Panel 9 Overview:
Program Evaluation

Over the last ten years, the role and sophistication of program evaluations have grown in parallel with the Summer Study. Conservation program evaluation has become a large field with many good practitioners. Those represented in this panel represent a variety of affiliations: universities, national laboratories, state agencies, public and private utilities, and private consultants. We feel the papers represented here are valuable contributions to the Summer Study. They were selected out of 70 abstracts. Nevertheless, many excellent evaluators could not be represented due to the time constraints and competing interests of the Summer Study. An entire conference on energy program evaluation is held every two years in Chicago (organized by Argonne National Laboratory) which should be of interest to those who find the papers in the panel to be useful.

The panel is divided into sessions each of which emphasizes one topic. In this introduction we will try to highlight the contribution of each paper to the session topic.

Exemplary Residential Program Evaluations is an ambitious title. Yet, most evaluators have seen so many residential weatherization evaluations that they would only be interested if there were something special about a new one. We believe that these are special. The Horowitz paper describes two straightfoward evaluations using mainstream approaches, but with significant differences between them. Questions arise about whether the methodologies selected may have influenced the results in a predictable way.

Kushler's paper is a classic example of the highly desired "evaluation-policy-evaluation" link. Over several years, evaluations tested policies; new policies were tried; and the new policies were evaluated.

While Michigan and Bonneville have examined the persistence and reliability of savings over three years, the Sumi and Coates paper presents an effort to follow weatherized cohorts for up to eight years. This paper should be read by everyone who feels uneasy about the reliability of conservation savings.

Appliance Efficiency Program Evaluation looks at evaluations of a furnace replacement program, a specific refrigerator rebate program, and a national overview of 10 appliance rebate programs and their evaluations. The Hall paper reflects how evaluations can provide a basis for serious policy decisions even if sample sizes are small and statistical significance is difficult to show. The Mystakides paper is intriguing in that they were able to collect so much information on a control group that they were able to indentify the efficiency of the refrigerators purchased outside the program area. McRae did a review of the evaluations done on appliance rebate programs to determine if a general statement could be made about the effect of "free-riders" on program economics. It will be interesting to see if all listners will share her reluctance to draw conclusions.

Issues in Evaluation Policy will provide a mix of methodology development (Newcomb) that can apply to other end use sectors, PUC policies to ensure quality evaluations (Prahl), and a humorous approach to a topic with a heavy title (Rothstein). This is a session that may attract planners and policymakers as well as evaluators.

Evaluation methodologies develop more quickly then they can be disseminated and replicated. One of the roles of the ACEEE Summer Study is to help speed the transfer of techniques and results among the participants and their agencies. Sophisticated Evaluation Methods is a session title that may confuse the authors if they consider their techniques to be noncontroversial. However, the use of two stage logits to estimate self-selection effects (Train), the estimation of free-rider effects in an efficient home certification program (Buchanan and Violette), and efforts to extrapolate evaluation results to a national scale (Carroll et mucho al) are not yet mainstream techniques. We anticipate that each of these presentations will raise issues that could involve the audience.

Take-back, rebound, amenity improvement, and the Khazoom effect are all terms for the apparent loss of conservation savings and they imply that behavioral changes occur after conservation actions. They are popular expressions in energy conservation parlance. Take-back has been characterized as an energy planner's ready explanation and an engineer's crutch when measured savings fail to match the savings predicted by auditors or analysts. The session entitled: Whither the Great Takeback Effect could be called "Wither the Great Takeback Effect" or "Khazoom Goes Boom". One paper (Ruderman) undermines the theoretical and mathematical basis for the effect, while Weihl and Ternes and Stovall attempt to measure the effect in terms of changes in temperature setting behavior with empirical measurement techniques. None of the authors (nor Hall from the Appliance session) report a substantial effect on other than an anecdotal basis. This session may make some people uncomfortable because the results seem to imply that long held beliefs may have been misplaced.

Data Analysis in Support of Evaluation. Evaluations require good data, often those which may not be easily obtained, and solid baseline data. Miller and Griffin describe the efforts of the California Energy Commission to establish a new database for planning while testing the reliability of previously reported survey data. Yoder and Schoch took an existing database and mined it further to help understand the variation in the success of the Hood River Conservation Project. Jan Moen, from Norway, explains how they combine measured data with engineering assumptions to evaluate their programs. Mark Jackson's paper may represent an evaluator's ideal situation--two years of detailed metering on five parameters with an experimental and a control group. But, even with this richness of data, the validity of the evaluation can be undermined by problems with sampling and sample attrition.

We feel that the papers to be presented in the Program Evaluation Panel are strong and represent the best of what is being done in this country. Nevertheless, do not expect perfect designs, fool-proof methodologies, unimpeachable results, or ideal samples. The best of evaluations will meet the difficult goal of being scientifically defensible, these papers should meet that test.

Kenneth M. Keating, Bruce Cody
Bonneville Power Administration