Saturday, November 4, 1:50 P.M. to 3:20 P.M.

Sponsored by the Topical Interest Group on Collaborative, Participatory & Empowerment Evaluation
Developments in the Science and Practice of Capacity Building
Chair: Abraham H Wandersman, University of South Carolina

This panel will address the measurement, implementation, and process of building capacity within two community initiatives. In this panel, the authors will introduce a method for assessing program evaluation capacity, findings from this measure, a vision for future program capacity and a process for improving program capacity. The tools and methods described by this panel contribute to the science of empowerment evaluation and capacity building by offering a theoretically based practice for assessing and improving program capacity.

Panelists: *How Do We Measure Capacity Building Activities? Directions for the Future*, Patricia B Keith & Melanie Dye,
                University of South Carolina

Our current knowledge base, theoretical foundations, and field work have lead us to ask many questions such as "How are we aiding clients to better use their skills, abilities, and resources for self-assessment, accountability, and quality program development?". This paper will present an overview of instruments used to measure the process of skill transfer among evaluators, program directors, service providers, and boards of directors. It will share findings about the "state of art" of measuring what we do in the field with our clients and their programs, and where we hope to go in the future.

The Past and Current Evaluation Capacity of Two Community Initiatives, Jessica Snell-Johns & Dana C Keener,
                University of South Carolina

This paper will present findings from an assessment of the evaluation capacity of two comprehensive community initiatives. The same methods will be used to assess the evaluation capacity of both initiatives, although different empowerment evaluation approaches have been utilized. This distinction will allow the authors to comment on the implications of taking a comprehensive quality programming approach in comparison to an approach that works to establish a self-evaluation system from the onset of the programming. In doing so, the authors will evaluate their own effectiveness as capacity builders, as well as contribute to the greater understanding of how empowerment evaluation may influence the capacity of community initiatives.

Establishing a Vision for Practitioners' Capacity To Implement Highly Effective Programs, Jessica Snell-Johns,
                University of South Carolina

In this paper, Schorr's seven attributes of highly effective programs will be applied as a framework for examining the current capacity of a collaborative partnership and as a vision for the partnership's future capacity to implement highly effective programs. This presentation will describe the use of these seven attributes as a way to guide the growing capacity of the Foundation for the Future (FFF) Project in Spartanburg, South Carolina. It is hoped that this framework for examining programs' implementation will contribute to the field of empowerment evaluation by identifying a set of common principles that create a vision for practitioners' capacity to implement highly effective programs.

A Dialectical Approach to Building Capacity among Practitioners, Dana C Keener, University of South Carolina

In an effort to develop concrete strategies for building evaluation capacity, this paper will introduce concepts derived from an empirically validated clinical treatment that utilizes a dialectical approach to improve the quality of life among clients. As in therapy, evaluators seek to facilitate a growth process among practitioners that requires a sensitive balance between accepting the current capacity of a given program while challenging program personnel to make changes that are necessary for achieving greater program capacity. This paper will present several dialectical strategies, coupled with case examples, which elucidate how this approach has been used in an evaluation of a

comprehensive youth initiative. In doing so, the author strives to provide clear and practical tools for evaluators to utilize in their effort to build evaluation capacity among practitioners.

Discussant: Margret Dugan, Empowerment Evaluation Institute

SESSION 556: Panel   **Room:** Akaka

Sponsored by the Topical Interest Group on Assessment in Higher Education
Fostering the Elements of Change: Graduate Education and the National Science Foundation
Chair: Conrad G Katzenmeyer, National Science Foundation

The Graduate Research Traineeship (GRT) program was funded by the National Science Foundation from 1992 through1995. 157 awards were made to 93 different institutions which have supported 1,834 graduate students for portions of their graduate training. NSF has provided over $85 million in trainee funding through this program.

Between 1996-1999, two-person teams consisting of a program expert from Abt Associates and an academic content specialist conducted 60 site visits to GRT projects. In this panel, we discuss findings from these visits within the overall context of science graduate education. What were the changes in graduate education that spurred GRT's creation? How was the program structured to respond to those changes--what were its stated goals? How successful was the program in reaching those goals? What are the lessons learned from the multiple experiments of the GRT projects, and what do they suggest for the future of science graduate education?

Panelists: *Graduate Research Traineeship Monitoring: The NSF Perspective*, Conrad G Katzenmeyer, National Science
                    Foundation

As a Senior Program Officer in the Division of Research, Evaluation, and Communication (REC) of the Nation Science Foundation, Conrad Katzenmeyer has oversight responsibility for the evaluation of multiple cross-cutting programs funded by NSF. Dr. Katzenmeyer will discuss the design and uses of program evaluation within the Foundation with specific reference to the Graduate Research Traineeship program, as well as other graduate education programs. The discussion will encompass REC's responsibilities in meeting NSF's data requirements for the Government Performance and Results Act (GPRA); assessment at the program, division, and Foundation levels; and reporting to Congressional committees.

Graduate Research Traineeship Monitoring Study: Project Overview, Anne M Chase, Abt Associates Inc

Dr. Chase was Abt Associates' project director for this monitoring study. She will first describe the development of the Graduate Research Traineeship (GRT) program within the context of changing perspectives on graduate education in science, mathematics, engineering, and technology and within the context of support for graduate education by NSF. She will next outline the structure and goals of the GRT program. Finally, she will describe the overall design for this multi-year monitoring study, the major research questions addressed through interviews with faculty and students in the project as well as university administrators, the approaches used in analyzing these qualitative data used, and the reports produced through the study.

Funding Stream or Catalyst for Redesign? NSF's Graduate Research Traineeship Program, Beth Gamse, Abt
                    Associates Inc

This presentation will discuss changes in graduate programs we observed through our visits to 60 Graduate Research Traineeship (GRT) projects. The majority of projects visited used the GRT funding to introduce innovations in their educational programs, and were able to fashion qualitatively different educational experiences for their students. Some projects used the GRT primarily as a funding stream for talented students, and did not alter their educational programming. One lesson is that there are differences across disciplines in the extent to which projects used the GRT as a catalyst for program redesign. Although the programs varied from engineering to computer science to conservation biology, it is clear that engineering projects relied upon the GRT as a support mechanism rather than as seed money for programmatic innovation-reflecting the demands of the labor market. It is also clear that the stronger

projects integrated a variety of experiences into their educational programs.

Talent and Diversity: The Graduate Research Traineeship Program and the Human Resource Base, Karen S
　　Johnston, Abt Associates Inc

This presentation discusses the role of the Graduate Research Traineeship (GRT) in fostering two key NSF goals in human resources development: 1) increasing the number of scientists trained in critical and emerging fields in science, mathematics, engineering, and technology; and 2) strengthening the nationÆs human resources base among underrepresented groups. Because they could offer generous support, most projects were able to attract a talented cadre of students who might otherwise have gravitated to different fields. Success in recruiting women and underrepresented minorities was mixed. The most successful GRT projects typically exhibited a strong commitment to recruitment of women and minorities, had faculty from underrepresented groups, and had well-established relationships with institutions serving underrepresented groups. The dearth of interested candidates from underrepresented groups was cited as a significant challenge to recruitment, which many project faculty believed could be ameliorated by strengthening k-12 and undergraduate pipeline programs in math and science.

Discussant:　Deh-I Hsiung, National Science Foundation

SESSION 557: Panel　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Room:** Koko
Sponsored by the Topical Interest Groups on Minority Issues in Evaluation & International and Cross-cultural Evaluation
Evaluation in American Indian and Native Hawaiian Communities
Chair:　　　Joan L LaFrance, Mekinak Consulting

As tribal and native communities develop their own educational institutions and take control of their own social and health service delivery, they are interested in developing evaluation capacity. This panel will discuss issues facing Indian and Native Hawaiian evaluators who work with native and reservation communities. The panel members have extensive experience working in tribal communities across the United States and the Native Hawaiian community. Their discussion will focus on the cultural and technical challenges facing evaluators with a special emphasis on how evaluation has been adapted to fit the situation. The panel members will also discuss what Indian and Native Hawaiian cultural worldviews have to offer the field of evaluation. Some of the issues to be discussed include: small numbers; confidentiality in close knit communities; adapting outcomes to fit the resources and realities; empowering communities to conduct evaluation and research relevant to their goals; evaluation within a native context.

Panelists:　*Building Capacity: Empowerment Evaluation in Reservation Communities*, Joan L LaFrance, Mekinak Consulting

The Lummi Reservation in Northwest Washington State is revising their Constitution to reflect tribal community values. The Constitution Committee has struggled with this challenge. The Committee has felt caught in a Catch 22 û unwilling to propose changes without knowing what the community wants, and not know what the community wants without proposing changes for public vote. Dr. LaFrance assisted the Committee with focus group and survey methods to learn from the community. The project involved young tribal college interns and transferred research skills to community members. This presentation will discuss this project and the types of fundamental questions facing tribal communities as they attempt to make their governing structures consistent with their traditions.

From Los Angeles to South Dakota, Liana Champagne, Independent Consultant

Dr. Champagne worked as an evaluator in the Los Angeles School District for a number of years before returning to South Dakota where she works as an external evaluation consultant. She will discuss the contrasts in working large, diverse, urban population and in small homogeneous tribal communities. The presentation will focus on the challenges facing evaluators in tribal settings such as: limitations in documentation; constraints on comparison designs; and difficulties in working with small numbers. She will discuss ways to adapt evaluation to these special situations.

Multi-site Evaluation in Native American Communities, Richard Nichols, ORBIS Associates

Richard Nichols is a principal in Orbis, a Native American owned consulting firm in Washington, DC. He has extensive experience as an evaluator of single-site and multi-site educational programs on reservations in the East, Mid-West and Southwest. His presentation will focus evaluating multi-site programs. He will discuss issues in using multi-site logic models in Indian Country. Of particular interest are the ways in which outcomes are defined in Indian communities. The history of low educational attainment, social and health problems and poverty cause grass roots program staff see program outcomes somewhat differently than many external funding organizations.

Evalution as Catalyst, Brenda Wong, Alu Like Inc

Evaluation offers opportunities to shape and focus programs serving Native Americans and Hawaiians. The panel will feature this role for evaluation in the Native Hawaiian community. Brenda Wong works as an evaluator for ALU KIKE, Inc., a Native Hawaiian Vocational Education Program. She will discuss how evaluation can serve as a catalyst in the planning and implementation of programs. She offers the perspective of an internal evaluator working with a service delivery staff serving Native Hawaiian community.

Discussant: Joan L LaFrance, Mekinak Consulting

SESSION 558: Panel   **Room:** Waimea
Sponsored by the Topical Interest Group on Quantitative Methods: Theory and Design
Alternatives to Randomization and "Placebos"
Chair: Lee B Sechrest, University of Arizona

Randomization is often quite difficult in evaluation studies and sometimes even impossible. Moreover, even when randomization is possible, the formation of appropriate control groups can be greatly frustrating. That is particularly so when it seems necessary to use the equivalent of a placebo control group. The case for such pseudo-placebo groups is often weak, and options may be available that avoid such comparisons or that eliminate the need for separate comparison groups at all. Conservation of resources and efficiency of information acquisition urge that our evaluation designs be kept as simple and as inexpensive as possible in light of data requirements to support inferences that are desired. This panel addresses problems of research design for evaluation studies in terms of the rationale for comparison groups, including placebo treatments, comparisons of interventions to normative data, innovative comparison groups, and the interplay of theory and method.

Panelists: *The Rationale for Comparison Groups*, Opher Caspi & Lee B Sechrest, University of Arizona

Two questions arise concerning putative effects of any intervention: are the effects real? and what is the mechanism by which the effects are produced? Whether the effects are to be regarded as real depends on ruling out chance and artifactual changes, i.e., threats to validity. Comparison groups can help to rule out threats to validity, but not all threats are necessarily plausible. Threats to validity are best answered by comparison groups that receive no treatment or that receive what is standard treatment. No-treatment comparisons are needed and are ethical only for interventions addressing novel problems. So-called placebo groups have nothing to do with threats to internal validity but do help to answer questions about the mechanisms responsible for changes, i.e., construct validity. Placebo is a generic term; in any given study, the specific mechanism representing an alternative explanation needs to be specified and a comparison group devised to represent that explanation. When mechanisms are not in question or alternative mechanisms are implausible, some comparison groups are irrelevant.

Normative Expectations as Standards for Comparison, Victoria Weldon, University of Arizona

Under some circumstances, normative data may be available against which to judge program effects with sufficient confidence that any other comparison group is unnecessary. The specific conditions necessary for acceptance of norm group comparison methodology have never been made explicit, but they can be derived from theoretical

considerations and from analysis of extant literature. This paper sets forth the specifications likely to make normative comparison group methodology acceptable and illustrates them with actual studies.

Innovative Comparison Groups, Rebecca J Hill, University of Arizona

The use of nonequivalent comparison groups is quite common in evaluation, but too often the groups are chosen opportunistically or without much thought to their actual usefulness in understanding and arguing for effects of interventions. Devoting more thought and imagination to the identification and assessment of comparison groups could result in considerably strengthened arguments for the effects of interventions. The beginning point should be more careful specification of the rival hypotheses actually constituting the greatest threats to validity and then a survey of characteristics of comparison groups that would best deal with the threat. Examples of innovative comparison group designs in the existing literature will be presented.

Theory and Method, Lee B Sechrest, University of Arizona

The effective and persuasive use of comparison methodology requires the interplay of theory and method. As will be argued here, theory is method in the sense that strong theory enhances the value of data supporting it and, therefore, makes methodological weaknesses more tolerable. The identification of appropriate comparisons by which to judge interventions depends on sound theoretical understanding of treatment mechanisms so as to enable evaluators to formulate comparisons that will both make plausible the hypothesis of interest and weaken the most plausible alternatives. Examples of both failures and successes of theoretically buttressing arguments will be discussed.

SESSION 559: Skill-building Workshop   **Room:** Niihau
Sponsored by the Topical Interest Group on Extension Education Evaluation
Evaluating Youth Development Programs Using Qualitative Methodology
Chair: Susan J Barkman, Purdue University
Presenters: Susan J Barkman, Purdue University
              Krisanna L Machtmes, Purdue University

Beginning professionals often are unfamiliar with using qualitative methodology to measure and document the impact of their programs. Participants will have an opportunity to view actual examples of how observations, interviews, existing documents, and audio-visual materials have been used to evaluate the impact of existing programs. The presenters will demonstrate how to use triangulation to establish validity.

Participants will learn step-by-step procedures and guidelines for designing qualitative collection methods to measure the impact of their programs. As programs begin to incorporate more technology it is important for educators to be able to measure and document this arena. Thus participants will learn how to analyze e-mail and on-line focus groups. Presenter will utilize a combination of computer presentations, videotape recording, and small group activities and interactive discussions. Program participants will break into small groups throughout the program to learn how to measure the impact of their programs.

SESSION 560: MultiPaper   **Room:** Hilo
Sponsored by the Topical Interest Group on Non-profit and Foundations Evaluation
What Would Be Defined as Quality Evaluation Research? Academics Versus Practitioners Do They Use the Same Definition? If No, What Are the Differences?
Chair: Stanley T Capela, HeartShare Human Services

Beginning professionals often are unfamiliar with using qualitative methodology to measure and document the impact of their programs. Participants will have an opportunity to view actual examples of how observations, interviews, existing documents, and audio-visual materials have been used to evaluate the impact of existing programs. The

presenters will demonstrate how to use triangulation to establish validity.

Participants will learn step-by-step procedures and guidelines for designing qualitative collection methods to measure the impact of their programs. As programs begin to incorporate more technology it is important for educators to be able to measure and document this arena. Thus participants will learn how to analyze e-mail and on-line focus groups. Presenter will utilize a combination of computer presentations, videotape recording, and small group activities and interactive discussions. Program participants will break into small groups throughout the program to learn how to measure the impact of their programs.

Presenters: *Community, Nonprofits, and Program Evaluation*, Elizabethann O'Sullivan & Steven Lesky, North Carolina State University

With the devolution of government services public agencies rely on nonprofits to deliver human services. The arrangements may be far more complex than a simple agreement between a nonprofit and a public agency. Rather, a network of public and nonprofit agencies may deliver a service. For example, clients in transitional housing may live in housing provided by one nonprofit and managed by another. Client rent may be paid by a county agency. Clients may rely on different agencies and programs for transportation, medical services, mental health services, and recreational activities. Current information on how to evaluate programs may be ill-suited to determining the effectiveness of a decentralized effort. The proposed paper will examine whether this devolution has hindered evaluations indicating if the community as a whole has benefited from a program or if we are stuck knowing how the individual components work. In the proposed paper we will examine the evaluation literature to learn if and how it considers programs designed and implemented by a network of independent agencies. We will also look at works that consider the evaluation needs of nonprofits. Among the questions we plan to examine in the review are: does the literature lead individual nonprofits to evaluate only their processes and outcomes? how should efforts that involve cooperation and coordination between two or more nonprofits be evaluated? who is responsible for looking at effectiveness of the overall program? what strategies should be used to evaluate a decentralized effort. The paper will examine 2 to 3 decentralized programs funded by a county human services department. The department is committed to results-oriented management and partnering with other community organizations. The purpose of this section is learn what evaluative efforts have been completed or are underway and identifying barriers and opportunities to evaluating the overall programs. The information will be gathered from county and agency documents and interviews with department administrators and the executive directors or program managers of affected nonprofits.

Understanding and Improvement, Carolyn Sullins, University of Illinois at Urbana-Champaign

The author will talk about the struggle with the use of a transformatory empowerment evaluation of a NFP consumer-run mental health drop-in center. The original intent was to help staff and consumer of a mental health drop-in center gain ownership and direction of their drop-in center by learning to evaluate it as a team and act on the results Both staff and consumers had told me that although there were built-in processes of empowerment evaluation, they amounted to little more than a complaint department. They were hoping that by becoming engaged in evaluation skills, the consumers would learn to take more ownership of their center and help it continually grow and develop. Ultimately, the question the author will pose is what we're doing - generating and integrating knowledge and action - evaluation? It does seem to be empowering and empowerment (via self evaluation) was the main goal of this project. Do the ends justify the means? Can I even call myself an evaluator in this role? Or should I call myself an "understander and improver?"

Evaluation as Retrospective Policy Analysis: Views of a Method(olog)ical Eclectic, Doug Fraser, Independent Consultant

The authror will talk about how he has dealt with the question of what is evaluation based on his experience in Australia.

Meeting the Evaluation Challenge in Nonprofit Sector Organizations, Sandra L Bozzo, Canadian Department of Justice

Nonprofit organizations are faced with an increasing demand on the part of public and private sector agencies and the public to demonstrate the effectiveness of their programs and activities. Within the nonprofit sector, however, organizations must first cope with capacity issues before they can adequately respond to these evaluation expectations. There are a number of evaluation resources that have been developed to address critical questions regarding evaluation and to guide nonprofit organizations through the evaluation process. Do existing evaluation resources sufficiently facilitate the evaluation process for nonprofits? Do these resources help organizations to understand the various approaches and methods available or to understand which approaches or methods work best in which contexts? Based on an assessment of the usefulness and applicability of existing evaluation resources for nonprofit sector organizations, there appears to be room for substantial improvement and training in this area. The lack of specificity in the reviewed material - in relation to, for example, choice of methodology and factors that need to be considered in making evaluation decisions - compounded with the reality that nonprofit organizations may lack the skills or financial resources to undertake evaluation activities point to the need for private and public sector agencies and nonprofit organizations, together, to take a more proactive and collaborative role in terms of capacity building in this area.

Discussant: Stanley T Capela, HeartShare Human Services

SESSION 561: Panel   **Room:** Puna

Sponsored by the Topical Interest Group on Evaluation Use
Roles for Program Logic in Interactive Evaluation Practice
Chair: John M Owen, University of Melbourne

Interactive evaluation is concerned with assisting program managers and deliverers to document achievements and make decisions about future directions of their programs. While the use of program logic has been shown to aid large scale evaluations, an issue is the degree to which the development of program logic can be of assistance in evaluations which are consistent with an interactive or developmental perspective. This panel provides three case studies of successful evaluations in which the development and use of program logic was integral to the process.Each of the cases describes the context of the study, the roles of the evaluator and the integration of program logic into the evaluation process. Through comparing and analysing the findings of these cases, some tentative generalizations will be synthesised as the basis for discussion and debate with the audience. It is expected that audience members may also bring examples of the use of program logic in other interactive settings. Adequate time will be allocated for audience participation to ensure that the these examples can be discussed and compared with those of the presenters.

Panelists: *Program Logic and a Basis for Ongoing Decision Making and Documentation: The Shared Action Project*, John M
            Owen, University of Melbourne

The Shared Action Project was a three year intervention designed to improve the quality of life for residents in a local community in a large country town. Based on funding from a philanthropic trust, the Project was based on the involvement of the community in decision making. Acting in close conjunction with the steering committee the author, acting as evaluator,developed a program logic during an early stage of the Project. The logic was unique in its clear separation of program and evaluative dimensions. Over time the logic was used for two purposes. The first as a basis for key staff to 'check off' the achievement of successive stages of the process. The second was to act as a framework for giving an account of the program as a whole. The process relied on the development of trust between the author and key staff and the acceptance of the evaluator as a 'critical friend'. In this case the program logic provided an important framework on which to base interactive evaluation procedures.

*Building Brighter Futures: Using Logic Models To Increase Utilization and Build Commitment in Community-based
            Programs*, Arnold J Love, Independent Consultant

The Building Brighter Futures Project (BBF) was initiated in 1994 in seven high-risk communities, as part of a 25-year

community-based primary prevention initiative. Specific objectives of the Project included reducing the incidence of low birth-weight babies, increasing parenting skills, reducing risk factors associated with abuse and neglect, improving child functioning, increasing community responsiveness to the needs of parents, and improving the effectiveness of community agencies and organizations. The Project was operated as a partnership among more than a dozen community agencies and services were provided jointly by "community parents" and professional staff from the partner agencies. This presentation describes the use of the Logic Model process to (a) to focus the evaluation, (b) evaluate the contributions of dynamic Project component programs, (c) strengthen staff and community ownership of the evaluation, (d) and improve the utilization of evaluation information by Project staff, partner agencies, and the broader community by linking evaluation findings to action.

Logic Modeling as a Tool for Guiding the Evaluation of a University-based Technology Integration Project, John A McLaughlin, University of Southern California

Program evaluation is a process that creates an information base that will reduce uncertainties about the project in the minds of key stakeholders in the project's design, delivery, and results. A successful evaluation will tell a program's performance story in a way that enables stakeholders to connect the program to the important problems it is trying to serve. The evaluation must inform stakeholders of specific program accomplishments and results that lead to solving the problems to which the program is aimed. More importantly, the evaluation should produce information about what worked and did not work with respect to the project mission. This paper will present the methodology that progressed through three stages:

1. Clarifying the program's espoused theory of change - in the eyes of key program staff what were the strategic aims of the program and how was it supposed to work in order to achieve these aims;
2. Documenting the program's theory in action - observing actual program results and how the program worked to achieve these results under what conditions;
3. Building explanations - trying to understand why things worked and did not work and recommending improvements in future practice.

Throughout this process, regular interactions were maintained with key program staff and their partners. Interim products were shared in draft to afford stakeholders an opportunity to reflect and comment on presentations and interpretations.

SESSION 562: Panel   **Room:** Kohala
Sponsored by the Topical Interest Group on Evaluation of Services for Special Needs Populations
Tracking Impact Data: National Network Evaluation Methods
Chair: Kenneth R Olsen, Mid-South Regional Resource Center

An overview outlining the evaluation efforts of a collaborative technical assistance network funded to help states systematically improve educational services to children with disabilities and their families. Participants will discuss the need for evaluation including Government Performance and Results Act (GPRA) requirements, funding office needs, advocacy needs, and internal use. A brief description of the national network will lead into the presentation of separate evaluation efforts. Representatives from three different Regional Resource Centers (Western, Mid-South, and Northeast) will present their current methods for gathering, reporting and using impact data. In addition, representatives will discuss their efforts to collaborate for improvement in evaluation and the implications for multi-site evaluation, reporting, and use. Through collaboration, continuous feedback, and increasing evaluation across the nation, this presentation suggests a more global approach when evaluating the effects of national networks and each participating site. They will discuss the potential for such evaluations to increase the capacity of the clients to review and revise their own work.

Panelists: *WRRC Evaluation Techniques*, Anita Pierce, University of Oregon

Anita Pierce will discuss the response of Western Regional Resource Center (WRRC), which provides assistance to Alaska, American Samoa, California, Commonwealth of the Northern Mariana Islands, Federated States of Micronesia, Guam, Hawaii, Idaho, Nevada, Oregon, Republic of the Marshall Islands, Republic of Palau, and Washington. Her presentation will have an emphasis on evaluation, not only for RRC internal use, but also to help

states review and revise their own systems.

MSRRC Evaluation Techniques, Mackenzie D Riney, Mid-South Regional Resource Center

Mackenzie Riney will present the current tracking methods implemented at the Mid-South Regional Resource Center (MSRRC). The MSRRC provides assistance to Delaware, Maryland, Kentucky, North Carolina, South Carolina, Tennessee, Virginia, Washington, D.C., and West Virginia. The emphasis of the MSRRC will be on linking a range of evaluation components into a comprehensive whole (i.e., input, process, output, outcome, and impact).

NERRC Evaluation Techniques, Patricia H Mueller, Northeast Regional Resource Center

Patricia Mueller represents the Northeast Regional Resource Center (NERRC), which provides assistance to Vermont, Maine, New Hampshire, Massachusetts, New Jersey, New York, Rhode Island, and Connecticut. She plans on describing how NERRC conducts needs assessments with State Education Agency Directors. She will discuss different techniques of collecting data and how NERRC is meeting the needs of its state through evaluation.

Discussant: Mackenzie D Riney, Mid-South Regional Resource Center

---

SESSION 563: MultiPaper   **Room:** Kona
Sponsored by the Topical Interest Group on Alcohol, Drug Abuse, and Mental Health
Substance Abuse Treatment and Criminal Justice
Chair: Robert G Orwin, Battelle Centers for Public Health Research and Evaluation
Presenters: *Fixing the Fallacy of Criminal Justice Indicators in Outcome Research*, Martha C Beattie & Hung Nguyen, Alcohol and
             Drug Services Research Institute; and Richard Speiglman, Public Health Institute

Criminal justice activity is difficult to interpret, although treatment outcomes often include counts of arrests and convictions to evaluate the extent of legal involvement. However, an arrest or conviction during or after treatment on a warrant or for a charge prior to treatment should not "count" as during- or post-treatment illegal activity. Clients might even work on resolution of outstanding warrants as part of treatment.Santa Clara County, CA's Department of Alcohol and Drug Services has begun an outcomes monitoring program using secondary data from other agencies within the county to examine the extent to which patterns of use of county services are altered following treatment. This paper will explain the process used to interpret criminal justice activity records to arrive at realistic interpretation of how patterns have altered. It will present preliminary findings from the use of this process, and compare the new approach to the more traditional method.

Results from a Three-year Follow-up of Outcomes for Three Drug Courts in Washington State, Gary B Cox, Charles
     Morgan & Linda Brown, University of Washington

The pupose of this project is to evaluate outcomes for persons referred to three drug courts in Washington State. Four naturally occurring groups in each court (graduates, failures, opt-outs [who were offered entry but declined], and ineligibles [referred to court but subsequently found to be ineligible]) were compared on several measures for two years pre and three years post drug court referral. Data were collected from existing state administrative sources, including earned income, arrests, convictions, use of several social services, and others. Results to date show that, generally speaking, graduates showed higher incomes and lower arrest rates than the other groups after entry to drug courts, but not before. Mortality rates and chemical dependency treatment patterns were similar across courts. Analyses of the other data sets are ongoing.

Motivation and Treatment Readiness in Criminal Justice Clients, Maureen L O'Keefe & Melissa Doffing, Colorado
Department of Corrections; Kelli J Klebe, University of Colorado at Colorado Springs; and G Edward Wensuc & Erica
J Boyce, Colorado Division of Criminal Justice

Motivation and treatment readiness were assessed in participants across two substance abuse treatment modalities, a prison-based therapeutic community (TC) and a community-based intensive residential treatment (IRT) program.

Voluntary participants included 500 convicted felons. Offenders completed two self-report questionnaires at treatment admission: the University of Rhode Island Change Assessment (E. McConnaughy, J. Prochaska, & W. Velicer, 1983) and the Circumstances, Motivation, Readiness, and Suitability (CMRS; G. De Leon, G. Melnick, D. Kressel, & N. Jainchill, 1994). IRT participants were also evaluated by clinicians on a 6-point motivation scale.

This study compares offenders' scores to scale norms and examines mean differences between the two modalities. Within and between scale relationships are explored. Further, the present research examines the level of agreement between clinician ratings and participant self-report. The utility of two motivation questionnaires with criminal justice clients is discussed.

The Effectiveness of Substance Abuse Treatment in Reducing Violent Behavior, Robert G Orwin & Michael Maranda, Battelle Centers for Public Health Research and Evaluation

The use of alcohol and other drugs has long been associated with acts of violence. Yet relatively little is known about the effect of treatment on specific violent behaviors. This paper examines the distribution of violent behaviors in the National Treatment Improvement Evaluation Study (NTIES), and examines the extent to which treatment appears to reduce violence for each measured behavior. It also breaks out these reductions by gender, treatment modality, and whether the client was a victim of violence in addition to being a perpetrator. Finally, it examines what client characteristics, services, and program characteristics distinguish clients who reduce violent behaviors from those who do not. Preliminary analyses suggest large reductions in violence among those clients actively violent during the pre-treatment period, with virtually no increase in violence among clients not actively violent during that same period. Implications for theory or practice: Advancing understanding of the role of drug treatment in reducing violence.

SESSION 564: Panel   **Room:** Ewa
Sponsored by the Topical Interest Group on Health Evaluation
The Programmatic and Evaluation Aspects of Outreach Efforts: A Case Study in Baltimore, Maryland
Chair: Anthony C Tommasello, University of Maryland

It has been said that: "Outreach, broadly defined, is a practice even older than the wandering ministry of Jesus" (Rowe, 1999). Others have cited the horseback doctors of the American frontier as early outreach workers. Outreach has evolved considerably and is seen now as an organized structured effort, staffed by clinic personnel, with specific locations at which the effort is conducted. Outreach is currently funded by federal grants in an attempt to connect with hard to reach populations of individuals in need of medical care and social services. Legitimate questions about outreach are being asked and the need for evaluation research to examine these questions is apparent. This panel presents a case study of outreach to homeless substance abusers in Baltimore, Maryland as it has evolved over the past five years. The clinical and research challenges of incorporating evaluation methodology into service delivery systems are presented for consideration and discussion.

Panelists: *Outreach to Homeless Substance Abusers: Challenges and Impact on Program Operations*, Laura M Gilis, Health Care for the Homeless

Since 1986, Health Care for the Homeless (HCH) in Baltimore, Maryland has been conducting outreach. In 1995, funding from the Center for Substance Abuse Treatment (CSAT) provided the resources to support a targeted outreach effort of HCH clinical teams engaging homeless people abusing substances and recruiting them into treatment based on the indigenous leader model of the National Institute of Drug Abuse (Booth and Weibel, 1992). In 1996, HCH expanded this model through funding by HRSA. Outreach recipients were found to have a higher incidence of HIV, IV drug abuse, non-IV drug abuse, and alcohol abuse than were those clients that walked in for services without the benefit of outreach. When outreach efforts to a targeted population succeed, system changes may be necessary to adjust to the growing proportion of outreach recipients. This presentation discusses outcome indicators for measuring the impact of outreach endeavors on the service delivery system

Evaluating the Effectiveness of Outreach at Health Care for the Homeless: Selecting and Measuring the Outcomes,

Anthony C Tommasello, University of Maryland

Health Care for the Homeless in Baltimore, Md. conducts outreach to homeless substance abusers at high risk for HIV acquisition. Outreach teams contact these individuals at soup kitchens, homeless shelters, and on the streets. These outreach recipients are compared to clients who have walked into the clinic for services without the benefit of outreach. Our data show outreach contacts are more likely than walk-in clients to be involved in substance abusing HIV risk behaviors. These findings are consistent with others doing similar work, supporting the success of outreach to connect with the target population. Outreach clients differ from walk-in clients only marginally in other areas, indicating that outreach recipients are a subset of the larger population of potential service recipients. Client level measures regarding service consumption are reported. Evaluation at the systems level to describe the impact of outreach on the dynamic of service delivery is needed.

The Evolution of an Outreach Initiative and Collaboration with an Evaluation Team, Indera Kotval, Health Education Resource Organization

HERO is a well-known community based AIDS agency in Baltimore, in existence since 1983. We have a constant inflow of HIV positive individuals seeking help. When the epidemic reached the substance abusing population we realized that outreach was necessary to approach them before they became infected. Starting with street outreach HERO progressed to mobile van outreach, an office on wheels. We take services to the community because individuals who are not within the system become infected and infect others. HERO is a service driven agency and did not want research to jeopardize program operations. Through the relationship with the evaluators in a SPNS/HOPWA project we developed an understanding of the benefits of evaluation. We see evaluation as a process for achieving a better understanding of our program through which we can create a more efficient service delivery system. In a new grant application HERO included the evaluators in the planning phase and integrated evaluation concepts into the project implementation.

Evaluation Practicality: Serving Client Needs While Stimulating Methodology Improvements, C Patrick Myers, University of Maryland

Client organizations engage in program evaluation studies for various reasons. Federal funding under "Knowledge Development and Application" grants require an evaluation component, participation in "cross-site" evaluation efforts is a common expectation of federal service grants, and service organizations desire data demonstrating to prospective supporters that their program is effective. Service providers may be wary of evaluation fearing that it hinders clinical procedures or imposes unnecessary and undesired burdens on clients. Evaluators should proceed with these sensitivities in mind. The evaluation team must work to develop a positive working relationship based upon mutual respect and trust so that the evaluation staff and program staff each are motivated to achieve the highest quality evaluation. As an example of how developing such a relationship improves evaluation methods two evaluation models are presented. The first was developed after funding was awarded. The second resulted from collaboration that began when the grant proposal was under development.

Discussant: Molly Engle, Oregon State University

SESSION 565: Panel   **Room:** Honolulu
Sponsored by the Topical Interest Group for Graduate Students
Advancing the Discipline: Evaluation Journal Editors Discuss Ways of Building Our Knowledge Base
Chair: E Jane Davidson, Claremont Graduate University

Because evaluation is a young discipline, evaluators are often trained primarily within fields such as the social sciences, rather than in evaluation per se. Accordingly, they may receive little guidance about contributing specifically to evaluation journals (as opposed to journals within their own fields). This panel brings together editors from four evaluation journals, and asks: (1) What does your journal look for that differs from what other journals (both within and outside evaluation) require? (2) What types of articles do you and your editors wish for but never see submitted? (3) Where are the gaps and opportunities for the new generation of evaluators to inject some innovation into the

evaluation literature? It is hoped that this session will not only encourage some of the field's new minds to start making a contribution, but will also highlight areas where contributions would be particularly valuable for expanding the discipline's knowledge base.

Panelists: *The American Journal of Evaluation*, Melvin Mark, Pennsylvania State University

Dr. Melvin Mark is the editor of the American Journal of Evaluation (AJE). He will be representing that publication, explaining the processes used for review and selection of submissions to AJE, and how they differ from the other journals. Dr. Mark will also be sharing his views on what kinds of contributions he feels would be particularly valuable to evaluation's growing body of knowledge, and the main opportunities he sees for innovation in the evaluation literature.

New Directions for Evaluation, Jennifer C Greene, University of Illinois at Urbana-Champaign

Dr. Jennifer Greene is co-editor of New Directions for Evaluation (NDE). She will be representing that publication, explaining the processes used for review and selection of submissions to NDE, and how they differ from the other journals. Dr. Greene will also be sharing her views on what kinds of contributions she feels would be particularly valuable to evaluation's growing body of knowledge, and the main opportunities she sees for innovation in the evaluation literature.

Evaluation and Program Planning, Jonathan A Morell, ERIM - Center for Electronic Commerce

Dr. Jonathan Morell is the editor of Evaluation and Program Planning. He will be representing that publication, explaining the processes used for review and selection of submissions, and how they differ from the other journals. Dr. Morell will also be sharing his views on what kinds of contributions he feels would be particularly valuable to evaluation's growing body of knowledge, and the main opportunities he sees for innovation in the evaluation literature.

Studies in Educational Evaluation, David Nevo, Tel Aviv University

Dr. David Nevo is the editor of Studies in Educational Evaluation. He will be representing that publication, explaining the processes used for review and selection of submissions, and how they differ from the other journals. Dr. Nevo will also be sharing his views on what kinds of contributions he feels would be particularly valuable to evaluation's growing body of knowledge, and the main opportunities he sees for innovation in the evaluation literature.

SESSION 566: Roundtables   **Room:** Kahuku
Evaluation in Higher Education and Adult Education Settings
(This session includes two 45-minute rotations of roundtables. The Host will ask the tables to rotate at 2:35.)
Host: Mark Hawkes, Dakota State University

Roundtable A (First Rotation): Quality Assessment in Higher Education Standards
Presenters: *The Quest for Excellence in Higher Education: Seeking the Best of Standards*, Mandana Hashemzadeh, Claremont
                  Graduate University

The purpose of this paper is twofold. On the grand scale, it aims to provide audiences in the domain of higher education with the best of standards upon which colleges and universities can be evaluated. More acutely, it aims to evaluate and compare the new set of standards that the Western Association of Schools and Colleges has recently introduced to assess institutions of higher learning in the western region of the U.S. to the previously utilized set of nine standards. The basis or the general provision is a relevant 'needs assessment' of the 'impactees' of post-secondary institutions. The methodology involves a triangulation among a thorough literature review on the subject of quality assurance and accreditation of universities and colleges worldwide; tailored interviews with domain authorities, expert valuators, and other informed observers of higher education; and a greater survey of other field patrons for the purposes of casting a wider net.

Development of a Nation-wide System of Quality Assessment in the Swedish System of Higher Education, Folke K
    Larsson, National Agency of Higher Education

Institutions of higher education in Sweden normally constitute independent public authorities which are accountable to the government for their activities, their quality and their quality management. During the 1990s a number of quality audits and program evaluations/quality assessments were performed by the National Agency for Higher Education. A number of these assessments also concerned the right to award degrees. Following a decision by the Swedish parliament, a comprehensive nation-wide quality assessment system will shortly be introduced by the Agency. The paper describes the evolution of such a national system where all subjects and programs will undergo quality assessments at least once every six years. The process of identifying appropriate indicators and aspects of quality in basic higher education and postgraduate training is dealth with as is the process of implementing the evaluation system.

Roundtable B (First Rotation): Innovative Approaches to Documenting Learning
Presenters: *Evaluating Learning That Lasts: Evaluations Perspectives To Serve Post-secondary Education Practice*, William H
    Rickards, Alverno College

Emerging directions in higher education pracctice involve an integrated construction of teaching, reflection, and inquiry organized around service to learners. As frequently envisioned (e.g., through the recent Carnegie Scholars program), the inquiry approaches share many characteristics with evaluation. For example, they will address recurring performances and patterns of practice in existing curriculum contexts. They will often be problem-based, with deeper purposes involving the improvement of service at the same time that they may further conceptual understanding. They will draw on existing assessment practices as well as other practice-based data collection to build an understanding of teaching and learning as well as judging the value of pracitces. This paper, then, examines a collaborative approach to studying teaching and learning that integrates course and program-based assessments, close analysis of faculty practices, and student perspectives on learning in a campus' own efforts to understand and improve practice. Examples are drawn from programs dealing with students at two levels-entry to college and the transition from general education to major coursework.

Working Backward: A Program in Search of a Theory, Margie E Law, University of South Alabama and Susan A
    Tucker, E & D Associates

How is an evaluation undertaken when there exists no program theory on which to base the evaluation? This session suggests a model for program theory development for a study abroad program in the context of higher education. The use of formative evaluation methodology was employed to reveal the program theory. Data collection methods included interviews, questionnaires, and extant data analysis. Findings indicate (1) common threads in observations among participants and administrators but not between the groups; (2) that program theory articulation can result from the evaluation process; and (3) program theory development facilitates program improvement. In conclusion, findings from the initial session of the program suggest that program theory articulation developed through the process of formative evaluation can induce program improvement.

Roundtable C (First Rotation): Formative Evaluation in Higher Education
Presenters: *Lessons from a Participatory Evaluation of a Freshman Year Program*, David D Wilson, Brigham Young University

In January 1999, a participant-oriented evaluation began of a major university's freshman experience program serving over 1,600 students. 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 program. Many participants collaborated to help gather and analyze data to answer a key administrator's question, "How is the program doing?" from their multiple perspectives. The study also drew upon data collected by several participants for their own purposes since 1993. 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. 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 university communities will be considered.

Evaluation of the California Future Leaders Program: An Early Academic Outreach Program at a Private Liberal Arts College, Marycarmen Kunicki & Kelly Fischbein, Claremont Graduate University

This report is a formative evaluation using the Key Evaluation Checklist (Scriven, 1999) focusing on the California Future Leaders (CFL) program within an academic outreach program at a private liberal arts college. CFL brings 11th graders from local high schools to the college campus for a day of workshops and other activities. This visitation provides students with the information they need to undertake the college application process, while motivating them to pursue post-secondary education. The process used in this research can be generalized to evaluations of other similar programs. This evaluation plan provides the program staff with the tools necessary to continue the evaluation process over the years. In addition, it exposes the staff to the benefits of evaluation.

Roundtable A (Second Rotation): Reform Initiatives: Capturing the Process
Presenters: *Reform in Evaluation of Higher Education: What Is the Story?* Judith A Burry-Stock, John A Dantzler & Joy F Johnson, University of Alabama

Managerial type program methods were employed to evaluate a reform effort in freshmen engineering at a southeastern major university. The program is part of a national effort the goals of which are to encourage engineering students to be: active learners, team members, part of a learning community, and to use higher level thinking skills in problem-solving. The evaluation model incorporated aspects of an evolving reform model called Regular Evaluation, Feedback for Organization, Resources, and Management, (REFORM). Qualitative and quantitative data were collected to support the national goals and objectives. A number of barriers were encountered during the evaluation. These barriers included the gamut of problems ranging from: the preordained national interpretation of evaluation; national and local politics; and the lack of knowledge of assessment and evaluation processes on the part of the program administrators, faculty and staff. All of these effected the program evaluation implementation. In order to accomplish program evaluation in higher education reform new methods need to be designed to capture the reform process. If these processes are not understood then, meaningful reform and the evaluation process suffer. The results of this program evaluation tell a superficial, but useful story.

Improving Systemic Reform in Mathematics Education through Collaborative Evaluation: A Case Example, Michelle L Batchelder, Austin Independent School District

The Austin Collaborative for Mathematics Education (ACME), funded jointly by the National Science Foundation (NSF) and the Austin Independent School District (AISD) is an initiative to improve mathematics education districtwide. The major goal of the initiative is to support teachers with long-term professional development as they implement inquiry-based curriculum resources and improve teaching strategies. The initiative also targets administrative support for systemic reform and school cultures.

Motivational Forces: Motivational forces for this evaluation include (a) accountability to NSF, (b) interest in what makes systemic reform work, and (c) the desire of local administrators to improve the effectiveness of program activities. Organizational Environment: The internal, formative evaluation evolves as the program director and evaluator negotiated a collaborative relationship that involved exchanging ideas backed by data through teamwork and advice, not judgment. Workforce and Professional Development: The evaluator adapts quantitative and qualitative research skills to the needs of program and district administrators. Resources and Supports: Resources include surveys, observation protocols, and interviews designed by NSF and other instruments locally designed for case study research. Learning from Experience: Part of the collaborative relationship involves directing the evaluation to make the job of program staff easier and altering the program informed by evaluation findings.

Roundtable B (Second Rotation): Evaluating Teaching

Presenters: *A Multivariate Approach to the Evaluation of Teaching*, Charles W Davidson & Margaret Faye Davidson, Claremont Graduate University

The evaluation of teaching has been a perennial problem for education decision makers and for teachers themselves. The arguments have usually gravitated to there being too many variables, with many of those being uncontrollable or unmeasurable, to obtain valid evaluations. With the arrival of the computer, the number of variables can no longer be considered a part of the problem. However, in the 30 years that computers have been common place, the evaluation of teachers has still not been seriously addressed. It is the purpose of this paper to show how the authors intend to use multivariate statistical analyses procedures with school, community and individual student data (presently being collected on a state-wide basis) to arrive at a valid assessment of the performance of individual teachers. Since education is the nation's largest industry, this study, following peer review, should raise the demand and capacity for evaluation to a new level.

Evaluation of Teaching: A Comparison of Student Ratings and Focus Groups, Christopher Migotsky, University of Illinois at Urbana-Champaign and Trav Johnson, Claremont Graduate University

Typical end-of-semester student ratings of instruction are employed in over 90% of institutions of higher education in the United States. While many faculty find the results informative and insightful, others often complain about their narrow scope that severely under represents instructors' teaching quality. When asked to recommend alternative assessment techniques, faculty frequently say something like "talk to my students-they'll tell you how good the class was." The implication is that the same student cohort will give different interpretations of course quality depending on the methodology used to collect the student reactions. The current session compares the results of student rating and focus group information for the same university classes. Similarities and differences in student views are highlighted and discussed. Additionally, survey data from involved instructors shows their reactions to the two sets of evaluative data. Results should be of interest to faculty, administrators, and evaluators in higher education.

Roundtable C (Second Rotation): Evaluating Teacher Professional Development Initiatives

Presenters: *The Evaluator as Facilitator of Teacher Capacity Building in a Longitudinal Evaluation of Students' Higher Order Thinking and Learning To Learn Skills*, Imelda R Castaneda & Terri Heidotting, University of Cincinnati

This is a research/evaluation of an arts-integrated program implemented in the midwestern part of the United States. The evaluators work hand-in-hand with the clients and participating teachers in tracking the students' higher order thinking and learning to learn skills. Issues of what constitute a defensible longitudinal study have been raised. The evaluators, the clients, and the teachers, forming the research team, collaborate in raising the issues concerned in finding out about transfer of learning and the students' performance and skills development. The very essence of the cooperative relationship between evaluators, clients, and teachers that involves reflection, analysis, decision-making, and action research is empowering. The evaluators, the client, and the teachers are all learning together. Evaluation is becoming client and teacher-led. The evaluation process evolves so that clients and teachers eventually perpetuate their own evaluative skills in this area.

Evaluators and Practitioners Working Together: The Evaluation of Teacher Professional Development, Michael Puma, Jacqueline Raphael & Jorge Ruiz-de-Velasco, The Urban Institute

Schools and school districts are under more pressure than ever before to increase student achievement. National experts agree that the standards-based reform effort is dependent on school staff's ability to teach students to master high academic standards. With teachers indicating that they do not feel prepared to teach these challenging standards, professional development programs have become a critical component in the standards-based reform movement. Yet scarce research or evaluation has proven a link between professional development programs and student outcomes.

With funding from the U.S. Department of Education, the Urban Institute has formed partnerships with five school

districts from across the country to help strengthen their capacity (internal and external) to evaluate their professional development programs, including effects on students. In this roundtable, we will discuss the successes and challenges encountered by our district partners and the lessons learned by all participants about increasing evaluation capacity.

SESSION 567: Panel   **Room:** Oahu

Sponsored by the Topical Interest Groups on State and Local Government & Evaluation Use
Performance Evaluation: An International Symposium on Global Capacity and Challenges
Chair: Mark E Ward, US General Accounting Office

This panel of experts in the field of evaluation will examine the global reach of new evaluation scenarios and capacities in the context of management reform initiatives. Employing a diverse "open door" examination of evaluation initiatives across the globe, panelists will discuss how each of several countries has adapted new evaluation stategies to assess performance in a broad range of programs, such as social services, employment, and economic development. Presentations will focus on evaluative experiences in Canada, Great Britian, Pacific Rim nations, among others. The assessment of results has become a demonstrably international prioriy, providing a world stage for informed discussion about enhancing the development, application, and value of emerging evaluative techniques and their use in facilitating managerial change. More than a passive exchange of invaluable information, this panel will provide a forum to sustain engaged discussion about the pragmatic application of new evaluative techniques in an increasingly international and interdependent environment.

Panelists: *An Overview of Performance Measurement: An International Rainbow of Diverse Approaches*, Nancy Kingsbury, US
                 General Accounting Office

Based in part on the global work of the U.S. General Accounting Office (GAO), this presenation will provide a rich landscape of the diverse approaches to performance evaluation, related challenges, and emerging capacities. How does evaluation focus on results in the context of managerial reform? What challenges do evaluators encounter in aligning management information, human capital, and other support systems with intended program results? How do new evaluation tecniques foster the efficient and effective allocation of resources? Based on GAO's global experiences, the information highlighted will identify key elements necessary to establish, maintain, and enhance the most revealing and useful forms of evaluation.

How Organizational Location Affects Evaluation's Capacity To Meet User Needs, Stan Divorski, US General
                 Accounting Office and John Mayne, Office of the Auditor General Canada

Evaluation is becoming increasingly important as countries around the world focus increasingly on results as the basis for managing government programs. It is becoming clearer that performance measurement and monitoring systems cannot provide all the information that governments will need regarding program impacts and the relationship between programming strategies and program outcomes. At the same time, this greater focus on results creates a tension between the need for information for accountability purposes and information for program improvement purposes. Drawing on experience in a number of countries, this paper presents a model showing how the utility of evaluations is influenced by the location of responsibility for evaluation. The model depicts the fit that can be expected between the needs of potential users and the kinds of information that evaluation is likely to produce depending upon where the responsibility is situated. The model's pertinence to recent experience in implementing GPRA is discussed.

Modernizing Government: Addressing Key Challenges in Strengthening Evaluative Capacity, Michael Whitehouse &
                 Nick Lacy, National Audit Office

Great Britian's National Audit Office (NAO) has continued to address evaluative challenges--Value for Money Development--in the context of modernizing government. Managerial change presents constant challenges for adapting new evaluative strategies that provide timely and reliable information to improve program efficiency and effectiveness. The presenters will describe key developments that prompted the need for organizational change, how the Value for Money Development iniative evolved, challenges in instituting this approach, and contributing factors in

enhancing NAO's evaluative capacity. The evaluative lessons and applications of NAO's experiences in Great Britian provide timeless insights that can be considered in other national, organizational, and institutional settings.

Discussant: Mark E Ward, US General Accounting Office

SESSION 568: MultiPaper   **Room:** Waialua
Sponsored by the Topical Interest Group on Pre-K - 12 Educational Evaluation
The Evaluation of Professional Development in Education
Chair: Roger J D Chesswas, WestEd
Presenters: *Building Capacity to Keep the Collective "Eye on the Ball" in Schools*, Kristine L Chadwick Nilsen & Rebecca C Burns,
               AEL Inc

This paper describes an evaluated curriculum alignment process that has been shown to promote self-evaluation and data-based decision making in schools. The Teaching/Learning Mapping Strategy (TLMS) is a professional development and curriculum alignment process that provides schools with both the vehicle and the path to increase and maintain high levels of student achievement. Although there are other processes in the technical assistance marketplace for accomplishing these goals, TLMS is a comprehensive strategy that uses teacher self- and group-reflection and evaluation, collaboration, an electronic curriculum mapping tool, and a collective focus on student learning to achieve a system-wide curriculum that is aligned with national, state, and local standards. TLMS has been implemented in four school districts. First year TLMS evaluation results suggest that the process works well to build school systems that are inclusive, collaborative, self-reflective, and which base their instructional decisions on the results from multiple sources of data.

Considerations In the Evaluation of Teacher Professional Development, Roger J D Chesswas & Jeanne Elliott,
               WestEd

The professional development of teachers is seen as a critical component of improving teaching and learning in schools. Standards have been developed to assess the delivery of professional development but not the outcomes. This paper will present a logic or theory of how we can evaluate a series of professional development outcomes for teachers. We will talk about assessing what teachers learn in professional development, what they value in their learning, how teachers change practice, and how changes in teaching impact student learning and achievement. The critical issues in this paper will be to try to identify how one can trace the direct impact of teacher professional development, and how best to collect data on impact.

*Determining the Impact of Educator Professional Development Programs Using a Hierarchical Assessment Level
     Model*, Steve C Cordogan, Illinois Mathematics and Science Academy

Determining the impact of professional development experiences for educators requires selecting appropriate assessments, realizing that those incorporating more authentic measures of impact often require substantially greater implementation efforts. In the absence of appropriate models guiding such decisions for the education field, this model was developed to identify a hierarchy of seven levels of authenticity of assessment, covering participant counts, satisfaction indicators, and measures of changes in knowledge skills, and attitudes. These components are integrated with the following assessment considerations: 1) self-assessed or externally measured, 2) focused on the short-term professional development experience or on longer-term applications within the teacher's school, and 3) targeted toward the educators, their institutions, and/or their students. The resulting matrix also identifies an array of specific measures for each level. The accompanying text further details the application of the matrix so that this model can be used by practitioners to guide their assessment choices.

Linking Teacher Professional Development to Student Outcomes: Results of a Mixed-method Program Evaluation, Jo
     Ann Doino-Ingersoll, Strategic Research and Carol Boyle, Lakeland Central School District

This paper will present the qualitative and quantitative methods used to evaluate the impact of a Goals 2000-funded k-12 teacher professional development project on teachers and their students. Called WorkNet 2000, the project was

developed by a central school district in New York State in collaboration with a university, and has just completed its fourth year of state funding. The results of three evaluation methods will be presented, including:
1)   teacher focus groups and administrator interviews;
2)   a teacher follow-up survey;
3)   New York State Regents test scores of students whose teachers had completed the project's key professional development courses vs. test scores of students whose teachers had not completed these courses.
The degree to which the findings link positive project impact on teachers to increased learning outcomes for students will be discussed, as will implications for evaluation methodology in professional development.

Assessing Impact: A National Study of the Evaluation Capacity of State Educational Agencies, Wendy Togneri, Sara
       Dewees & Lois Haid, AEL Inc

Despite an emphasis on accountability, organizations receiving federal dollars have long been able to renew funding with minimal proof of progress. In an effort to make federally funded programs more accountable, Congress enacted the Government Performance and Results Act (GPRA) in 1993. GRPA requires agencies and programs to set results-oriented goals and evaluate the effectiveness of their activities.

One program under GPRA oversight -- the Teacher Training Program (TTP)*-- was created to assist states and schools with providing professional development to K-12 teachers. This paper draws from a national evaluation of the TTP, which assessed the ability of states to self-evaluate by setting measurable goals, gathering data, and analyzing programmatic impact. The study reveals that states and school districts are experiencing considerable difficulty in setting and assessing progress toward measurable goals. The paper will share 1) the results of the study and 2) the processes used by evaluators to assess state competence in setting performance indicators and measuring progress. Several evaluation rubrics will be shared, and their use will be described.

The paper will focus on the following interrelated topics.
- Self-evaluation: success in setting performance indicators
- Increasing the capacity of state and school staff to gather and analyze data
- Policy implications related to federal and state accountability

*Teacher Trainng Program is a psuedonym.

Discussant: Debbie Zorn, University of Cincinnati

SESSION 569: MultiPaper   Room: Waianae
Sponsored by the Topical Interest Group on Alcohol, Drug Abuse, and Mental Health
Perspectives on Building Evaluation Capacity in Local Health and Human Service Organizations
Chair: John F Stevenson, University of Rhode Island

Despite an emphasis on accountability, organizations receiving federal dollars have long been able to renew funding with minimal proof of progress. In an effort to make federally funded programs more accountable, Congress enacted the Government Performance and Results Act (GPRA) in 1993. GRPA requires agencies and programs to set results-oriented goals and evaluate the effectiveness of their activities. One program under GPRA oversight -- the Teacher Training Program (TTP)*-- was created to assist states and schools with providing professional development to K-12 teachers. This paper draws from a national evaluation of the TTP, which assessed the ability of states to self-evaluate by setting measurable goals, gathering data, and analyzing programmatic impact. The study reveals that states and school districts are experiencing considerable difficulty in setting and assessing progress toward measurable goals. The paper will share 1) the results of the study and 2) the processes used by evaluators to assess state competence in setting performance indicators and measuring progress. Several evaluation rubrics will be shared, and their use will be described.

The paper will focus on the following interrelated topics.
- Self-evaluation: success in setting performance indicators

- Increasing the capacity of state and school staff to gather and analyze data
- Policy implications related to federal and state accountability
*Teacher Trainng Program is a psuedonym.

Presenters: *Profiling Practice as a Catalyst for Capacity Building*, Paul Florin & Dana Scott Mills, University of Rhode Island

Describes how "profiling" program evaluation plans and practices from 14 diverse prevention agencies (funded by the Safe and Drug Free Schools and Communities state allocation) was used to develop working relationships and focus tasks in a project designed to enhance evaluation capacity. The procedures used to gather information and create profiles for each agency along eleven dimensions (five of evaluation design and seven of evaluation implementation) are briefly reviewed, while a summary profile across all 14 agencies is displayed as a simple but useful "audit" of evaluation practice among this group of organizations. Knowledge transfer, technical assistance and capacity building directions are described which were derived from both a "diagnostic" approach of assessing how agencies performed on each profile dimension, as well as a "consumer" approach of expressed needs as defined by the agencies themselves. Concludes with reflections on options for the locus of "evaluation capacity" within a prevention system.

Influencing and Measuring Evaluation Capacity in Human Service Organizations, John F Stevenson, Dana Scott Mills & Marco Andrade, University of Rhode Island

This paper reports on a project intended to enhance evaluation capacity within local RI agencies (14 programs). Extending from the needs assessment described in FlorinÆs paper, this work entailed: (1) the provision of a series of three evaluation workshops for agency personnel; (2) ongoing technical assistance to the agencies as they designed and conducted their program evaluations; (3) enhanced evaluations of three exemplars; and (4) continuing efforts to assess the changes in agency capacity for evaluation. The evaluators developed a list of "Steps on the Path to Successful Evaluation." Methods to measure progress in evaluation capacity enhancement have included (1) evaluator ratings of evaluation capacity on the "Steps" checklist; (2) agency staff ratings on the "Steps;" and (3) pre- and post-workshop ratings by attendees of their confidence in skills related to the "Steps." Results and conclusions are reported.

Building Capacity in United Way Agencies: A Customized Workshop Approach, Jim Emshoff, Georgia State University; and Steve Erickson & Julie Chambliss, EMSTAR Research

Staff from 68 programs, representing 59 organizations that are funded by United Way, were invited to participate in evaluation workshops. The workshops were designed to build evaluation capacity and in linking specific measures to their logic models. Programs were grouped according to their intended outcomes and their current evaluation capacity. Fifteen separate groups were organized around 13 indicators and skill levels. Participants submitted a response to a needs assessment regarding their evaluation capacity. Based on this assessment, a separate curriculum was created for each of these fifteen groups. The curriculum was divided into two half-day workshops separated by approximately one month, with homework assignments to be completed in the interim. Each participant was given a packet of resources customized to their intended program outcomes. Quantitative and qualitative data will be presented on the capacity needs reported, the experience of providing and receiving technical assistance, and lessons learned in the process.

Building Evaluation Capacity through a Research Methods Advisory Service, David R Thomas, University of Auckland

The presentation will report experiences of providing a research methods advisory service for health researchers. A primary focus is developing evaluation skills and competencies among researchers and health programme staff. The core services provided are regional workshops, development of locally-grounded evaluation procedures, a resource database, and consultation with research teams to assist development of evaluation plans and proposals. The service emphasizes the important of qualitative data and works with both mainstream and ethnic minority groups. Some of the key issues which have emerged are: ensuring evaluation procedures are culturally appropriate, facilitating self-monitoring in community programmes, developing realistic local indicators of outputs and impacts, the role of evaluations in assisting funding renewal, and using evaluations as a social change intervention. Policy issues and

cost-benefit perspectives on state-funded evaluation capacity building will be discussed, and implications for the effective delivery of health and welfare services will be outlined.

SESSION 570: MultiPaper   **Room:** Molokai

Sponsored by the Topical Interest Group on International and Cross-cultural Evaluation
Cross-cultural Transferability of the Program Evaluation Standards
Chair: Mahesh S Patel, United Nations Children's Fund
Presenters: *The Appropriateness of Joint Committee Standards in Eastern Asian Settings: Case Studies of Four Eastern Asian Countries (South Korea, Japan, Taiwan, and Thailand)*, Soojung Jang, Syracuse University

The Joint Committee Standards (JCS) has been used in educational evaluations in U.S, but they are not recognized as international standards yet. This study examines the appropriateness of JCS in consideration of cultural factors of South Korea, Japan, Taiwan, and Thailand. Each category of JCS is investigated in how external evaluators should change them when conducting the international evaluations with domestic staffs or clients of these countries. In international contexts, change is more often an unplanned reaction to external events. The researcher conducts survey research targeting government officer, educators, business men from the four Eastern Asian countries who have ever been involved in evaluation projects in their home countries. Especially she focuses on Propriety Standards closely related to the cultural factors.

Usage of the PES in Africa, Craig Russon, Western Michigan University

The African Evaluation Association Conference produced a strong consensus that a set of African Evaluation Guidelines, adapted from the PES, should be produced. Experiences with use of the PES in their current form have been mixed. Purchasers of evaluation services are often unaware of the PES and have limited time for an education in their utility. The PES were used in support of a UNICEF Country Programme Evaluation in Zambia and an Immunization Programme Review in Kenya. Progress tends to be piecemeal. The current focus is to get the PES included in model evaluation protocols that are repeatedly used.

Cross-cultural Transferability of the Program Evaluation Standards, Sandy M Taut, University of Cologne

The paper addresses the cross-cultural transferability of the Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1994). Based on cross-cultural psychological and anthropological literature, cultural values are identified and discussed for each of the 30 Standards. Examples are given as to the cross-cultural differences regarding these values. The analysis includes a review of the relevant literature dealing with the Standards' transferability. The author concludes that particularly Utility and Propriety Standards have limited applicability in cultures differing from their North American origin. Hence, evaluators from other cultural contexts should be aware of the underlying values limiting the Standards' cross-cultural applicability.

SESSION 571: Demonstration   **Room:** Maui

Sponsored by the Topical Interest Group on Qualitative Methods
Finding and Owning Your Voice as an Evaluator: Lessons for Evaluation from Creative Non-fiction and Autoethnography
Facilitator: Michael Quinn Patton, The Union Institute

The traditional third-person passive style of academic and research writing disguises the voice of the person writing. In contrast, the "personal factor" at the center of Utilization-Focused Evaluation supports the importance of evaluators finding and owning their voices. The importance of "voice" has become a central feature of creative nonfiction. Drawing on my experiences in writing a creative nonfiction autoethnography, Grand Canyon Celebration: A Father-Son Journey of Discovery (Prometheus Books, 1999), I will demonstrate how articulating the voice of the evaluator can enhance evaluation reporting. The challenges of voice-centered evaluation reporting will be explored, including pros and cons.

Evaluation Checklists as a Vital Tool for Strengthening Evaluation Practice
Chair: Arlen R Gullickson, Western Michigan University

Checklists are useful tools for evaluators because they are versatile, mnenomic, reduce the chances of overlooking important factors, reduce biases such as halo effect, and increase the defensibility of evaluation findings. In the past, a number of evaluators have developed and applied evaluation checklists for contracting, guiding, and assessing evaluation work. Recently, the Western Michigan University Evaluation Center has identified a growing need for and interest in using such checklists for program, product, and personnel evaluation and consequently has launched an NSF-supported project. Its aims are to compile existing checklists, develop new ones, and make them available to evaluators through a user friendly web section. It will also provide an in-depth examination of the logic and methodology of checklists, including guidelines for appropriate development and use. The project's team will report on its aims and achievements and invite AEA members to become involved as contributors, reviewers, and users.

Panelists: *The Logic and Methodology of Evaluation Checklists*, Michael Scriven, Claremont Graduate University

Some people see checklists as simple laundry lists of tasks, and many persons regularly use such hastily constructed devices. However, as illustrated by airline pilots' employment of checklists to assure that aircrafts are safe to fly, a checklist can be crucial to guaranteeing an operation's success and thus should be carefully crafted. Accordingly, checklists for judging products, programs, and personnel should be grounded in sound logic and methodology. In addressing this need, Michael Scriven will define the vocabulary of checklists, criteria for constructing and judging checklists, conditions for appropriate use, procedures for reaching conclusions, and steps for creating sound checklists. He will show that evaluation checklists should be relevant, complete, parsimonious, clear, and consistently pitched to an appropriate level. Important uses of evaluation checklists will be identified as providing analytical push, designing and contracting for evaluations, monitoring implementation, justifying conclusions, and metaevaluation.

The 10 Commandments, Constitutional Amendments, and Other Evaluation Checklists, Daniel L Stufflebeam, Western Michigan University

Sound checklists can have profound applications, like those associated with codes and laws. The subject project will provide evaluation checklists for program evaluations, personnel evaluations, and product evaluations. Within each category, specific checklists will focus on professional standards for evaluation, steps in the evaluation process, key dimensions of a comprehensive summative evaluation, and key dimensions of particular areas such as teaching, administration, and continuing education. Daniel Stufflebeam will present a conceptual framework for viewing different types of checklists and will present exemplars, including his metaevaluation checklist keyed to the Joint Committee Program Evaluation Standards; his checklist keyed to the steps of focusing, collecting, organizing, analyzing, synthesizing, and reporting evaluation findings; his Contracts Evaluation Checklist; and Scriven's Key Evaluation Checklist and Product Evaluation Checklist. It will be shown that evaluators often should apply a combination of checklists in examining a given evaluation.

The Evaluation Checklist Web Site, Lori A Wingate, Western Michigan University

In order to make sound evaluation checklists more widely available and facilitate discourse on the development and use of checklists, the Western Michigan University Evaluation Center created the Evaluation Checklist Project Web Site. This site is designed to be a clearinghouse for evaluation checklists. It will also provide information about the logic and methodology of checklists and guidelines for use. Drawing on the project team's substantial experience in developing and applying checklists, the Evaluation Checklist project Web Site is an important resource for the evaluation community. Lori Wingate will describe the contents, organization, and interactive features of the site. She will invite AEA members to become involved in reviewing and gathering materials for the site and in using the provided checklists.

Discussant: Arlen R Gullickson, Western Michigan University