists, health care professionals, and government policy makers have focused their attention on agencies that provide needed health and wellness services to the elderly--but in a disconnected, uncoordinated way. Questions have been raised about the various types of organizational and structural relationships (interorganizational linkages) that are most appropriate to better serve the elderly's needs. There are specific concerns about funding streams and access to limited resources, duplication of services, gaps in service delivery, and the appropriate roles for various agencies to play in a coordinated system. This study evaluated a Pennsylvania demonstration project that was designed to coordinate the delivery of health and social services to the elderly by developing linkages between area agencies on aging and managed care organizations. As an evaluation of the process involved in forging these linkages, the results of the study highlight the different models used, the key features of the planning process, the benefits and risks as identified by the partners, the factors that facilitated the linkages, and the challenges the partnering organizations faced. In conclusion, recommendations are made to others considering the establishment of such interorganizational linkages.

**Roundtable D (First Rotation): Evaluating Innovations in Health Service Delivery**

Presenters: *A Qualitative Process Evaluation of Integrated Care for Persons Living with HIV*, Miles A McNall & Tomas Soto, The Core Center

The literature is replete with descriptions of various models of hospital-based integrated care for persons with HIV. Such descriptions are often presented with the results of process evaluations, including figures for patient enrollment, retention, and types and amounts of services provided. Very few discussions of process evaluation results address the challenges involved in establishing multidisciplinary patient care teams. Our study is a three-year federally funded process evaluation of the integration of multidisciplinary outpatient services to persons with HIV. The integration combines primary care with mental health, case management and chemical dependency services. A variety of qualitative process data, including provider and patient perceptions of integrated care, are being collected to evaluate the implementation of integrated care at intervention and comparison sites. Challenges in conducting the evaluation will be discussed.

*Evaluation Research as a Tool for Developing an Evidence-based Health Care Collaboration*, Lyndee Knox, Tracy Norton, Sabra Hitchcock, University of Southern California; Anna M Malsch, Claremont Graduate University; and Petite Konstantin, Haven House Inc

The current paper discusses an evaluation used to develop an evidence-based healthcare collaboration between the University of Southern California, Department of Family Medicine and Haven House, a safe house for victims of domestic violence. A pre-program needs assessment is conducted to determine the health needs of victims of domestic violence and common barriers they face when accessing health care. This information is collected with the goal of informing and determining the training needs of USC Family Medicine residents and will also be used to guide the design of service provided at Haven House. An implementation evaluation of the collaboration between USC's Family Medicine resident training program and Haven House and an impact evaluation of the collaboration are conducted. This paper demonstrates the utility of evaluation as a collaborative tool between two organizations, and has interesting implications for both theory and practice in the field of evaluation.

**Roundtable E (First Rotation): Abstinence-based Programs for Teens**

Presenters: *Abstinence-only Education: Examining Results from the Field*, Gerald Halpin & Glennelle Halpin, Auburn University; and John M Trent, Southern Christian University

Adolescent pregnancy rates are higher in the United States than in any other industrialized nation. Further, teens are contracting sexually transmitted diseases at an alarming rate. Add other psychological, social, and emotional effects of teen sexual activity, and the health risk becomes blatantly obvious. To combat these problems, Congress authorized in 1996 $250 million for abstinence-only education grants. Being reported are results from the evaluation of one of these funded programs. Participants were seventh- and ninth-grade students in city/county school districts in a southeastern state who were taught by classroom teachers lessons from a modified Sex Respect curriculum (seventh grade) and Abstinence Only: Send a Clear Message (ninth grade). Outcome evaluation using results from the Youth Survey measuring federally mandated objectives indicated that the program was effective. Both objectives-based subscale scores and self-rated likelihood of remaining abstinent until marriage were significantly higher after program participation. Implications for theory and practice are discussed.

*Abstinence Education as a Response to Teens Having Babies: The Oklahoma Experience*, Tom James & Joshua Stockley, University of Oklahoma

Every 17 hours, a baby is born to an Oklahoma mother not old enough to drive. Every three hours a baby is born to a mother not old enough to vote. The appropriate response to teenage sexual activity continues to be a topic of scientific and political debate. This paper reports findings from the first two years of an evaluation of the Oklahoma Department of Health Abstinence Education Program. The program was instituted to reduce teen pregnancy and adverse health effects of teen sexual activity. Six local programs have provided abstinence education to 3000 students in 80 school districts across 12 counties. The evaluation focuses on documentation of the operation and activities of the local program sites and short- and long-term program outcomes. Data are collected through site visits and interviews with program staff; from program records; and from pre-test, post-test, and follow up instruments administered to treatment and comparison groups.

**Roundtable A (Second Rotation): Challenges in Assessing Implementation of Health Service Programs**

Presenters: *Using Evaluation Tools and Results in Performance Measurement and Accountability*, Sharon K Stout, Community Well

An evaluation of the Montgomery County Maryland Department of Health and Human Services, Infants and Toddlers Program was designed to test whether paper and automated systems files records on referrals to the program, assessments, services provided, and outcomes were recorded consistently and appropriately. The recorded mental health outcomes were to be assessed using goal attainment scaling. Review and analysis revealed that few outcomes (mental health or other) were scored. The evaluation results were shared with stakeholders, and recommendations developed. In addition, the evaluation results and database were shared with the Early Childhood Initiative Planning Group and other stakeholders and used to develop recommendations on data sharing, performance measures, outcomes, and accountability for agencies and service providers addressing families with at-risk infants and toddlers ages 1-3. This paper summarizes the evaluation and the process of developing the Early Childhood Recommendations.

*How Much Is Enough? Measuring Duration, Intensity and Service Mix in a Prenatal Care Program*, H Lindsay Guyn, Barrington Research Group Inc

One of the more important challenges in a program evaluation is to determine how successfully an intervention was delivered to the target population. Measuring the level of implementation at an individual level can provide a more focussed estimate of program impact. This paper describes the efforts made to develop a service delivery index for the Canada Prenatal Nutrition Program. To develop such an index, the duration and intensity of the interventions had to be taken into account, and more importantly, the fact that each participant could receive a different set of prenatal care services (which included food supplements, dietary counseling, child care, and several others). The derived measures of intervention level were then used in multiple regression analysis with birth outcome as the dependent variable. Results showed that there was a significant increase in infant birth weight for high levels of prenatal care.

**Roundtable B (Second Rotation): The Role of Evaluation in Planning Health Community Promotion Efforts**

Presenters: *Evaluating Community-based Coalitions for Health Promotion: What We Know, What We Don't Know*, Naomi G Penney, Cornell University

Community-based coalitions have been used as a strategy for health promotion over the past several decades. Although the literature is replete with articles evaluating coalitions there are still some significant gaps. This roundtable session will present a review of the current literature and propose areas that need future research. To stimulate discussion an evaluation will be presented which attempts to fill some of the gaps in the literature. The evaluation to be presented is a mixed-method evaluation designed to help health educators and community health planners work better with their target communities and engage them in the community health planning process.

*Strategies for Change*, John B Bare & Lizabeth L Sklaroff, John S and James L Knight Foundation

The public and private sector both spend hundreds of millions of dollars each year in attempt to add value to society and improve general quality of life. However, organizations increasingly want to know the level of impact their money and activities are having on their target areas. Knight Foundation's evaluation team is gathering information on community needs and nonprofit capacities to assess its communities' current status and help guide plans to create change. The presenter will discuss the different methods foundations can use to assess community conditions to successfully plan, produce and measure change. The presenter will also discuss how an organization working within a larger system can document and attribute change. Participants will be encouraged to share their organization's strategies for planning and measuring change.

**Roundtable C (Second Rotation): Using Meta-analysis and Metaevaluation To Examine Health Interventions**

Presenters: *Are Drug Prevention Programs for Adolescents of Color Effective? A Meta-analysis*, Katrina L Bledsoe, Claremont

Graduate University

The proposed paper presentation seeks to provide insight through meta-analytic procedures about the effectiveness of drug prevention programs for adolescents of color. In general, adolescents are at risk for drug use and abuse and designing effective prevention programs for this population has been a growing concern for government, health and school organizations. Adolescents of color are considered more at-risk because of disproportionate rates of health problems (e.g., cirrhosis of the liver) and high rates of problem behavior (e.g., violence). This meta-analysis explores the effectiveness of drug prevention programs in terms of common approaches used for adolescents of color, the impact of these programs on attitude and behavior change, and the effectiveness of culturally-focused programs versus non-culturally focused programs. This information is valuable for both practitioners and theorists in their quest to increase the capacity of the design and evaluation of effective drug prevention programs for adolescents of color.

*Applying Summative Metaevaluation to Public Health Evaluation*, Nalin Johri, CARE-India

This paper presents a summative metaevaluation of CARE-India's Mid Term Evaluation Quantitative Survey (MTS) of the Integrated Nutrition and Health Program (INHP). At the mid point of the INHP a MTS was instituted to assess progress from baseline. The metaevaluation of the MTS attempts to provide a fair assessment of the quality of this survey. For purposes of the metaevaluation, the procedure adopted was to focus on three identified indicators. The steps followed were - 1) Specify the objectives of the INHP; 2) Specify the objectives of the MTS; 3) Comparison of the design of the MTS and Baseline Surveys; 4) Comparison of the results from the MTS and CARE-India's Qualitative Assessment; and 5) Assessing the MTS on the Standards and Criteria established for evaluations. Findings and Discussion of Results: Overall, on the standards for quality evaluations, the MTS performed creditably, in spite of shortcomings on certain criteria, largely on account of inadequate attention to procedures and weak program documentation and contextual analysis. Based on this metaevaluation, the overall quality of the MTS can be objectively vouched for. It is recommended that the objectives of this evaluation be defined to include the information needs of all stakeholders and provide program and context information; adhere to rigors of laid down procedures and have alternate sources for triangulation of results; study cost-effectiveness of the MTS; report results in a more user-friendly form with charts and graphs; address issues of adequate staffing; undertake a summative and formative metaevaluation. The paper makes a strong case for applying standards and criteria to public health evaluation and leads to the cultivation of self-evaluation skills in an effort to 'Increasing Evaluation Capacity' in non-governmental organizations.

**Roundtable D (Second Rotation): Measuring Substance Abuse Treatment Effects in Different Settings**

Presenters: *A Progress in Treatment Measure as an Evaluation Tool*, Kelli J Klebe & Dina L Dunn, University of Colorado at Colorado Springs; and Maureen L O'Keefe & Joe Stommel, Colorado Department of Corrections

A progress in treatment measure, the Therapeutic Community Client Assessment Survey (TCCAS; De Leon & Kressel, 1997), was used as both a program evaluation tool and a clinical tool in a prison therapeutic community for substance abuse. The TCCAS has 14 subscales to measure behaviors and attitudes that are related to progress in treatment. There is a client version with 117 items and a staff/counselor version with 14 items. Items are rated on a 5-point Likert-type scale. There were 334 residents in treatment who were assessed monthly by the staff and self. There were 1014 resident ratings and 1154 staff ratings. This paper compares ratings between clients and staff, looks at changes that occur during treatment, and looks at the relationship between progress in treatment and length of stay in program. The use of this type of measure as an evaluation tool is discussed.

*Cost-effectiveness of Residential and Outpatient Substance Abuse Treatment*, G Nicholas Braucht, Charles S Reichart & Lisa J Geissler, University of Denver; and Michael W Kirby & Nancy R VanDeMark,Arapahoe House Inc

In an evaluation of the relative cost-effectiveness of residential and outpatient substance abuse treatment programs, a heterogenous sample of 309 clients were randomly assigned to either residential, outpatient, or usual care treatment. Based on ASAM clinical criteria, 90% of clients assigned to usual care were then given residential treatment and 10% were given outpatient treatment. All study clients were assessed at baseline and again at 6 and 12 months after baseline on a wide range of outcome measures. The average cost of outpatient treatment ($578) was significantly less than residential treatment ($2,283). Both groups of clients showed dramatic improvement over time on a wide variety of outcomes, but there were relatively few statistically significant differences between residential and outpatient groups. Although these few outcome difference almost always favored residential treatment, they were generally small in magnitude. Given the large difference in the cost of these two types of treatment, the cost-effectiveness of residential treatment was considered to be questionable.

**SESSION 231: Think Tank** **Room:** Oahu

Sponsored by the Topical Interest Group on Cluster, Multi-site and Multi-level Evaluation

**Principles for Evaluating Comprehensive Community Initiatives**

Facilitators: David M Chavis, Association for the Study and Development of Community
Kien Lee, Association for the Study and Development of Community

Discussant: Linda Bowen, National Funding Collaborative on Violence Prevention

Comprehensive Community Initiatives (CCI's) are a major approach to prevention and the promotion of community capacity and well being in a wide variety of social and health related areas. There have been several articles and publications about the challenges in conducting evaluation of CCI's. Few offer practical advice on how to conduct evaluations of these complex, multifaceted systems-level initiatives that are based on the collaboration of a wide variety of community institutions and leaders. Practitioners and funders find themselves even less prepared to "evaluate" the approach and methods of their evaluators. In order to assist evaluators, practitioners, funders and community leaders in comprehensive crime prevention efforts, a list of 28 principles was developed based on an extensive literature review, interviews with 12 leading evaluators of CCI's and then an additional review by a group of thirty evaluators and practitioners. In this session, following a description of the development, participants will break up into four groups to discuss the four groupings of principles and answer the following questions: What capacities do evaluators needs to implement these principles? What capacities do practitioners, funders and community leaders need? Other comments on the principles? The director of a national CCI support organization will be the respondent**.**

**SESSION 232: Panel** **Room:** Waialua

Sponsored by the Topical Interest Group on Pre-K-12 Educational Evaluation

**Monitoring and Measuring the Effectiveness of Systemic Reform in K-12 Mathematics and Science Education: Issues and Innovative Practices**

Chair: Bernice T Anderson, National Science Foundation

The National Science Foundation (NSF) has aimed to catalyze in a comprehensive way the educational delivery systems for the improvement of teaching and learning in precollege mathematics and science education. NSF's Directorate of Education and Human Resources (EHR) launched its Statewide Systemic Initiatives (SSI) program in 1991 and in 1994, expanded its systemic programming to urban and rural areas through the Urban Systemic Initiatives (USI) program and the Rural Systemic Initiatives (RSI) program. To measure and substantiate the outputs, outcomes, and impacts of its systemic reform programs, NSF has funded monitoring, evaluative, and impact studies. The panel will focus on the challenges and effective approaches in designing and interpreting data within comparative and/or causal structures for attributing results to systemic reform efforts. Each presenter will share preliminary findings and lessons learned about measurement and assessment issues in systemic reform.

Panelists: *Study of the Impact of Statewide Systemic Initiatives*, Norman L Webb, Council of Chief State School Officers and Iris R Weiss, Horizon Research Inc

The main goal of this three-year project is to study the outcomes and impacts of Statewide Systemic Initiatives (SSI) on student learning, curriculum, and policy. The following key questions are being addressed by this study. What has been the impact of statewide systemic initiatives on student learning and other important variables? What lessons have been learned about designing, implementing, evaluating and supporting statewide systemic reform? Data from the National Assessment of Educational Progress (NAEP) have been analyzed to develop profile for SSI states and non-SSI states and to draw inferences about the impacts of the SSIs, particularly in the area of improved performance for minority students. Case studies are being conducted to generate theories and competing hypotheses about systemic reform in mathematics and science. A model of systemic reform has been developed and is being refined to depict levels of reform and paths of influence within and across levels.

*A Large-scale Educational Evaluation and Assessment - How Reform Works: An Evaluative Study of National Science Foundation's Urban Systemic Initiatives*, Jason J Kim, Systemic Research Inc; Rolf Blank, Horizon Research Inc; Pendred Noyce, Noyce Foundation; and Lloyd Richardson, University of Missouri at St Louis

A large-scale, three-year study is being conducted to determine the impact of National Science Foundation's (NSF) Urban Systemic Initiative (USI) program. The evaluative study is analyzing the effectiveness of the USI program across all 21 school districts based on an inferential causal model using the process drivers (policy; curriculum and instruction; assessment; professional development; convergence of resources; broad-based support; leadership and governance) and outcome drivers (student outcomes; teacher quality). The project's components and outcomes include: 1) Key Indicator Data System (KIDS) to compile five years of qualitative and quantitative data to assess annual and cumulative progress of 21 school districts; 2) Survey of Enacted Curriculum to monitor classroom instructional practices, curriculum content, materials and technology, and teacher preparation; 3) site visits to analyze changes in policy and practices; and 4) dissemination of a new paradigm for large-scale, cross-site, longitudinal rubrics for statistical evaluative modeling.

*SSI Impact Study: Merged Research Model of Systemic Change*, Frances P Lawrenz & Douglas Huffman, University of Minnesota

This evaluation project is incorporating the drivers of systemic changed suggested by the National Science Foundation and factors derived from research on school change into an empirical model to help explain the impact of student achievement that can be accomplished through a systemic reform effort. Data collection is underway, using a variety of methods including student, teacher and principal questionnaires, interviews, observations, and secondary analyses of student achievement data. Explanatory models are being constructed for schools in three SSI states that have had significant interactions with the SSIs and for schools in the three states that have not had interaction with the SSIs. In addition, a model will be produced for students in a state which never had the SSI program. The multi-stage, multi-state comparisons will serve as the basis for identification of the unique aspects that can be attributed to participation in the SSI.

*Summary of Systemic Initiative Core Data Elements Findings Over Time: Lessons Learned*, Gary Silverstein and Laurie Somers, Westat

Since 1997, the Division of Education System Reform (ESR) has used the Core Data Elements (CDE)-a web-based collection system-to compile quantitative data from three Systemic Initiative (SI) programs, i.e., Rural Systemic Initiative (RSI) Program, Statewide Systemic Initiative (SSI) Program, and Urban Systemic Initiative (USI) Program. This online system is designed to obtain standardized information about the implementation and impact of SI-related activities. Benefits to having the systemic initiatives annually report a common set of core data elements include: 1) Information from the projects can be easily aggregated to provide a snapshot of the systemic initiatives. 2) ESR staff use the data to monitor overall and project-specific trends over time. 3) The relative status of the three programs can be compared. Findings from three collections have been summarized, as well as lessons learned and recommendations regarding the use of web-based systems to collect standardized data from SI programs.

Discussant: Bernice T Anderson, National Science Foundation

**SESSION 233: Panel** **Room:** Waianae

Sponsored by the Topical Interest Group on State and Local Government

**Working Beyond Our Borders To Increase Evaluation Capacity in Government**

Chair: Rakesh Mohan, Washington State Joint Legislative Audit and Review Committee

A number of U.S. legislative auditors and evaluators have been active in working with their counterparts in other countries. These international working relationships have provided opportunities to share valuable experiences in the field of government oversight and program evaluation. The panelists will discuss the lessons learned from such interactions and how these interactions have been useful in building evaluation capacity in government.

Panelists: *Program Auditing: Crossing Boundaries*, Marion M Higa, State of Hawaii

Hawaii's mid-Pacific location has afforded me the opportunity to come in contact with audit delegations from Thailand, China, and Korea. My involvement with the National Conference of State Legislatures has enabled me to travel to Germany and Japan. In the case of Germany, the contact was not so much with auditors as with parliamentarians and their role in stimulating economic development. In the case of Japan, I spent some time with prefectural auditors. Their audit structure appears to be very reflective of their culture as a whole. From all these experiences, I conclude that the concept of program auditing, as contrasted with financial auditing, is a nascent concept but one that more and more countries will adopt, albeit slowly.

*Sharing Lessons Learned To Advance Evaluation of Performance in the Pacific Islands*, Cheryle A Broom, New York State Metropolitan Transportation Authority

Evolving democracies, economic development necessities, cultural and environmental pressures, and unique isolation yield many challenges to the governments of Guam, American Samoa, the Commonwealth of the Micronesia, and the Republic of Palau. Through the Pacific Islands Training Initiative (PITI), a joint program of the U. S. Department of Interior and the Graduate School, USDA, the author provided training and technical assistance in government oversight and program evaluation to representatives of these governments. The author found the opportunity to share experiences and advice was of particular value to the participants. The presentation will highlight how the PITI approach advanced lessons learned that diverse participants could use to improve performance in their respective governments. Further, she will provide some tips for those serving in a consulting capacity to increase evaluation capacity across borders.

Discussant: Craig Russon, Western Michigan University

**SESSION 234: Panel** **Room:** Molokai

Sponsored by the Topical Interest Group on International and Cross-cultural Evaluation

**Evaluating Poverty Reduction in International Development Programs**

Chair: Rema N Balasundaram, The World Bank

This panel examines the results of evaluations of poverty reduction programs and policies pursued by major international development ogranizations. The papers present cross-country experiences and lessons in the context of less-developed countries.

Panelists: *Evaluation of Poverty Reduction Programs: Lessons from Cross-country Experience*, Kalanidhi Subbarao and Gloria Rubio, The World Bank

This paper compares the experience in evaluating poverty reduction programs across less developed countries. It summarizes the main lessons from these evaluations in terms of methodological approach, information base, and policy relevance of results. The paper also examines the extent to which they were aimed at and successfully contributed to build institutional capacity.

*Evaluation Capacity Building and Gender*, Gita Gopal, The World Bank

This presentation will provide results of an evaluation of the World Bank's Policy on Gender. It will address issues of relevance and implementation of the Bank's Gender Policy and will explore the development of gender-related indicators at the project and country level. The presenter will discuss the challenges faced in developing a methodology to assess gender issues and the inherent importance of the development of sound indicators which can be fed into development work.

*The Grassroots Development Framework of the Inter-American Foundation: An Evaluation Tool*, Emilia Rodriguez Stein, Inter-American Foundation

The Grassroots Development Framework (GDF) is a tool to measure the outcomes and impacts of grants supported by the Inter-American Foundation. It allows the Foundation and its grantees to report on inputs and outcomes of project activities and show the results of community-based development interventions. It measures results that can be readily quantified, as well as intangibles that can be observed and inferred. The GDF is based on the experience of the Foundation, its grantees, and In-Country Service (ICS) contractors, and assesses outcomes and impacts at three levels: 1) individuals and households; 2) community organizations and networks; and 3) policies and practices at the national level.

Discussant: Ian C Davies, Office of the Auditor General of British Columbia

**SESSION 235 Roundtables** **Room:** Maui

**Conference Committee Spotlight: Leading Edge Issues in Evaluation Theory and Practice**

*(This session includes two 45-minute rotations of roundtables. The Host will ask the tables to rotate at 11:05.)*

Host: To be announced

**Roundtable A (First Rotation): Obligations to Stakeholders: Informed Consent and Human Subjects Review**

Presenters: *When Did We Stop "Informing" in the Informed Consent?* Rhonda A Bohs & Lauren Strano, DLR Consultants Inc

University evaluation and research departments are required to submit all research projects and most evaluation projects through Institutional Review Boards in order to protect the rights of persons involved in these endeavors. However, evaluators and researchers working as independent consultants do not face this requirement. This roundtable discussion will focus on the need for evaluations working within independent consulting companies to ensure that client/subject rights are maintained, especially when working with grassroots organizations with little experience with research and evaluation ethics and confidentiality.

*Creative Strategies for Overcoming Barriers to Human Subjects Compliance Approval with Youth in the Juvenile Justice System*, Lance E Schnacker, Northwest Professional Consortium Inc

This presentation focuses on creative strategies for overcoming barriers to obtaining human subjects compliance approval with Institutional Review Boards (IRBs) when the subject population is youth in the juvenile justice system. IRBs view this population as particularly vulnerable because of their status as minors without parents as legal guardians and susceptible to legal risk in the disclosure of criminal behavior. Key issues discussed include (a) who advocates for the youth when the state is the legal guardian, and (b) how disclosures of certain categories of information are perceived to put the youth at potential risk. Several case studies are examined as examples of where problems occur and creative solutions are presented.

**Roundtable B (First Rotation): Data Care**

Presenters: *Examining Confidential and Sensitive Data: Making Connections with Criminal and Juvenile Justice Agencies for Program Evaluation*, Maureen H Rumptz, Juliette R Mackin, Robbianne T Cole & Lisa M Lucas, Northwest Professional Consortium Inc; and Nahama Broner, New York University

Successful harvesting of sensitive and confidential data from criminal and juvenile justice agencies can be a unique challenge compared to obtaining other public records. Focusing on the actual project work of the presenters in the areas of domestic violence, jail diversion program for persons with co-occurring disorders, and juvenile justice programs for youth with substance abuse problems, this presentation will provide useful guidelines for working with local criminal and juvenile justice agencies. Techniques for establishing relationships with key players in data management, maintaining local and state policymaker interest in the research, engaging their support in promoting your full access, and dissolving barriers to acquiring data will be discussed. An examination of the usability of existing criminal and juvenile justice data, how to match across agencies, and suggestions for additional data collection outside of existing criminal and juvenile justice databases will be included.

*Data Quality Standards: Institutionalizing Attention to Data Quality Issues*, Natalia Pane, American Institutes for Research and Steven W Zwillinger, United States Department of Education

This session will present the Data Quality Initiative developed by the U.S. Department of Education (ED) to assess the reliability and credibility of program performance data and to foster improvement across ED's data systems. Driven by the rationale and demands of GPRA, the Department has developed seven Data Quality Standards. These standards provide a framework for program managers to monitor and report program performance and for officials to assess data quality progress and to plan for long-term data quality improvements. The standards are designed to promote discussion and thereby improvement of data systems and are being integrated across all ED programs. We will present the policy context, the Data Quality Standards/Initiative, and possible implications for states, other grantees, and their evaluators. Audience members will be asked to comment and share examples of data improvement or coordination initiatives. Results from the first year of implementation will be presented.

**Roundtable C (First Rotation): Examining the Profession**

Presenters: *The Social Psychology of TIGs: Are We Holding Evaluation Back?* Ryan M Quest & E Jane Davidson, Claremont Graduate University

Research within social psychology has demonstrated that groups often attempt to establish their own superiority by derogating outgroups. In recent history, the advancement of the field of evaluation was stinted by an internal debate regarding qualitative and quantitative methods until the value of multiple method evaluation and critical multiplism (Shadish, 1993) was finally recognized. The present paper explores the social psychology behind this tendency to present various evaluation approaches as mutually exclusive, and the extent to which it impedes the advancement of the field. In an attempt to show the value of moving beyond these false dichotomies, we present an example of how theory-driven evaluation might be combined with a seemingly incompatible approach to forge a tool more powerful than the sum of its parts. Suggestions are made regarding the potential for increased inter-TIG collaboration.

*The Use Patterns of Evaluation Literature and Theory in the Practice of Program Evaluation*, Shahpar Modarresi, Prince George's County Public Schools

Variation is seen as a fact of life when one studies the field of program evaluation. The practice of program evaluation is a field of diverse theories, literature, and methods that have separate roots and reflect different ideologies. Each theory or method has its own advocates in practice. There are different groups in the field who are influenced by the different types of evaluation literature and theories. The purpose of this paper is twofold. First, to produce a model of evaluation literature and theories reflecting the current practice of program evaluation. Second, to identify groups of evaluators who are influenced by different types of evaluation literature and theories in practice. To address these objectives, a survey of people working in the area of evaluation was conducted. The survey sample was composed of 500 evaluators selected from the American Evaluation Association membership directory.

**Roundtable D (First Rotation): Implications of Capacity Building**

Presenters: *Evaluating Program Sustainability Efforts: A Guided Discussion*, Michael C Fagen, University of Illinois at Chicago

This roundtable will discuss the evaluation of program sustainability efforts. The lack of conceptual clarity regarding program sustainability presents key challenges to the evaluator. Should the evaluator track program continuation, program effects, or changes to individual or organizational capacity? Discussing this question will start a beneficial dialogue for practitioners evaluating program sustainability efforts. This dialogue will be enhanced by the presentation of a case study: The Aban Aya Youth Project, a school and community based prevention program for African American adolescents. The effort to sustain Aban Aya centers on parent educators who are trained to teach the Aban Aya curriculum. While this effort builds parent capacity to conduct prevention programming, adolescent-level program effects may be diluted by parent implementation. This potential conflict between building capacity and maintaining program effects represents one of the fundamental dilemmas for evaluators of sustainability efforts, and will be a focus of this roundtable discussion.

*Increasing Evaluation Capacity: Is It Always Good for Everyone?* Tenzing Donyo, The California Endowment

University evaluation and research departments are required to submit all research projects and most evaluation projects through Institutional Review Boards in order to protect the rights of persons involved in these endeavors. Academic Units at the University of Illinois at Urbana-Champaign have been heavily involved with developing outcomes assessment plans for the past two years. All eighty-three units have submitted their highly individualized plans at this point. Phase two (occurring this past year) involves departments working to refine and implement their particular plan. The authors have worked at both the campus and departmental levels, as well as with individual faculty to facilitate these unit's efforts. Their facilitation has spanned all eight UIUC colleges, and disciplines ranging from engineering to women's studies. The main emphasis for discussion will be the lessons learned (political as well as practical) from facilitating this large-scale effort toward implementation of highly individualized departmental outcomes assessment plans. Additionally, specific measures and methods of collecting outcomes assessment will be presented. This will be of interest to all involved in the very current trend of student outcomes assessment in higher education. However, evaluators and researchers working as independent consultants do not face this requirement. This roundtable discussion will focus on the need for evaluations working within independent consulting companies to ensure that client/subject rights are maintained, especially when working with grassroots organizations with little experience with research and evaluation ethics and confidentiality.

**Roundtable E (First Rotation): Instrument Administration and Response Rates**

Presenters: *A Technological Approach to Increasing Accessibility and Use*, Darlene F Russ-Eft, AchieveGlobal Inc and Regina L Atwood, San Jose State University

Job demands typically result in low response rates within organizations leading to limited generalizability, validity, and use. The purpose of this project was to create an instrument available in different media for assessing behavioral performance of customer service personnel. This instrument could then be used for training needs assessments, as well as evaluations of transfer of trained skills to the job. It was based on five dimensions identified from 2100 critical incidents gathered in North America (Russ-Eft, Berrey, Boone, & Winkle, 2000). Validation took place with 885 respondents from 10 different organizations, covering various industries including financial services, business services, manufacturing, wholesale trade, and higher education. The overall Cronbach alpha was .94. Factor analyses confirmed the underlying dimensions. By making the instrument available through paper-based, email-based, and web-based versions (with the web version providing an option for a summary report), accessibility and use have been increased.

*Maximizing Survey Response Rates in Community-based Prevention Program Assessment*, Scott W M Burrus, Northwest Professional Consortium Inc

Yes, it's possible to experience mail survey response rates in the 80 to 90 percent range. In "Maximizing Survey Response," Scott W.M. Burrus, Project Coordinator for NPC Research, Inc., will explain how they obtain successful response rates in community-based program surveys. After this brief explanation a practical discussion on how to maximize survey response rates in mail, telephone, and web-based surveys, focusing on the actual project work of members of the group, will consume the remainder of the roundtable session. A great discussion on how to maximize resources to produce the highest response rates possible!

**Roundtable F (First Rotation): Adapting to the Context: The Realities of Evaluation**

Presenters: *Good Evaluation Can Be Unsystematic and Feature Data from the Fringes*, Morris K Lai & Donald B Young, University of Hawaii

An evaluation that is not very systematic and includes data from many sources, some of which are from the fringes, should not necessarily seem discordant given that our ways of knowing and learning are akin to such an approach. We learn about many things through diverse, often unsystematic, activities such as trying, observing, asking, discussing, and listening to little bits at a time. Many interactions may prove to be essential to our learning. By also including relevant standards (e.g., content and evaluation standards), an evaluation has an even better chance of achieving an exceptionally high level of understanding of a program and its impact. We will report on such an evaluation of a professional-development program conducted over a four-year period. That evaluation proved to be noticeably more powerful than more commonly done ones that were more systematic and based on a small number of major data sources.

**Roundtable A (Second Rotation): Behind the Scenes of Evaluations**

Presenters: *Management of Evaluation Studies: Common Problems and How to Avoid Them*, Stephen K Dietz, Westat

Professional Evaluators need more than good research skills -- they also need good management skills if they are to be successful project directors or principal investigators. This paper focuses on important management methods that have proven useful in keeping studies within budget and on schedule; keeping clients and sponsors informed and satisfied; coordinating the tasks, staff, and data effectively; understanding contractual obligations, ensuring the quality and usefulness of findings; and communicating the results in a manner that will be well received. This paper is presented by a Senior Vice President of Westat, who is both a manager and a researcher. He has taught management seminars and has been a frequent presenter at AEA on a wide range of quantitative and qualitative evaluation methods. He shares many valuable lessons learned from more than 30 years of evaluation project management.

*Medicaid Outreach Evaluation: An Annotated Bibliography of Traditional and Internet Resources*, Cheryl Meyer, Carl Brun, Katherine Cauley, Carla Clasen, Barbara Fuller, Bill Mase, Lisa Rambaldo, Kelly White, Tanya Willis & Betty Yung, Wright State University

A team of faculty from medicine, psychology, nursing and social work have been engaged by the Ohio Department of Human Services to conduct a state-wide evaluation of Medicaid Outreach initiatives in eight-eight counties. The evaluation methodology features a unique approach to the literature review which led to the development of an annotated bibliography on Medicaid Outreach Initiatives. The bibliography includes references from traditional sources, and the internet, providing indexing, cross-referencing and annotation strategies which respond to the complexities of hyper-links and validating data from sources which do not follow a traditional peer review process. To increase evaluation capacity in the 21st century, these literature review skills are increasingly important. This paper presentation will both review the process of developing and provide a sample of the product created, an annotated bibliography on Medicaid Outreach Initiatives across the county.

**Roundtable B (Second Rotation): Looking at Logic Models**

Presenters: *Logic Models: Friends or Foes?* Lauren Strano & Rhonda A Bohs, DLR Consultants Inc

The use of logic models within program evaluation has become very "trendy" in recent years. However, experiences with logic models has lead many to report they are not utilized by program staff and supervisor and are merely relevant for evalaution purposes. This roundtable discussion will focus on the use of logic models as tools within evaluations. The discussion will highlight ways to develop logic models that meet needs of the program manager, evaluator, and funding agency.

*TOP's (Targeting Outcomes of Programs) Usefulness As a Program Logic Model*, S Kay Rockwell, University of Nebraska and Claude F Bennett, US Department of Agriculture

A teaching/learning website, "Targeting Outcomes of Programs (TOP): A Hierarchy for Targeting Outcomes and Evaluating Their Achievement" is available at http://deal.unl.edu/TOP/ . TOP uses a seven-level hierarchy (commonly known as Bennett's hierarchy) to expand upon "inputs/outputs/outcomes" in the program planning and evaluation process. TOP applies to single agency and collaborative interagency programming. The website helps conceptualize program and evaluation plans to target outcomes, track progress toward reaching the outcomes, and evaluate the degree to which outcomes are achieved. The first section helps programmers understand theoretical concepts. The second is an application section that walks programmers through each level of the hierarchy so they can develop worksheets that include pertinent information for evaluating the extent to which outcomes are achieved. After a brief overview of TOP, the discussion will focus on pro's and con's of using TOP as a generic program logic model. What are the benefits? What are the drawbacks?

**Roundtable C (Second Rotation): At the Intersection of Technology, Learning and Evaluation**

Presenters: *Evaluating Wildwood Junior High School: Using Online Simulations in the Teaching of Evaluation*, D Michelle Hinn & Bertram C Bruce, University of Illinois at Urbana-Champaign and Kevin M Leander, Vanderbilt University

This presentation will report on the design and implementation of an online graduate evaluation course, which incorporated a simulation of a junior high school. The simulation, implemented in order to provide a "whole class" evaluation experience, was centered on a request by the fictional "National Education Foundation" to conduct an evaluation of a school district's use of their funds to support the use of educational technologies at Wildwood Junior High School. During this presentation, the presenters will share information about the course design, including an account of the decision to use a simulation environment for online learning and the continued evolution of the simulation as the course progressed. Additionally, the successes and limitations of using a simulation activity in an online course will be discussed, drawing on course data from multiple sources, including the archives of synchronous and asynchronous communication tools used by the instructors and students throughout the course.

*Using Electronic Listservs in Post Secondary Program Evaluation*, Susan A Tucker & Katherine Whiteley, E & D Associates

USDE's Fund for the Improvement of Post Secondary Education recently launched a listserv to 'support' evaluation of distance education programs.  Focusing on the Learn Anytime, Anywhere Projects (LAAP), this electronic forum is tasked to assist a national network of external evaluators as well as project directors and other relevant stakeholders in designing and implementing ongoing improvements over the three to five year life of project funding.  The goal of this session is to share learnings revealed by an analysis of the email exchanges over an eight month period.  We will focus on the following issues and strategies generated to solve them:

1. what are critical questions and indicators/standards to include in designing evaluation plans?
2. how can cost-effectiveness of distance education be conceptualized and  managed across all the projects?
3. how can evaluation findings be used to coalesce diverse public and private sector stakeholders as well as ground public policy for future distance education programming?

**Roundtable D (Second Rotation): Meeting the Evaluation Needs of Higher Education Institutions**

Presenters: *Increasing Evaluation Capacity and Use: Collaboration with Stakeholders*, Summers Kalishman, Jan Mines & Lisa Sema, University of New Mexico

As a result of a curriculum change, our medical school had  a need for outcome data about its graduates.  To maximize resources and minimize additional effort, we collaborated on a modification of an existing evaluation from the Associate Dean's office to graduate medical program directors asking them to evaluate the skillfulness of our graduates. We expanded the evaluation effort by sending a parallel questionnaire to our graduates in training asking them to self-assess their competence and to assess the preparation they received from the medical school curriculum. We reported one year summary data to curriculum stakeholders who requested additional evidence, ratings from a comparison group of first year graduates from other schools in similar residency programs.  We incorporated this request into the next year's data gathering and report.  We plan to study the validity of these data by obtaining secondary information from national residency licensure examinations and local performance assessments of residents.

*Helping Units Implement Their Departmental Level Outcomes Assessment Plans*, Cheryl D Bullock, Lizanne DeStefano & John C Ory, University of Illinois at Urbana-Champaign

Academic Units at the University of Illinois at Urbana-Champaign have been heavily involved with developing outcomes assessment plans for the past two years.  All eighty-three units have submitted their highly individualized plans at this point.  Phase two (occurring this past year) involves departments working to refine and implement their particular plan.  The authors have worked at both the campus and departmental levels, as well as with individual faculty to facilitate these unit's efforts.  Their facilitation has spanned all eight UIUC colleges, and disciplines ranging from engineering to women's studies. The main emphasis for discussion will be the lessons learned (political as well as practical) from facilitating this large-scale effort toward implementation of highly individualized departmental outcomes assessment plans. Additionally, specific measures and methods of collecting outcomes assessment will be presented. This will be of interest to all involved in the very current trend of student outcomes assessment in higher education.

**Roundtable E (Second Rotation): Examining Group Problem Solving through Different Lenses**

Presenters: *Evaluating a Participatory Action Research Project: Tensions and Opportunities*, Gretchen B Rossman, University of Massachusetts at Amherst; Alex Risley Schroeder, Holyoke Community College; and Rose Sackey Milligan Peace Development Fund

This paper discusses the opportunities and tensions inherent in evaluating a federally-funded participatory action research project. This PAR project engaged adult learners in investigation and critical reflection about the impact of welfare reform, immigration reform, and the changing nature of work on their achievement of educational goals. Tensions occurred between the focus of the project and federally-driven evaluation purposes. In addressing these tensions, opportunities were created to enrich participants' critical reflection and refine project evolution. This paper chronicles these tensions and opportunities and furthers the discussion about the need for congruence in purpose between program and evaluation goals, especially in a PAR setting.

*Evaluating Team-based, Problem-solving Interventions in Organizational Settings*, Mark G Wilson, David M DeJoy & Robert J Vandenberg, University of Georgia

The purpose of this presentation is to discuss the evaluation of a comprehensive healthy work organization intervention. The intervention is based on a theoretical model that hypothesizes that organizational level action impacts one or more of three domains (job design, organizational climate, job future), which produce changes in employee perceptions and expectations, ultimately impacting organizational effectiveness (financial as well as health and well-being of employees). A data-driven, problem-solving intervention process was designed that starts with a healthy work organization audit examining the variables delineated in the model. From this, a team-based problem-solving process is used to develop, implement, and evaluate a customized action plan for the worksite for addressing the problems identified by the audit. The process is designed to incorporate multiple intervention strategies at multiple levels. This presentation will discuss evaluation methods for the intervention seeking means to incorporate new, innovative strategies.

**Roundtable F (Second Rotation): A Demonstration of Action Evaluation**

Presenters: Jay Rothman, Action Evaluation Research Institute
Victor J Friedman, Ruppin Institute

This demonstration describes 'action evaluation' a method for integrating knowledge production into the design, implementation, and assessment of social-educational programs. Action evaluation builds on an earlier vision of program evaluation as social experimentation but also incorporates concepts from conflict resolution and action science for promoting productive learning. These include program theories of action, participation, goal inquiry, stakeholder deliberation, team building, conflict engagement, and web-based research. This demonstration will present the basic concepts of action evaluation, walk participants through the action evaluation process, including use of the website, and provide illustrations from case studies. It will conclude with a discussion of both the advantages and disadvantages of action evaluation as well as current frontiers in its on-going development.

**SESSION 236: MultiPaper** **Room:** Kauai

Sponsored by the Topical Interest Group on Assessment in Higher Education

**Evaluating Instruction in Higher Education**

Chair: Christina S Ballantyne, Murdoch University

Presenters: *Evaluating Teaching: Toward a Pluralistic View of Good Teaching*, Daniel D Pratt & John B Collins, University of British Columbia

Current strategies for teacher evaluation assume a single, universal view of effective teaching and rely most often on student surveys as their primary means of gathering data. Yet, two decades of grounded research involving several hundred teachers of adults has revealed that instructors view their roles and responsibilities in relation to one or more of five teaching perspectives: transmitting content, apprenticing learners, developing critical thinking, nurturing self-concept, or transforming society (Pratt, 1998, Kember, 1997). Our research has operationalized these five perspectives into a 45-item Teaching Perspectives Inventory on which normative results are available for nearly 1,500 teachers of adults. Teachers are spread across the five perspectives, but most hold only one or two of these perspectives as ôdominant,ö suggesting the need for multiple rather than single, views of effective teaching. This paper presents and reviews these results, then discusses implications, showing how the TPI can be used for evaluations that help guide self-reflection on practice for evaluative purposes.

*Are Student Ratings of Instructional Quality Valid? A New Approach to a Frequently Asked Question Which Still Lacks an Answer*, Bettina Greimel, Vienna University of Economics and Business Administration

Student ratings of instruction are used throughout higher education. While hundreds of studies have generally concluded that student ratings are highly reliable and sound measures of teaching quality, one area has not been adequately addressed-the perceptions of students. How students approach and interpret rating forms strikes at the heart of the substantive aspects of construct validity. The current study delves into the underlying perceptions that students have related to ratings of instruction by drawing on both student self-reports (interviews & surveys) and students' public thought processes while completing the rating forms (think aloud protocols). The consequences of student ratings are high: tenure, promotion, and merit decisions are based partly, and sometimes solely, on results of student evaluations. The paper will highlight evidence related to the substantive validity of student ratings. It should be of interest to those who seek to maximize appropriate interpretations and encourage thoughtfulness of student responses. An abundance of research has been carried out over the past thirty years to determine whether student ratings of teaching are a valid indicator of instructional quality. Meta-analyses (such as Cohen 1981) have determined the correlation between student achievement and student ratings to be about $r = 0.47$. This major result supports the fact that students distinguish between teachers on the basis of how much they have learned in a course. However, the findings of a multitude of studies are mostly inconsistent - sometimes due to methodological differences - and the debate is still controversial. The aim of the paper is to summarize and discuss the main research results in this area as well as to present a recently designed methodological approach to investigating (once again!) this crucial question of educational evaluation. It also intends to overcome any common methodological problems and pitfalls of evaluation research.

*Utilization of Evaluation Methods and Techniques by Faculty*, David Laverny-Rafter, Minnesota State Colleges and Universities

In order to effectively evaluate their courses and programs, university faculty have shown an interest in administering a wide variety of evaluation and assessment techniques and methods. Today, faculty are demonstrating a willingness to move beyond traditional course evaluation surveys of students and to include more innovative assessment tools such as pre- and post-testing of student knowledge, external peer reviews, concept mapping, portfolio analysis, etc. This evaluation innovation is based on a desire for more meaningful formative evaluation data to improve teaching and for summative data to document learning to external stakeholders (e.g. administrators, accredidation organizations). This paper is based on research that examined evaluation techniques used by faculty in state universities and in community and technical colleges throughout Minnesota.

*Student Perceptions of Student Ratings: Construct Validity Revisited*, Christopher P Migotsky, University of Illinois

Student ratings of instruction are used throughout higher education. While hundreds of studies have generally concluded that student ratings are highly reliable and sound measures of teaching quality, one area has not been adequately addressed-the perceptions of students. How students approach and interpret rating forms strikes at the heart of the substantive aspects of construct validity. The current study delves into the underlying perceptions that students have related to ratings of instruction by drawing on both student self-reports (interviews & surveys) and students' public thought processes while completing the rating forms (think aloud protocols). The consequences of student ratings are high: tenure, promotion, and merit decisions are based partly, and sometimes solely, on results of student evaluations. The paper will highlight evidence related to the substantive validity of student ratings. It should be of interest to those who seek to maximize appropriate interpretations and encourage thoughtfulness of student responses.

Discussant: Mark E Troy, Texas A&M University