

## **Saturday, November 4, 8:00 A.M. to 9:15 A.M.**

### **SESSION 500: Plenary III**

**Room: Lanai**

#### **The Role of International Associations in Building Evaluation Capacity**

Moderator: Donna Mertens, Gallaudet University

Presenters: Mahesh Patel, African Evaluation Association (AfrEA)  
Jean-René Bibeau, Canadian Evaluation Society (CES)  
Ralph Straton, Australasian Evaluation Society (AES)

## **Saturday, November 4, 9:30 A.M. to 11:00 A.M.**

### **SESSION 501: Presidential Strand Alternative Format**

**Room: Lanai**

Sponsored by the Topical Interest Group for Graduate Students

#### **Cultivating the Grad-student Garden: Elders of the Field Share their Wisdom and Provide a Networking Opportunity**

Chair: Deborah L Wasserman, The Ohio State University

Discussants: Michael Quinn Patton, The Union Institute  
Daniel L Stufflebeam, Western Michigan University  
Laura C Leviton, Robert Wood Johnson Foundation  
Michael Scriven, Claremont Graduate University  
Joseph S Wholey, University of Southern California  
Linda Mabry, Indiana University

Building Evaluation Capacity, the theme of this year's AEA conference involves continual infusion of the field with new research, new ideas and new energy. The AEA Graduate Student Association TIG, comprised of AEA members currently in graduate degree programs, recognizes that students constitute one of the organization's principal mechanisms to build evaluation capacity. This panel/roundtable discussion features six evaluation "elders" who have contributed to the combined wisdom of the field will speak about the topic of their choice related to tools, skills, attitudes and values future evaluators need to cultivate. Forty minutes of presentation will conclude with a 10-minute open discussion followed by a 40-minute roundtable session whereby graduate students (and others) will meet in small groups with the speakers. Grad students will also have the opportunity to request contact/networking with evaluation professionals not present on the panel. Because of the interest this session will generate and its relationship to the conference theme, we will be nominating it for the Presidential Strand. A fairly large room will be laid out with chairs around six discussion tables and a head table. "Elders" will begin at the head table and proceed to their assigned discussion table. At each will be a sign-up sheet whereby grad students who need further contact can list their name, research interest and specific request of "elder" (e.g. meeting or emailing re. a specific topic). At each table will be an additional sign-up sheet on which grad students can list names of other evaluation professionals they are needing to contact, along with topic of interest and specific request, and the session coordinator/chair will facilitate the contact.

### **SESSION 502: MultiPaper**

**Room: Akaka**

Sponsored by the Topical Interest Group on Distance Education and Other Educational Technology

#### **Evaluating the "Online" Learning Experience**

Chair: Tara Knott, Evaluation Resources Inc

Presenters: *Utilization-focused Needs Assessment: A Case Study of Adult Learners' Web-based Distance Education Needs*,  
Denise J Flowers, & Joyce Woodruff, University of South Alabama; and Susan Tucker, E & D Associates

Distance education has moved from the deprived stepchild of the 70's to the prized date with the Prince. To meet the increasing demand in higher education, the dynamic needs of disparate stakeholders must be heard and transformed into viable programs. What are important questions to include in a utilization-focused needs assessment? Experts (i.e. evaluators, adult educators, instructional designers, FIPSE-LAAP recipients, and web vendors) were asked to generate critical questions and needs in order to design and deliver successful web-based instruction during planning

and start-up phases in a college. On-campus participants included current web-adult learners, potential adult-web students, adult students who have dropped out, faculty, staff, and instructional designers. Campus responses were congruent with those revealed by web practitioners, however additional practical and idiosyncratic issues emerged regarding student access, readiness, and avoid attrition. These findings will be shared as well as a matrix of "holistic" evaluation questions which receive wide support across stakeholders.

*Evaluating Online Learning Initiatives from Multiple Perspectives*, D Michelle Hinn, Alaina Kanfer & X Christine Wang, National Center for Supercomputing Applications

Published evaluations of online courses and programs have largely focused on including faculty and student perspectives but rarely include the perspectives of the support staff a key stakeholder group. This presentation will focus on an evaluation, aimed at including the perspectives of students, faculty, and staff, of a graduate course that integrated onsite and offsite students and prototyped an Internet-based two-way synchronous audio/video communications system. The evaluation approach was qualitative and included onsite observations, online chat transcript analysis, and post-course semi-structured interviews. The presenters will discuss the results of the evaluation with regard to the experiences of the instructors, support staff, and students who participated in the course, highlighting the convergent and divergent perspectives within and between each group. The implications of these results will also be discussed, which include the challenges and realities of online course management as well as inter- and intra-group communication issues that can affect the successful implementation of online learning initiatives.

*Lessons from an Evaluation of Eleven Web-based Courses*, David D Williams, Brigham Young University

During fall 1999, an evaluation was conducted of eleven web-based university courses, all designed for completion in one semester. Based on Patton's Utilization Focused Evaluation and Stufflebeam's CIPP models, a variety of players were invited to participate in clarifying stakes they held in the courses. Several participants collaborated to help gather and analyze data. Questions addressed technical feasibility, instructional quality, individualization issues, workload considerations, completion and achievement rates, student profiles, and attitudes of students and faculty toward the experience. This presentation will summarize the evaluation processes used, the roles participants played, the findings of the study, how the participants used the results, and plans for modifying and continuing the study of these and other courses. Concerns as well as accomplishments associated with the participant-oriented evaluation approach used will be explored in light of growing relevant literature. Implications for building participant-oriented evaluation into online course development processes will be considered.

*An Illustration of the Need for Situational Responsiveness in Technology-related Evaluations*, David D Williams & Todd C Wilson, Brigham Young University

The proliferation of technology is creating the need for greater situational responsiveness in evaluators. While evaluations that deal directly with technology are not the only types that call for adaptation, they are becoming increasingly more common. Traditional methods of evaluation can be effective, but may need to be extended and reconsidered in light of technological advancements. Greater emphasis must be placed on situational adaptation, often calling for expanded roles of evaluators. Evaluators must become effective inquirers able to quickly familiarize themselves with current technologies, the effects they may have on evaluations, and the changing interests of stakeholders. We will illustrate the need for situational responsiveness in technology-oriented evaluations by presenting the case of such an evaluation that we performed. In describing the conditions of our evaluation we will briefly discuss some of the ideas of Michael Patton. We will then present findings and recommendations based on our experience.

---

**SESSION 503: MultiPaper**

**Room: Koko**

Sponsored by the Topical Interest Group on Assessment in Higher Education

**Multiple Perspectives on the Evaluation of an Undergraduate Course in Psychology: A Comparison of Wholey, Cronbach, and Chen**

Chair: Jennifer M Camacho, University of Illinois at Chicago

The purpose of this session is to compare the methodologies and frameworks of three evaluative theorists: Wholey, Cronbach, and Chen. Each presenter simultaneously evaluated an undergraduate writing course in psychology from the perspective of one of these theorists and will present her findings and her recommendations. Presenters will also discuss the substantial differences in these outcomes as a function of which theorist was chosen.

Presenters: *An Evaluability Assessment of a Psychology Writing Course*, Tina Taylor-Ritzler & Jennifer M Camacho, University of Illinois at Chicago

According to the methods proposed by Joseph Wholey, an evaluability assessment of an undergraduate writing course in psychology was conducted. During an evaluability assessment a program's ability to be meaningfully evaluated is determined. The present evaluability assessment involved reviewing program documentation and conducting semi-structured open-ended interviews with faculty who made administrative decisions regarding the writing course to determine program intent. These data were used to develop a model of how the program was supposed to work. Interviews were also conducted with instructors who had taught the writing course in order to determine program reality. Data obtained from these interviews were compared to the program intent model. Results revealed fundamental discrepancies between the stated intent of the program at the administrative level and the way instructors actually conducted the course. Evaluability assessment results, recommendations that were made to the department, and other applications of Wholey's evaluation approach will be presented.

*An Evaluation of a Writing Course: Cronbach's Perspective*, Erin Hayes, University of Illinois at Chicago

An evaluation of an undergraduate writing course was conducted from the perspective of Lee Cronbach. Cronbach advocated investigating the process by which programs were successful: what were the mediating processes by which programs achieved their goals? This evaluation was the first step in a Cronbach model, as he suggested conducting a number of smaller evaluations in order to reach the central goal of generalizing to unobserved persons, settings, treatments, and outcomes. Classroom observations and interviews were conducted in an attempt to begin uncovering the mediating processes that facilitated student success or failure. Semi-structured interviews with a number of open-ended questions allowed interview participants to discuss whatever course issues they found relevant. Results revealed that administrators, professors, and teaching assistants each had different visions about the class. Further, these discrepancies were acknowledged by students, and they felt it impeded their progress. Specific recommendations that were made for both the classroom and course policy, as well as the limitations of Cronbach's approach, will be discussed.

*An Evaluation of An Undergraduate Course: Chen's Perspective*, Bianca Wilson, University of Illinois at Chicago

Chen (1983) advocates a theory-driven approach to program evaluation. This approach to program evaluation is distinct within the history of program evaluation inasmuch as it calls for a movement away from the experimental theory of evaluation practice as well discouraging reliance solely on the kind of experiential, pragmatic knowledge that constitutes much of what is known about program evaluation. The goal of this program evaluation project was to determine and clarify the normative program theory underlying the Psychology Department's effort to teach writing in the discipline. Data collection methods included examining archives, course catalogs, and syllabi, and conducting interviews with Department of Psychology Executive Committee members and faculty who taught the writing course. The resulting program theory will be presented as well as comments regarding the appropriateness of evaluating the program from Chen's theoretical framework.

---

#### **SESSION 504: MultiPaper**

**Room:** Waimea

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

#### **Combining Evidence**

Chair: Georg E Matt, San Diego State University

Presenters: *Interpersonal Conflict Management Programs for Youth: A Synthesis of Treatment-effectiveness Research*, Wendy M

Garrard, Vanderbilt University

The tide of public concern about aggression and violence among youth has risen dramatically in recent years. Consequently, decision-makers in schools and other community-based youth services are rushing to implement programs assumed to reduce interpersonal conflict. Because there is relatively little practical information available about the effectiveness of conflict management programs, it is an important time for evaluators to learn about the characteristics which differentiate the programs that are most, and least, effective. This meta-analysis examines over 30 years of treatment effectiveness research on programs aimed at interpersonal conflict skills such as anger management, impulse control, and social problem solving. The important characteristics of these programs are summarized, with an emphasis on what can be learned that promotes useful evaluation for this genre of programs.

*Qualitative Meta-analysis: A Systematic Method for Evaluating and Synthesizing Bodies of Literature*, Jennifer L Mersman & E Jane Davidson, Claremont Graduate University

To date, two well-known methods exist for synthesizing large bodies of literature. A literature review involves the use of professional judgment to evaluate individual studies, and a subsequent synthesis (conclusion) step, also based on the reviewer's judgment, using a heuristic that is seldom made explicit. In contrast, meta-analysis uses a very systematic method of combining the results of quantitative research in order to reach overall conclusions about actual effect sizes in various populations, and under various conditions. This paper will present a third option that builds on the best points of the other two. 'Qualitative meta-analysis' is a methodology for systematically evaluating and then synthesizing both quantitative and mixed-method research. We will present a list of criteria to evaluate research quality, and then demonstrate how these can be applied in a step-by-step method that results in an overall conclusion about the quality of the body of research.

*Getting to the Bottom Line: A Method for Synthesizing Findings from Mixed Method Program Evaluations*, Andrew McConney, Andy Rudd & Bob Ayres, Western Oregon University

Evaluators concerned more with pragmatics than with competing epistemological paradigms have brought multi- and mixed method evaluations into common practice. Program evaluators commonly use multiple methods and mixed data to capture both the breadth and depth of the evaluand and to strengthen the validity of findings. However, multiple or mixed methods may yield incongruent results, and evaluators may find themselves reporting conflicting findings to program staff, policymakers, and other stakeholders. Our purpose is to offer a method for synthesizing the findings from multi or mixed method evaluations to reach defensible evaluation (primarily summative) conclusions. The proposed method uses a set of criteria and analytic techniques to assess the worth of each data source or type, and what each says about program effect. Once on a common scale, simple math allows synthesis across data sources or types. The method should prove a valuable tool for evaluators across a range of experience.

*Using Pretest Effect Size to Measure Selection Bias in Meta-analysis*, William R Shadish, University of Memphis

Meta-analysts frequently include data from nonrandomized experiments, but such data can yield biased effect size estimates. This paper shows how pretest effect size can measure selection bias in nonrandomized experiments, illustrating with data from a meta-analysis of psychotherapy. Pretest effect sizes are computed on outcome variable(s) taken before treatment begins. Pretest effect sizes should average zero and be homogenous in randomized experiments, and they are in our data. If nonrandomized experiments are unbiased, their pretest effect sizes should be similar; but they are not in our data. In fact, the difference in effect sizes between nonrandomized and randomized experiments at pretest is nearly equal to their posttest difference. Further, pretest effect sizes in nonrandomized studies where clients self-select into treatment are significantly lower than in five studies where clients were put into conditions by another nonrandom mechanism; and the selection bias is plausibly explained. Implications for meta-analyses that include nonrandomized experiments are discussed.

Chair: Kim Sabo, Innovation Network NYC

This ninety-minute panel discussion will describe how Innovation Network's core values and evaluation strategies are used in four different contexts in order to build organizations' capacity. The panel will include four presentations and a discussion (i.e., questions and answer period). Each presenter will illustrate specific implementation strategies and lessons learned, discussing practical issues in strengthening and evaluating capacity building. Individual presentation topic areas will include:

1. Evaluating Technical Assistance for Building Organizational Capacity: Linking Literature and Practice
2. The Robert Bowne Foundation: Building grantee evaluation capacity through the use of participatory evaluation coaches.
3. Enhancing Organizational Effectiveness Through the Learning Circles Project: What have we learned?
4. Building Evaluation Capacity Through Cyber Space: Is it possible?

Besides learning practical approaches and techniques, the audience will also participate in an active discussion about capacity building, identifying best practices, barriers and implications for the future. Based on presentations, the audience will discuss how and to what extent evaluation practice can become part of ongoing organizational development.

Presenters: *Evaluating Technical Assistance for Building Organizational Capacity: Linking Literature and Practice*, Mary Ann Scheirer, Independent Consultant and Colette Thayer, Innovation Network

This paper will explore how 'capacity' is used in the organizational literature as well as its meaning for technical assistance providers in the nonprofit sector. Based on InnoNet's study of Best Practices in Nonprofit Capacity Building, this paper will describe issues faced by technical assistance providers and recipients, such as how needs are diagnosed and what criteria are used to evaluate capacity building. The paper will draw upon InnoNet's work evaluating a national environmental organization whose mission is to increase the effectiveness of regional, state, local, and grassroots organizations working on environmental issues. This paper will provide evaluators with an understanding of capacity building within the literature as well as on the ground.

*The Robert Bowne Foundation: Building Grantee Evaluation Capacity through the Use of Participatory Evaluation Coaches*, Kim Sabo, Innovation Network NYC

This paper will describe the Robert Bowne Re-imagining Afterschool Initiative and its capacity building outcomes. This four-year initiative has sought to build the capacity of its grantees by providing them with literacy specialist, program mentors and participatory evaluation coaches (P.E. Coaches). Further, the Robert Bowne Foundation has supported capacity building within and between P.E. coaches by providing training and supporting ongoing professional development. This paper will draw on the meta evaluation of this initiative and pay particular attention to the capacity building efforts and outcomes effected by P.E. Coaches. Margo Bailey, Ph.D.

*Enhancing Organizational Effectiveness Through the Learning Circles Project; What Have We Learned?* Margo Bailey, Independent Consultant

This paper will address the challenges encountered and successes achieved through The Learning Circles Project (TLC Project), a nonprofit capacity building effort led by InnoNet and funded by the Fannie Mae Foundation. Through formal and informal participatory training sessions, peer to peer learning opportunities, mentoring, and referrals, TLC Project seeks to enhance the organizational effectiveness of participants. The paper will include findings from InnoNet's evaluation of TLC Project based on data collected through diagnostic tools, site visits, training evaluation forms, and focus groups.

*Building Evaluation Capacity through Cyber-space: Is It Possible?* Colette Thayer, Innovation Network

InnoNet's interactive web-site is designed to guide organizations and evaluators through the development of program, evaluation, and fundraising plans as well as provide them with a toolkit of data collection instruments and analysis advice. Evaluators will learn how InnoNet's web site can serve as a resource for them as well as their clients through

the paper's description of the site and InnoNet's vision for its future. In addition, the paper will share lessons learned throughout the site's development phase and will provide an analysis of feedback collected from users.

---

**SESSION 506: Skill-building Workshop****Room: Hilo**

Sponsored by the Topical Interest Group on Independent Consulting

**Bids and Budget Proposals for Independent Consultants**

Facilitator: Donna Blair, Camp Blair Consulting Inc

How do you price your work? Are you covering all of your costs? Will your government budget proposal meet federal requirements? This session will cover different methods of setting prices and some important information for consultants hoping to bid on federal contracts.

---

**SESSION 507: Panel****Room: Puna**

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

**Lessons Learned from Multiple Cross-site Evaluation Studies**

Chair: J Fred Springer, Evaluation Management and Training Associates Inc

This Presentation will present lessons learned from multiple federally-funded cross-site evaluation studies conducted by EMT Associates and Macro International. The Presentation will consist of brief overviews of each cross-site study and inter-related presentations by principals on the various Cross-Site Evaluation teams. More specifically, presentations will address issues related to the areas of: a) outcome data; b) process data and program level analysis; and c) dosage data. The research base for these presentations are multi-site evaluation studies that all combine data from self-report surveys, intensive site visits to participating programs, and records of program contact for each youth. Drawing from our collective implementation experience, examples of challenges in the identified topic areas and the ways in which they were resolved are identified and discussed. Many of the data collection, management, and analytic procedures adopted by the study teams provide important experience and potentially useful methods and techniques for other prevention evaluators.

Panelists: *Project: Youth Connect: Lessons Learned from the Evaluation of a National Mentoring Program*, Rachele C Espiritu, Evaluation Management and Training Associates Inc and Cheryl Bildner, Macro International

The Project: Youth Connect (PYC) cross-site study is designed to determine the effectiveness of a mentoring/advocacy program for youth (9-15 years old) and their families. Approximately 3,000 program and control youth 9-15 years of age in fifteen program sites will participate in this study. The main objectives include the prevention, delay, or reduction of substance abuse among youth and increases in school bonding/academic performance, family functioning/bonding, and life management skills. The study collects outcome data from youth, parents, teachers, and school records; data on utilization of services (dosage) for each program participant; and data on program level variables through site visits and monthly telephone interviews. Examples of strategies used to create common data collection instruments, dosage forms, and site visit protocols; approaches to data management and monitoring; and analytic techniques employed on the outcome, dosage, and site visit data will be presented during this session.

*Common Purpose, Common Measures, Diverse Programs: Lessons Learned From the Parenting Adolescent Cross-site Study*, Shari Golan, SRI International and Mary Nistler, Evaluation Management and Training Associates Inc

The Parenting Adolescents (PA) Cross-Site Study examines whether and how intervention programs improve the lives of low-income pregnant and parenting teens and the factors that promote or hinder those outcomes (e.g., TANF participation, social support). Key outcomes being examined include alcohol and other drug use, school completion and achievement, occurrence of repeat pregnancies, and parenting stress. Common baseline and follow-up surveys are being administered to approximately 3,000 program and comparison youths in 10 programs. Over a two-year duration, study participants will be surveyed four times. Dosage data are being collected so that differential exposure to program activities can be assessed. Data on program level variables have been collected through site visits and program surveys, allowing the identification of program characteristics that contribute to positive outcomes. This

presentation will focus on the strategies used to create common data collection instruments and to conduct meaningful analyses given the diversity of programs.

*Children of Substance Abusing Parents (COSAPs) Study: Lessons Learned from the Evaluation*, Jack Hermann, Macro International

This multisite evaluation is designed to assess the effectiveness of intervention strategies for children and their substance-using parents. The study involves the local evaluation and program staff from 14 study sites. Each site targets a specific age group of youths and provides certain core interventions to the children and their parents. Program duration and intensity vary across the study sites. The study collects 1) outcome data from participants and comparison subjects, 2) dosage data (service utilization) on each participant, and 3) program-level data. Participatory evaluation principles can be applied to multisite studies like this. Variation in local program design, context, and ages of youths require flexibility and complexity in the cross-site evaluation design. Many design features, including outcome and dosage measures, were determined through stakeholder consensus. However, the consensus process resulted in design complexities that have important implications for data analysis.

*CSAP's National Cross-site Evaluation of High Risk Youth Programs: Methods, Challenges, and Findings*, Elizabeth Sale & J Fred Springer, Evaluation Management and Training Associates Inc

The Cross-site Evaluation of the High Risk Youth Demonstration Program (HRY) is a five-year study that examines selected high risk youth, female adolescent and replication initiatives funded by CSAP in 1994 and 1995. A total of approximately 50 programs will be evaluated, with approximately 6000 participants and 4000 comparison youth. Data Analysis focuses on assessing the effectiveness of High Risk Youth interventions in attaining intended outcomes generally, and in establishing effectiveness of alternative intervention strategies across programs. Detailed site visit and program information protocols support independent measures of intervention design and implementation strength that contribute to cross-level statistical analyses. Detailed dosage data documenting contact time by category of prevention strategy further elaborate outcome analyses. The growing sophistication of high risk youth programs and strength of this evaluation design hold promise for making this evaluation the strongest scientific study of prevention effectiveness to date.

Discussant: J Fred Springer, Evaluation Management and Training Associates Inc

#### **SESSION 508: MultiPaper**

**Room: Kohala**

Sponsored by the Topical Interest Groups on Collaborative, Participatory and Empowerment & PreK-12 Educational Evaluation

#### **Beyond the Schoolhouse: Evaluating School-community and School-agency Collaborative Initiatives**

Chair: Richard Phillips, Eastern Washington University

Presenters: *A Stakeholder Self-assessment Strategy for Milwaukee Wisconsin's 21st Century Community Learning Centers Initiative*, Daniel V Folkman, University of Wisconsin at Milwaukee and Susan Stuckert, Milwaukee Public Schools

The US Department of Education is sponsoring the 21st Century Community Learning Centers (CLC) initiative. Milwaukee has received federal grants supporting 19 centers located throughout the central city. The Milwaukee CLC initiative is being implemented through a citywide network of over 70 stakeholder groups called the School Community Integrated Services Network (SCISN). This paper outlines Milwaukee's collaborative evaluation strategy. The case study documents how the SCISN Evaluation Committee is guiding a self-assessment process involving data collection, analyses, and development of action steps to improve program performance by stakeholders at the 19 CLC sites. The role of the CLC management support team is discussed in terms of its ability correlate CLC activities with student academic achievement while also addressing issues of sustainability beyond federal funding. The paper concludes with a discussion of the multiple roles played by the evaluator, and the challenges faced when implementing a collaborative stakeholder self-assessment process.

*School and Community Collaboration: Evaluating a Partnership To Improve Student Health and Wellbeing*, Pamela M Jakwerth & B A Laris, Education, Training, and Research Associates

School and community partnerships are becoming a dominant feature of violence prevention and health promotion programs. It is critical to systematically evaluate these partnerships and define effective models. ETR Associates is the local evaluator for the federal Safe Schools/Healthy Students Initiative being implemented in the San Francisco Unified School District (SFUSD). One component of the Initiative is the formation of a school and community partnership to improve the health and wellbeing of students. The SFUSD interagency committee, comprised of representatives of public agencies and the school district, serves two functions. It works as a planning body to assist SFUSD staff in the design and implementation of the Initiative. In addition, the committee will help link schools to professional development, services, or resources that address school needs. ETR's evaluation of the partnership uses a combination of data sources, including surveys, focus groups, records, meeting notes, and observations. The evaluation focuses on fidelity, quality, satisfaction, and impact.

*So Where Do We Go From Here? Evaluating a Partnership between Public Schools and a Department of Health and Welfare*, Richard Phillips, Eastern Washington University

This paper describes learnings resulting from evaluating a partnership between two systems under pressure to show effectiveness defined in ways that seem to be mutually exclusive. In this case, 53 public school districts and a state department of health and welfare formed a partnership to provide service to children and families at risk of child abuse or out-of-home placement. The process of designing and carrying out an outcome evaluation of this partnership resulted in a number of poignant learnings:

- a) common evaluation terms have a politically laden context and meaning;
- b) system norms play a significant role in how partnership programs are actually evaluated
- c) a child-centered focus has the potential to form a bridge between systems

This evaluation experience has a number of implications for evaluation practice. First of all, it helps clarify the potentials of empowerment evaluation, and starts to redefine the types of goals that participatory evaluation might plausibly seek when working with multiple systems. In addition, it highlights the potential of considering the clients themselves, no matter what age, as agents of change across family, school, and social service systems. The relationship of this paper to the conference theme is straightforward: The capacity of evaluation to impact systems in partnership seems to be in direct proportion to the ability of the evaluators to build trusting relationships between the systems by focusing on the intervention progress of individual clients.

*Conducting a Participatory Evaluation of Safe and Drug Free Schools and Communities Program in the Austin Independent School District (AISD)*, Martha Doolittle & Marci Ryan, Austin Independent School District

For a number of years, school district staff have been conducting participatory evaluations of drug and violence prevention programs. A team of program leaders guides campus staff in their efforts to reduce the use of drugs and violence and increase protective factors among students. The school district's evaluation is meeting both federal, state, and local reporting requirements and is a resource for campus-based service delivery. A team approach helps facilitate district- & campus-level collaborative program design, delivery and evaluation. Through the use of developed methods, both campus needs and the grant objectives are being addressed. Multidimensional approaches to program evaluation are utilized that are inclusive of all stakeholders. Theory and research are brought into meaningful practice in schools. Shared evaluation roles present an example of new paradigms in evaluation

practices. The team tries to balance the need to report with the need to be a resource to constituents. As district-level program evaluators, we are building evaluation capacity among district-level and campus-level staff using the following:

- \* fostering a learning community
- \* information dissemination
- \* participatory evaluation
- \* tool building for program planning, implementation and evaluation
- \* opportunities for feedback.

Discussant: Abraham H Wandersman, University of South Carolina

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

**Model-guided Analyses in a National Multi-level Evaluation Study: The ACCESS Program**

Chair: Frances L Randolph, Massachusetts Institute of Technology

The Center for Mental Health Services in HHS has been conducting a multi-site, multi-level evaluation demonstration study to determine successful community-based strategies for impacting homelessness among persons with serious mental illnesses. The study, Access to Community Care and Effective Services and Supports (ACCESS) Program, was designed to test the hypothesis that making service systems more integrated would result in better clinical outcomes for homeless persons with serious mental illness. The study was conducted in 18 communities across 9 states and involved systems-level, site-level, and client-level evaluations. This panel will present the final results of this 5-year study and discuss the model-guided analyses that were used.

Panelists: *Measuring Strategies for Systems Integration Change*, Joseph J Coccozza, Policy Research Associates

The evaluation of the ACCESS Program sought to collect and analyze both client outcome and interorganizational data that quantified the extent of systems integration. Absent from the general body of research on systems integrations has been the attempt to measure the practical strategies implemented by communities in their attempt to better integrate their service delivery systems. The purpose of this presentation is to describe how these strategies were defined, measured, and documented. The measures and procedures for collecting and scoring them will be discussed along with a brief description of the findings on the type and nature of strategies employed at the ACCESS sites.

*Can Service System Integration Be Improved? Evidence from the ACCESS Program*, Michael O Calloway, University of North Carolina at Chapel Hill

This presentation focuses on the first question underlying the ACCESS program: Can social service delivery systems be integrated? The data consist of implementation indices of system change strategies as well as interorganizational network indices that reflect the actual levels of system integration achieved by the 18 ACCESS communities. The network indices are based on data from over 1,000 organizations. Both the implementation and system integration measures are available at 4 time points (1994, 1996, 1998 and 2000). Mixed models for site and organization levels of analysis with time dependent covariates are employed. Policy and practice implications will be highlighted along with directions for further research.

*Impact of Efforts To Improve Outcomes for Homeless Persons with Serious Mental Illnesses through Service Systems Integration*, Robert Rosenheck, Northeast Administration Evaluation Center

This study evaluated the relationship between efforts to improve the integration of service systems and outcomes among homeless people with serious mental illnesses. The ACCESS Program provided funds to nine sites to implement strategies designed to promote systems integration. Integration sites, along with nine control sites, also received funds to support outreach and assertive case management to 100 clients per year. Standardized outcome and service use data were obtained at baseline, 3 and 12 months from 7,224 clients across 4 annual cohorts at the 18 sites. Results will be presented comparing client outcomes between integration and comparison sites across the 4 cohorts.

*Sectoral Analyses of the Effects of Systems Integration in the ACCESS Program*, Joseph Sonnefeld, ROW Sciences Inc

This presentation reports on secondary analyses of the relationship between service system integration and client outcomes in the ACCESS Program. Analyses will be presented that examine ties among the subset of organizations that provide services that may be most relevant to improved outcomes for the subgroup of clients with self- or clinician-reported substance abuse problems. System-level variables include egocentric interorganizational network measures using the sectoral matrix as the analysis dataset. Individual control variables will be included in a multi-level model.

**SESSION 510: Alternative Format****Room: Ewa**

Sponsored by the Topical Interest Group on Health Evaluation

**Building Community Capacity to Evaluate**

Chair: Kathleen M Hebbeler, SRI International

Presenters: Stacie S Cherner, SRI International

Dorothy Meehan, Sierra Health Foundation

Lori Juszak, Grassroots for Kids

Shirley DickKard, Yuba Community Collaborative for Healthy Children

Community Partnerships for Healthy Children (CPHC) is a Northern California initiative designed to improve the health and well-being of young children through collaboration. One objective of the initiative has been to build community capacity to plan, implement, and evaluate actions undertaken to address issues related to children and families. An experimental aspect of the initiative is that each community is responsible for its own evaluation. The funder, the Sierra Health Foundation, contracted with SRI International to conduct the overall evaluation of CPHC and to provide technical assistance to each community around defining outcomes for children and evaluating their strategic plan. These grassroots efforts at evaluation were challenging, frustrating, and even overwhelming but they also produced some good quality evaluations with useful findings. The funder, the evaluator/technical assistance provider, and representatives of two community collaboratives responsible for their evaluations will share their perspectives on this experiment in evaluation capacity building. Another focus of this panel/demonstration will be how to assist non-evaluators to actively participate in or conduct evaluations. Providing technical assistance on evaluation requires that evaluators add a new set of skills to their repertoire. For the past five years, SRI has helped 26 community-based collaboratives identify child and family outcomes and corresponding indicators; formulate questions about program implementation; and collect, analyze and report findings. This intensive capacity-building effort involved training sessions, materials development, and individual on and off site consultation. By serving as a technical assistance provider, we learned a lot about how to communicate about evaluation, what is reasonable to expect, where the pitfalls lie and how to overcome at least some of them. We will present techniques that we found to be effective, share the training materials developed, and discuss important considerations related to evaluation capacity building. We will also highlight problems that could not be satisfactorily addressed.

**SESSION 511: Alternative Format****Room: Honolulu**

Sponsored by the Topical Interest Group on Evaluation Use

*(The second part of this session is to be held on Saturday at 11:10 in Honolulu.)***Evaluative Inquiry and Transformative Learning: The Theory and the Practice**

Facilitators: Rosalie T Torres, Developmental Studies Center

Hallie Preskill, University of New Mexico

This alternative format, extended session will address both the theory and the practice of conducting evaluative inquiry within organizations to facilitate transformative learning. The first half of the session will cover three topics: (1) the role and importance of transformative learning in the workplace, (2) how an organization's infrastructure mediates transformative learning, and (3) the specific processes and phases of evaluative inquiry which can be used to facilitate transformative learning. The second half of the session will address the practical aspects of implementing evaluative

inquiry within an organization. Participants will have the opportunity to analyze and problem solve with respect to these aspects of their own evaluation practice: (1) diagnosing the organization's infrastructure, (2) using a clinical approach which is both incremental and iterative; (3) articulating, formalizing, and making time for learning processes; (4) using responsive methods; and (5) spanning traditional boundaries between evaluation work and program work. In order to address both theory and practice, and provide session participants an opportunity to examine and discuss their own evaluation practice, we are requesting that this session consist of 180 minutes in total, and be extended over two 90-minute sessions. As proposed, this extended session is a combination of a demonstration session and a skill-building session. While participants will get the greatest benefit by attending both sessions, those who can attend only one will still take away useful information.

## **SESSION 512: Roundtables**

**Room:** Kahuku

### **Perspectives on Teaching, Learning and the Contexts of Education**

*(This session includes two 45-minute rotations of roundtables. The Host will ask the tables to rotate at 10:15.)*

Host: To be announced

#### **Roundtable A (First Rotation): Museum and Parks Evaluation Initiatives**

Presenters: *The Common Ground of Cluster Evaluation: The Role of Instrumentation in the P.A.R.K.S. National Evaluation Project*, Elizabeth J Hall & Emmalou Norland, The Ohio State University

The author presents the instrumentation used to collect evaluation data during site visits to several U.S. National Parks as part of the P.A.R.K.S (Parks as Resources for Knowledge in Science) project, a national evaluation of thirty-six collaborative, K-12, park-school science education programs. The purpose of this presentation is to share with the audience how instrumentation was utilized to capture the unique challenges and successful strategies of each park program, as part of a cluster evaluation. This presentation is a follow-up to last year's Panel Discussion introducing the same project, and a precursor to next year's presentation of results.

*Museum Explorers Year V: Beyond Models of Excellence*, Yuqiu Cheng, Georgia Hughes, Arron Service, Chad Briggs & Jack McKillip, Southern Illinois University at Carbondale

The Museum Explorers (ME) Project at the Southern Illinois University Museum is part of a statewide program known as Museums in the Classroom (MIC). MIC was initiated five years ago as part of a \$46.2 million allocation by the Illinois General Assembly to support efforts to improve student learning and performance through the use of technology. To date, over 400 Illinois elementary and secondary schools have participated in MIC. Applied Research Consultants (ARC) has been employed by SIU Museum to evaluate the ME project for five years. In the first three years, a one-group pretest-posttest design was utilized to assess gains in computer skills, interest in art/history, and attitudes toward learning for both students and teachers. In the fourth year, the evaluators adopted multiple case study methodology to find out how excellence of outcomes was related to characteristics of implementation. As a result, the evaluators developed a "Models of Excellence" for the ME Project, which could be applied to future ME Project participants. This year (the fifth year), ARC continues to utilize multiple case study methodology to examine and refine last year's "Models of Excellence" combining both qualitative and quantitative research methodologies. Subject Matter Experts (SME) evaluate the final products of the ME Project participants by answering the questions on the new scales developed by ARC to determine if students are utilizing reading and math. SMEs compare the products of this year and corresponding products of previous years so that evaluators can determine if students' reading and math skills have been improved because of the ME Project. There are three implications of the current evaluation for developing and advancing evaluation capacity and utilization. First, factors contributing to success (excellence) in the ME project are found so that a generalizable model of excellence can be used to guide future participation of schools. Second, the current evaluation determines whether the ME Project can benefit students by improving their reading and math skills, and which skills and how they can be improved. Third, Illinois State Board of Education can have information with regard to MIC on what learning standards it can help students meet.

#### **Roundtable B (First Rotation): Accessible Evaluation: Resources for Conducting Program Evaluation**

Presenters: *Hands-on Evaluation Workshops: Building Capacity for Cooperative Extension in Arizona*, Allison L Titcomb, University of Arizona

The Institute for Children, Youth and Families at The University of Arizona has, as one of its main purposes, an Evaluation Core. The ICYF offers training workshops in evaluation for Cooperative Extension Faculty and others (e.g., Tools for Evaluation Planning, Survey Development, Healthy Adolescents/Healthy Relationships). These workshops have offered examples of evaluation planning based on logic models, CYFERNet State Strengthening Guide, and other general models for evaluation. The purpose of the workshops is to promote integrated program evaluation across multi-county sites, to improve extension agents' knowledge and skills in the area of program evaluation, and to strengthen the cooperative activities engaged in by the participants. This roundtable discussion will share sample workshop handouts (e.g., logic models, key questions for evaluation planning), description of web-based activities used for follow-up, as well as lessons learned from the delivery and expectations of these workshops as methods for building capacity for multi-site evaluation.

*Outcome Measurement Toolkit for Resident Camps*, James A Riedel, Girl Scouts of the USA & Melanie A Hwalek, SPEC Associates

Outcome measurement is an important part of program evaluation. For decades, nonprofit organizations have conducted output evaluations. Though evaluation data are useful for program development and improvement, it took the catalyst of funders' recent calls for outcomes to move many organizations to collect it. The current project resulted in the development of instruments for measuring, across sites, initial and intermediate outcomes of Junior Girl Scout resident camp. Through a collaborative effort, an outcomes measurement "tool kit" was developed by an external evaluator, with continuous involvement of program administrators/deliverers and internal evaluators. Although resident camps are attended by thousands of children annually, few evaluation instruments were previously available for measuring the specific impact of camp experiences. This venture has relevance for increasing evaluands' capacity in developing evaluation skills and instruments and producing sustainable evaluations. The toolkit's development process, contents, and expected effects on changes in camp program activities will be described.

#### **Roundtable C (First Rotation): Longitudinal Evaluation of Education Programs**

Presenters: *A Longitudinal Evaluation of the Science Research in the High School Program*, Dean T Spaulding & Dianna L Newman, State University of New York at Albany

This paper will present methods, results, and issues related to a three-year, longitudinal evaluation of a Science Research Program in the High School. This instructional method, known as inquiry-based science, stresses the interactive relationship between the students working through the scientific process. As part of this evaluation, a longitudinal methodology was established that documented growth over the course of three years. Variables examined for growth included: teacher perceptions of the program and student outcomes, students perceptions of growth and skills acquired, and administrators documentation of resource changes along with systemic change. In addition, individual student portfolios were examined over the three year time period as a means of assessing higher level program solving and inquiry-based skills. Presented in this paper is also a discussion of methodological issues surrounding the documentation and procedures used in gathering and examining both qualitative and quantitative data from a variety of sources throughout the longitudinal process of the evaluation.

*The Graduate Student Experiences Assessment Project: Establishing a Longitudinal Evaluation System*, Claudia Figueiredo, John Berry & Emmalou Norland, The Ohio State University

A utilization-focused evaluation process, aimed at assessing graduate students' experiences, was established by the Student Services and Graduate Studies Office, School of Educational Policy and Leadership at Ohio State University (OSU). Throughout the 1999-2000 academic year, a stakeholder team was formed and designed a longitudinal evaluation system. This group was composed of students, faculty, and staff representing the several sections of the School, as well as the College of Education and the OSU Graduate School. This presentation will describe the development of the longitudinal evaluation system, as well as the components of this system: instruments of data collection, basic timeline and uses of information by stakeholders. The longitudinal evaluation system was created from evaluation questions and other inputs from the stakeholder group, and addressed students' experiences throughout the different phases of these experiences.

### **Roundtable D (First Rotation): Evaluating Programs for School Guidance Counselors**

Presenters: *Evaluation Training for School Counselors*, Tamara M Walser & Paulette Running Wolf, Washington State University

School counselors in Washington state must be able to design, implement and evaluate Counseling and Career Guidance Programs (CCGP) as mandated by the Washington State Administrative Codes for school counselors. CCGPs include four components: 1) curriculum, 2) individual planning, 3) responsive services, and 4) system support. Graduate students in the Counseling Psychology program at Washington State University are required to develop a CCGP plan given the scenario that they have just been hired as a school counselor in a small, rural, k-12 school district. Counseling Psychology and Educational Psychology faculty collaborated to integrate CCGP evaluation into the development of the CCGP plan: The goal was to provide the students with evaluation skills that would be relevant to their future careers. This integration of evaluation training as well as resulting implications for the teaching of evaluation, particularly as it relates to the specific skills required of those in education, will be presented.

*Sustaining Evaluation Capacity for School District Guidance and Counseling Programs*, Michael S Trevisan, Washington State University and Michael Hubert, Yakima School District

The recent emergence of a comprehensive, developmental, guidance and counseling (CDCG) program has given program evaluation a priority in the day-to-day task demands of K-12 school counselors. Trevisan and Hubert (1999) reported successful steps in implementing school district evaluation infrastructure for CDCG programs. The purpose of this paper is to extend the work of Trevisan and Hubert and document strategies for sustaining evaluation capacity, thereby transforming an evaluation development project from implementation to institutionalization. These strategies include conducting regular feedback to key district stakeholders, obtaining limited evaluation assistance external to the guidance and counseling staff, developing multi-year evaluation tasks and timelines, and mentoring new counseling staff to the central role program evaluation plays in the maintenance of quality CDCG programs as well as their responsibilities for ensuring sound evaluation. These strategies will provide interested evaluators with the repertoire needed to assist school districts develop and maintain CDCG evaluation support.

### **Roundtable E (First Rotation): Teaching the Teacher: Evaluation of Programs for Teacher Technology Training**

Presenters: *Does Training Achieve Traction?: Evaluating Mechanisms for Sustainability from Teacher Technology Training*, Catherine A Nelson, Jennifer Post & William Bickel, University of Pittsburgh

How and why does professional development take root and grow in a school, ultimately changing teaching and learning? This evaluation examines the long-term impacts of a foundation-sponsored teacher technology training program in a large urban school district. A diverse sample of ten schools was studied a year after the training intervention, with a focus on identifying mechanisms for building school-level momentum from the district-wide program. The analysis considers factors such as formalization and institutionalization of the program at the site level, development of support structures to help teachers translate new skills into classroom practice, mandates and incentives, and changes in both faculty and student culture. Using our previously reported framework for evaluating the classroom impact of teacher technology training (Nelson and Bickel, 1999), we identify barriers to and strategies for achieving and sustaining training transfer which are potentially significant in a variety of evaluation contexts.

*Computers in the Classroom: Evaluating a Professional Development Program for Teachers*, Karen Williams, St Paul's Episcopal School; Glennelle Halpin & Gerald Halpin, Auburn University

Teachers in today's schools must be technologically literate for the integration of computers in their classrooms is no longer a luxury but a necessity. While pre-service teachers are being prepared to use the tools of the 21st century, many in-service teachers lacking training in technology are finding the acquisition of these skills more difficult. Professional development programs addressing the technical needs of in-service teachers must be designed and evaluated. In this study, 42 teachers in elementary through high school participated in a four-module self-paced computer training program, and their level of learning and attitudes toward computer technology were evaluated. Their program participation resulted not only in mastery of the technology but also in global attitude changes as well as increased comfort with computers. Described are the computer training program employed, the processes used to

evaluate it, and the implications of the results obtained.

### **Roundtable A (Second Rotation): Focusing on Students at Risk**

Presenters: *Strengthening the Implementation of Block Scheduling at a High Performing High School: Focusing on Students at Risk*, Astrid Hendricks Smith, W K Kellogg Foundation, and William E Bickel, University of Pittsburgh

A high performing high school in Western Pennsylvania has continued to support an effort to strengthen its implementation of Block Scheduling. As part of an ongoing evaluation process that began in 1995, new activities will focus on two constituency groups, teachers and students. Issues related to how students at academic risk can be best served, and various aspects of student-teacher interactions in the classroom remain will be of primary focus. The following presentation will discuss the new evaluation activities and early progress made toward three broad outcomes:

- \* identify specific aspects of faculty and student concerns related to the two issues of at risk students and student-teacher interactions in the classroom;
- \* understand the relationship between faculty and student concerns and the "flexible block scheduling" initiative (if any); and,
- \* identify strategies for improving school performance in these areas whether suggestions for improvement involve block scheduling matters or not.

*A Mixed-method Evaluation of a Structured Mentoring Reading Tutorial Program for "At-Risk" Students*, Chi-Keung Chan, Teri D Davis, Cheryl A Videen, & Gale Mason-Chagil, Saint Paul Public Schools

Since November 1998, a structured mentoring reading tutorial program, Helping One Student To Succeed (HOSTS), has been adopted by several schools in the St. Paul Public Schools (SPPS) to help elementary students who are "at-risk" for reading. This evaluation employs a mixed-method approach to examine the attitudes of the school administrators, teachers, mentors, and participants towards the program, as well as the benefits of mentoring on reading attitudes and reading achievement of the participants. This presentation describes issues related to the background and methodology of the evaluation, the findings of the study, and how the mixed-method evaluation provides useful information for the schools to improve the administration of the mentoring program so as to benefit the targeted students. Finally, the presentation concludes with a discussion of how this mixed-method evaluation increases the evaluation capacity by establishing an evaluation framework on the structured mentoring program for the SPPS.

### **Roundtable B (Second Rotation): Does Character Education Work?**

Presenters: *Evaluating Character Education Outcomes*, Bethany A Page, University of North Carolina at Chapel Hill

The North Carolina Character Education Partnership, comprised of the North Carolina Department of Public Instruction and 3 of the 4 largest school districts in the state, has worked to create, implement, and disseminate a model character education program. Since its inception in 1996, the initiative has been evaluated. As part of this evaluation, a character education (CE) checklist derived from past principal surveys was used to measure school level implementation of CE. The results from the checklist were then compared to school violence data, achievement data, and results from a student climate survey. Evaluation instruments and results will be shared.

*A Two-year Evaluation of an Elementary Character Program*, David A Payne, University of Georgia and Cheryl Love, DeKalb County Georgia School System

In 1997 the Georgia Board of Education required that a list of 36 character education concepts (e.g., honesty, responsibility, respect) be taught in all of the state's public schools. Each school was responsible for creating or adopting their own instructional interventions. Two experimental schools in the present report adopted a word-a-week approach. A new character education concept (CE) or quality was introduced and defined during homeroom of each new school week. National and local exemplars of each quality were presented and the concepts integrated into the ongoing school program (e.g., social studies, English). Reading and writing tasks associated with the CE concepts

were undertaken. A pre-post treatment contrast quasi-experimental design was created to gather data at grades two (n = 130) and six (n = 180). Custom-made instruments were created to assess (a) student comprehension of CE concepts, (b) endorsement of character qualities, (c) attitude toward school, (d) parent opinion, and (e) discipline referrals. The same data were gathered for both years of the project. Two major findings thus far are:

- (1) A statistically significant increase in participating students' knowledge of the understanding of character education concept/value words.
- (2) A lower incidence of suspensions (in-school and out-of-school) was noted for students in participating schools.

### **Roundtable C (Second Rotation): Evaluating Contracted Services in the Education Arena**

Presenters: *Elements of Specific Contractual Services Evaluation*, Stephen J Ruffini, Baltimore City Public School System

Private companies have been hired to take control of entire public schools while other companies have been hired to provide a specific service. Individual speech and language, occupational therapy, and physical therapy services have been provided to special needs students in an urban school system. Concerns for accountability prompted an initial descriptive evaluation. This addressed simple concerns about the number of goals attempted and the level of progress. The design of a subsequent evaluation has raised a number of considerations for any agency that works with a profit-making contractor. Important considerations for discussion include client selection and the nature of the goals. Another consideration is whether to set a standard to gauge success. Also to consider is the level of data that the contractor maintains. This raises questions about secondary analysis, data accuracy audits, and collaboration.

*When Evaluation and Marketing Merge: For-profit Educational Management Organizations and the Manner in Which They Market and Evaluate Their Education Programs*, Gary J Miron, Western Michigan University

Charter schools have provided an efficient path for private educational management organizations (EMOs) to enter the public school sphere. Close behind the EMOs are venture capitalists and private investors. Since the education market in the U.S. is estimated to be more than \$6 billion per year, and since it repeats itself every year, its no wonder that the private sector is so interested in this traditional public domain. Charter schools by design have to be accountable, and the EMOs that operate them conduct the student assessment and program evaluation activities. Parallel with this activity the EMOs are involved with marketing their product. Increasingly, public boards that contract with EMOs as well as the public at large have been questioning the legitimacy of the evaluation activities conducted by the for-profit companies. The material for this paper is based upon evaluations conducted by The Evaluation Center: statewide evaluations of charter schools in Michigan, Connecticut, Pennsylvania, and Illinois, and an in-depth secondary analysis of student performance data in schools operated by Edison Schools Inc. The paper includes examples of how marketing information is passed off as evaluation. The paper also contains recommendations for safeguards and contractual arrangements to limit the merging of evaluation and marketing information.

### **Roundtable D (Second Rotation): Implications of Culture in Educational Teaching and Testing**

Presenters: *When Quality Goes beyond the Standard: Is Standardized Assessment Resulting in Social Discrimination and Inhibiting Innovation?* Rosalina Barrera, Robert T Jimenez & Edith J Cisneros-Cohernour, University of Illinois at Urbana-Champaign

Efforts for increasing the validity of student assessment resulted on the new Standards for Educational and Psychological Testing (APA, 1999). Although the new standards have included new considerations regarding the assessment of diverse children, some controversial issues remain. Issues regarding the inappropriateness of testing for making decisions about student promotion need attention when assessment results are inadequately used as measures of student achievement\*. Because the validity of the assessment is usually ignored, more attention is needed to the intended and non intended consequences of the assessment for students, teachers and schools. Consequences such as the diminishment of instructional quality and unfair practices that result in denying students educational opportunities need special attention as well as how the emphasis on high stakes assessment may be limiting teachers opportunities to use innovative methods that could improve the educational opportunities for all children.

In this paper we address some of these issues by presenting three case studies of schools with a diverse student

body where highly committed teachers introducing new teaching alternatives are facing the challenges of using traditional assessment procedures that do not fit their innovative teaching methods.

---

\* Stake, R. (2000). )., Evaluation and Assessment: Discriminatory practices against minorities?

*Cultural Synchronization Theory: Implications for Program Design and Evaluation*, Sonja M Poole, University of California at Berkeley

Cultural synchronization (CST) is the theory that students, particularly students of color, would perform better academically if they can identify with the approach, structure and attitudes of the teacher and institution. Despite an emphasis in urban communities on connecting cultural foundations with pedagogy and institutional focus, relatively little is known about how these links are made and what they look like. In this paper, I explore the relationship between CST and actual program practices and the implications for program development and evaluation. Drawing from my experience studying and evaluating an Afrocentric middle school in Oakland, California, I show that the use and acceptance of CST leads to unique program practices that are distinct from other programs. In addition, evaluating these programs requires evaluators to go beyond "cultural sensitivity." The findings support calls to recruit and train more evaluators who represent the communities of the programs evaluated.

### **Roundtable E (Second Rotation): An Overview of Museum Evaluation**

Presenters: Stephen C Bitgood, Jacksonville State University  
Amy Cota-McKinley, University of Tennessee

This session will describe and give examples of one of the most prevalent models of evaluation used in museums today. The model incorporates user evaluation during all major stages of exhibition and program evaluation. During the planning stage, visitor input includes such things as feasibility of the topic in terms of interests, knowledge of the topic area; misconceptions of the exhibit topic, relevant attitudes, and preferences for communication media. This information is then incorporated into the final goals and objectives of the exhibition. During the preparation stage, visitor input is obtained by testing mock-ups of exhibits and/or program components. Changes are made based on this visitor input. Final construction is implemented only after the mock-up is performing at an acceptable level. This formative evaluation process is particularly important when interactive ("hands-on") devices are being developed since design complexity is increased for such exhibits. After the exhibition (or program) is installed (or implemented), visitor input can help to identify problems associated with sight lines, sensory competition among exhibit elements, unclear educational messages, etc. While formative evaluation tends to focus on intra-exhibit factors (how each part of a display such as text, instructions, illustrations, and objects functions in unison), post-installation (remedial) evaluation is generally more concerned with inter-exhibit factors (how all of the displays function together). A final type of evaluation (summative) assesses a completed exhibition (or program) without any intention to improve the program. Such an evaluation may serve to provide information about whether to replace an exhibition, consider a renovation, or to write a report for the funding source.

---

### **SESSION 513: Panel**

**Room: Oahu**

Sponsored by the Topical Interest Groups on State and Local Government & Evaluation Use

### **Partnerships in Government Evaluation**

*(The second part of this panel is to be held on Saturday at 11:10 in the Oahu ballroom.)*

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

No one single evaluator or evaluation organization has all the answers when dealing with complex evaluation projects! Partnerships with other evaluation entities can help build evaluation capacity. In particular, these partnerships are useful in government settings to manage limited financial and staffing resources of evaluation offices and to address varied and often conflicting needs of stakeholders. The panelists will bring several perspectives on the use of private and university sector expertise in public sector sponsored evaluations.

Panelists: *Using Private Sector Expertise in Public Sector Sponsored Evaluations*, George F Grob, US Department of Health

and Human Services

This presentation will present the advantages and disadvantages of using private sector evaluators for public policy and program evaluations. It will describe the circumstances under which contractors, grantees, and private foundations can make important contributions to promote effective public decision making. It will raise and suggest answers to important questions, like: Whose evaluation is it anyway? How long can the public wait for it? What if the sponsor and the evaluator disagree? What are the most effective ways to procure outside help? In general, this session will help public evaluators decide how to effectively use private sector resources, and will help private sector evaluators learn how to effectively make contributions to public policy making.

*Using University Sector Expertise in Public Sector Sponsored Evaluations*, John M Owen, University of Melbourne

For evaluation to be meaningful and to ensure that policy makers retain a commitment to using evaluation as an aid to decision making, we need to make every attempt to ensure good standards of practice. Evaluators in universities can and should make a contribution to this end. Due to their position, they are especially privy to the latest epistemological and methodological knowledge through access to journals and links with colleagues who are at the cutting edge of evaluation work. This paper considers ways in which university evaluators could influence the quality of practice. These include: providing advice to government agencies planning evaluations, assisting with the selection of consultants, providing frameworks for others to undertake evaluation, and working in partnership with neophyte evaluators. Recent examples of the use of these strategies by the author will be critically reviewed.

*Partnerships that Increase Public School System Capacity*, Maria D Whitsett, Austin Independent School District

As the public sector is increasingly expected to "do more with less," forming partnerships with private sector entities becomes functionally necessary. In this speaker's experience, private sector partnerships with public school systems have proven fruitful in enhancing evaluation capacity. It must be acknowledged that more than one end is served by securing private sector assistance with evaluation and related functions. For example, external evaluators may provide added credibility to an evaluation due to their expertise in a specialized field, or due to their apparent impartiality towards a given program. Enhanced credibility, in turn, may be associated with greater leverage to effect program improvements and make funding decisions. Perhaps the best scenarios are those where private sector partners, either through voluntarism or contractual agreements, have identified procedural efficiencies and helped create systemic changes that led to self-sustaining evaluative activity by public sector staff. Specific examples will be shared with the audience.

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

#### **SESSION 514: Panel**

**Room: Waialua**

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

#### **Internal Evaluation: Challenges to and Strategies for Maximizing Utility across a Range of Settings**

Chair: Elizabeth A Cooley, WestEd

This panel session consists of four presentations that discuss the issues and tradeoffs that appear, along with strategies for addressing them, when conducting internal evaluations within or across educational R&D settings. The first three speakers will explore these issues from the perspectives of conducting internal evaluation: 1) across multiple projects that share a similar focus within an R&D organization; 2) across an entire R&D agency that comprises many diverse programs, projects, and sites; and 3) across the national network of 10 Regional Education Laboratories. The fourth speaker will integrate the first three perspectives by highlighting the similarities and differences between the issues encountered, and by speaking to the common themes and lessons learned. Audience input and sharing of perspectives will be invited throughout the session.

Panelists: *Internal Evaluation of Cross-project Capacity-building Efforts*, Stephanie B Wilkerson, Mid-continent Research for Education and Learning

As part of McREL's contract with OERI, researchers formed long-term partnerships with regional educators in order to enhance individual and district capacities to facilitate systemic reform. McREL researchers engaged intensive site participants in a collaborative research process and as a result, provided participants with the opportunity to increase their capacities for implementing their districts' reform initiatives. As part of McREL's internal evaluation efforts, an evaluation study was developed to assess the impacts of these researchers' capacity-building efforts on participants in the intensive sites. The results will be used to inform McREL researchers as they plan future research work, and will provide impact data regarding the overall effectiveness of the collaborative research process. This study exemplifies how internal evaluation can function to meet the information needs of both external and internal stakeholders and at the same time, remain relevant, meaningful and collaborative.

*Internally Evaluating the Impact of a Regional Laboratory*, Elizabeth A Cooley, WestEd

Serving as the internal evaluator for WestEd has raised a number of substantive and procedural issues. First, WestEd is an R&D organization with a strong culture of academic autonomy. As a result, institutional evaluation can trigger concerns about programs' independence. Second, the programs themselves report to a variety of funders, each with its own specific evaluation requirements. Program managers are often reluctant to burden their staff or clients with additional evaluation demands, and in some cases may view such efforts as superfluous or too costly. Third, the organization is large and complex, comprising 13 programs across 11 sites; thus, identifying the right questions to ask about its functioning is challenging. This presentation will focus on both the issues we encountered and the efforts we made to address them, with an analysis of the degree to which we have been successful.

*System-wide Performance Indicators: Tail Wagging the Dog?* Roger J D Chesswas, WestEd

Externally required, internally conducted reporting on Laboratory system performance carries the potential to be either an opportunity or an opportunity cost. On the one hand, it provides an opportunity for staff within the 10 Regional Laboratories to think together about what constitutes important and useful evaluation. The opportunity cost, however, comes when what must be gathered in the way of performance data turns out to be less than meaningful or pertinent. Using the REL system-wide performance indicator system as an example, this presentation will explore how Labs might manage their reporting requirements on performance and leverage what they do to collect such data more efficiently, while at the same time addressing other critical, cross-cutting performance questions that are commonly shared.

*Common Challenges and Lessons Learned across Three Internal Evaluations: Putting the Pieces Together*, Naida Tushnet, WestEd

This presentation will shed light on the different dimensions along which the first three types of internal evaluation varied, e.g., the size/degree of complexity of the entity being evaluated, the extent to which the evaluation questions were internally derived or externally imposed, and the degree to which each one's respective audiences had compatible information needs. Common issues, challenges, and tradeoffs will be highlighted, and cross-cutting lessons and insights will be discussed as they pertain to serving as an internal evaluator across a range of settings.

---

## **SESSION 515: Panel**

**Room: Waianae**

Sponsored by the Topical Interest Group on Qualitative Methods

### **Using Focus Groups To Refine Outcome Indicators and Improve Services in Vulnerable Populations**

Chair: Aja Tulleners-Lesh, Azusa Pacific University

Outcomes research has traditionally emphasized quantitative measures while qualitative research has been used predominantly for process evaluation. In this panel presentation the use of focus groups has been expanded to explore the significant dimensions of outcomes in social and health programs. Presentors will describe the advantages of utilizing focus groups to refine and improve indicators of program effectiveness across a range of geographic areas and diversity of populations. The presentation will highlight the often unconsidered approach of giving voice to underserved populations in the refinement of outcome indicators and the improvement of services.

Panelists: *Listening to Patients' Voices: Health Services Improvement in a Substance Abuse Program*, Martha Ann Carey, AECOM

As part of an ongoing effort to improve services to patients in a community based methadone maintenance program in the South Bronx, this project explored patients' perceptions and problems in transitioning to a recently available medication for treatment of heroin addiction. Supporting patients in their decisions is an important component of care. Outcomes in substance abuse methadone programs focus on a continuum of recovery. Clinical experience in recent research suggest the importance of patient involvement in the identification of components of their recovery. Challenges of working with this population include distrust, vulnerability, medical comorbidities, poverty, housing, and unemployment. Focus groups were used to examine barriers to appropriate decision-making and to identify and refine patients' perspectives of their recovery.

*The Use of Focus Groups To Define Non-traditional Indicators of Outcomes in Drug Treatment*, William L Nigh, TRC

Outcomes in drug treatment have consistently emphasized decreased criminal activity and abstinence as the premier indicators of a successful program. While federal level GPRA indicators support these measurements and are mandated reporting requirements, they are not the equivalent of evaluating program effectiveness. However, when non-traditional indicators have been used as outcome measures (such as improved family contact and quality of life) the results have been disappointing. Frequently the initial surge of improvement in quantitative measures are not sustained as the treatment progresses. The use of focus groups has resulted in the reconceptualization of program outcomes and even the design of services to more accurately represent the unique components of recovery.

*The Use of Focus Groups in Managing Asthma as a Chronic Health Problem*, Aja Tulleners-Lesh, Azusa Pacific University

Outcome measures of programs that intervene in chronic health problems often look to stabilize the health condition and to reduce resource utilization, such as hospitalizations, rather than effecting a cure. Families are often the crucial component in maintaining this stability. The use of focus groups with families of asthmatic children illustrates both the multiple dimensions of care needed to reduce asthma episodes as well as the integration of the entire family unit in the management of asthma. Differences emerged between Spanish speaking and English speaking populations as well as males and females in their abilities to handle medications, treatment, fears, and the reduction of triggers for asthma episodes. Poverty and lack of knowledge as well as language differences also influenced their ability to manage treatment and reduce resource utilization. These findings served to refine and improve the subsequent indicators of program effectiveness.

Discussant: Kevin P Mulvey, Center for Substance Abuse Treatment

---

## **SESSION 516: MultiPaper**

**Room: Molokai**

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

### **Integrating Quantitative and Qualitative Methods in International Evaluation: It's the Way To Go but Often Tough To Implement**

Chair: Michael Bamberger, The World Bank

This session is based on the recent publication 'Integrating quantitative and qualitative methods in the evaluation of development projects', which reviews experience in the use of integrated quantitative and qualitative methods in the evaluation of World Bank and UNICEF projects. Case studies describe the use of integrated approaches in poverty analysis, education, health, water supply, and sanitation projects; and programs for mothers and children. While there are significant benefits to be obtained from integrated evaluations, the review found that the use of these approaches in international evaluation is frequently constrained by logistical, budgetary, professional and interdisciplinary issues. Guidelines will be proposed for increasing the effective use of integrated evaluation approaches in developing country contexts.

Presenters: *Integrating Quantitative and Qualitative Methods in International Program Evaluation: Lessons from World Bank Experience*, Michael Bamberger, The World Bank

There is a growing recognition of the benefits to be gained from combining quantitative and qualitative methods in development research. A number of areas are identified in which the World Bank, in common with other development agencies, is making increasing use of integrated quantitative and qualitative research approaches. There is no clear distinction between quantitative and qualitative methods and it is more helpful to consider data collection and analysis methods as being located on a quantitative-qualitative continuum. Many research designs use a combination of quantitative and qualitative methods at different stages of the research cycle. The major characteristics of quantitative and qualitative approaches are illustrated with examples drawn from the 7 case studies presented in the report. The presentation concludes with guidelines for developing an integrated research methodology that ensures that full integration of quantitative and qualitative methods is achieved in the analytical framework and at all stages of the research process.

*UNICEF's Use of Multiple Methodologies: An Operational Context*, Mahesh S Patel, United Nations Children's Fund

Evaluation methodologies are selected according to the problem itself, the management decision faced and the follow through required. Complex problems may require use of multiple methodologies: some quantitative, some qualitative. A range of informational inputs and hence of evaluative methods may be needed for operational decisions and policy formulation. Required information can range from statistically valid survey data to opinions of leading politicians as obtained by key informant interviews. The most cost-effective evaluation strategy may be to optimally mix qualitative and quantitative methods, rather than use only one in a pure form.

Discussant: Jim Rugh, CARE

#### **SESSION 517: Panel**

**Room: Maui**

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

#### **The Program Logic Model: Does It Make a Difference or Impact Program Performance?**

Chair: Stanley T Capela, HeartShare Human Services

Given the emphasis placed on outcome measurements and the increasing use of the program logic model, the panel will debate the value of the model. Specifically, does it make a difference or is it a process that is designed to do nothing more than meet funder and accreditation requirements? Based on their experience, the panel will offer insight on how they used the model, identifying useful techniques as well as barriers they experienced in implementing the model within the organization. In their discussions they will also focus attention on how to use the model to build evaluation capacity. Ultimately, the information obtained from this panel will help participants assess the value of using the program logic model as well as identify techniques that will help strengthen the value of the model as it is implemented within the organization.

Panelists: Melanie A Hwalek, SPEC Associates

The author has used the program logic model extensively over the years and will offer insight on the experiences of her clients that have used the model. Focus will be on the strengths and weaknesses of the model as well as some of the barriers that were incurred during the implementation stages at client sites.

Jim Emshoff, Georgia State University

The author has used the program logic model quite often over the years. The focus of his talk will be on his experiences working with clients on the use of the logic model in evaluation and how he helped clients create them.

Leslie J Cooksy, University of Delaware

Drawing on experience using logic models to frame evaluations of welfare-to-work, education, services integration and

international agricultural research programs, my presentation would first describe the role of the models in planning evaluations and reporting evaluation findings. For example, logic models have enabled me to focus evaluation on outcomes that were appropriate for the program's level of development. In just one example, I was able to persuade the evaluation clients that it was inappropriate to test the effectiveness of a low intensity educational program solely on the criterion of changes in children's test scores. As a result, more realistic outcomes, such as changes in classroom behavior, were included in the evaluation. The second part of the discussion would address the concerns that I have about program logic models, such as their condensation of complex programs in complex environments into simplistic flow charts and the need to regularly update logic models to keep them consistent with program changes.

Deborah Helitzer, University of New Mexico

The panelist will discuss the utilization of logic models in public health program evaluations. Emphasis will be placed on organizational resistance to logic models and helping stakeholders think through the process. Based on the panelist's experience, she will provide techniques that were successful in overcoming barriers and implementing the logic model in an organizational setting.

Discussant: Andrea A Anderson, Aspen Institute

## **SESSION 518: MultiPaper**

**Room: Kauai**

Sponsored by the Topical Interest Group on Theories of Evaluation

### **Perspectives on Dialogic Practice in Evaluation**

Chair: Melvin M Mark, Pennsylvania State University

The notion of dialogue has begun to capture the attention of evaluators in various ways. Dialogue is considered as a means of conducting various kinds of participatory and collaborative evaluations, as a tool for achieving fair and just deliberation of values, and as a framework for reconceptualizing the evaluation enterprise. This session will explore these various ways of conceiving of dialogue in evaluation practice. It will offer new, critical perspectives on the 'dialogue about dialogue.'

Presenters: *The Importance of Dialogue in Local Evaluation*, Thomas A Schwandt, University of Illinois at Urbana-Champaign

The modern state was born of the wish to create and maintain order. Its mission is to shape society and to mold citizens to reflect a rational social design. It aims to tame and domesticate disorder and ambiguity. Evaluators play a key role in the modern state as the providers of expert knowledge necessary to shape and control the natural and social environments in this fashion. The postmodern condition suggests this utopian dream is unattainable. In contrast to the quest for certainty and control, postmodernism evinces profound respect for social differences, local traditions, diversity, disorder, ambiguity, and uncertainty. This social condition demands a changing role for the evaluation expert. The concept of dialogue figures prominently in the reconfiguration of the evaluation expert. This paper will argue the case for reconceptualizing evaluation in conditions of postmodernity and will explicate a conception of dialogue as essential to the new responsibility of evaluators.

*Dialogue as Evaluation Practice*, Jennifer C Greene, University of Illinois at Urbana-Champaign

The concept of dialogue is a common feature of evaluative discourse. Dialogue is viewed as a democratizing practice that honors pluralism while advancing relationships and communications that are respectful, open, and equitable. Dialogue can also present evaluation stakeholders with a postmodern set of dynamic ambiguities and pluralistic perspectives that invoke reflection, conversation, and enhanced understanding of the Other(s). Significant as the concept of dialogue is, it is most often construed as an individual feature or facet or characteristic of evaluation. Most often, dialogue is envisioned as a framework or a forum within which to present our evaluation findings. What would evaluation look like if construed more holistically as a dialogue or a dialogic practice? What if we didn't just offer our

findings for dialogic engagement, but conducted our entire evaluation dialogically? What would dialogue mean as evaluation practice? And would such a conception of dialogue be meaningful and warranted?

*Deliberating without Rationality*, Sandra Mathison, State University of New York at Albany

In democratic societies, deliberation is a primary means for determining what is an appropriate course of action. Deliberation as a process is deeply embedded in a humanist perspective with its attendant valuing of rationality, i.e., through the right use of reason, truth and knowledge can be discerned. This paper will explore ways of decoupling the strategy of deliberation from its root presumptions of rationality. Further, modes of deliberation rooted in moral values, intuition, and lived experience will be explored.

*For Whom and to What Purpose Do We Need Dialogue in Evaluation?* Ove Karlsson, Marladalen hogskola

This paper will address the question of whose interests dialogue serves. One conception of dialogue, advocated in the recent work of House and Howe, views dialogue mainly as a tool for reaching decisions that are well reasoned, considered and reflected upon, and accomplished in a democratic process with all stakeholders. The central function of dialogue is procedural, to legitimize the decision making process. A second conception of dialogue is more substantive, personal, and transformative. It is concerned with the process whereby parties to a dialogue confront and engage their respective values, dispositions, and so forth and come to a mutual understanding. This paper will explore these two conceptions of dialogue and attempt to locate them in current evaluation practice.

Discussant: Melvin M Mark, Pennsylvania State University