

# THE INFLUENCE OF PROGRAM THEORY-BASED EVALUATION ON THE USE OF EVALUATION INFORMATION: A LONGITUDINAL STUDY

Sheri Hudson, Ralph G Straton and Richard Cummings

*Murdoch University*

## INTRODUCTION

The ability of theory-based evaluations to expand understanding of how a program works, in addition to providing information regarding whether a program works, is claimed to lead to an increase in the usefulness of program evaluation information (Chen, 1990, 2004; Pawson and Tilley, 1996). It is surprising; therefore, that little empirical research has been undertaken to investigate the influence of theory-based evaluation on the utilisation of the information from an evaluation study. In fact, Rogers et al. (2000) suggest a lack of “real-world test and applications” (p.3) of theory-based evaluations indicates evaluation practitioners are working from professional intuition, rather than empirical evidence. The aim of the research presented here was to address this shortcoming in empirical research by investigating the extent to which program theory use influences the utilisation of the information provided by an evaluation study.

## PROGRAM THEORY-BASED EVALUATION

One of the primary differences between the *modus operandi* of an evaluation underpinned by the use of a program theory and that of other evaluation approaches is the basis upon which decisions about the evaluation study are made. There are two types of program theory (*implementation and causative*) which are very different in nature. Program *implementation theory* addresses the variables affecting how a program is implemented, whereas a program’s *causative theory* identifies the causal links between the predictor variables which govern a program’s application, and its intended effects (Scheirer, 1987). Although they are interrelated, it is essential to identify and distinguish between these two types of program theory because it clarifies the focus of an evaluation study and analysing both in an evaluation provides a fuller picture. In addition, it is difficult to differentiate

---

*The Influence of Program Theory-Based Evaluation on the Use of Evaluation Information: A Longitudinal Study*

*Sheri Hudson, Ralph G Straton and Richard Cummings*  
*Paper Presented at the Australasian Evaluation Society Conference, Canberra, 2009*

program causative theory failure from program implementation theory failure unless it can be shown that a program was faithfully implemented in accordance with the intended program implementation theory (Bickman, 1987).

This paper outlines the model underlying the present investigation and the design of the study, presents the findings and frames the discussion of the findings around three research questions. The main focus of the investigation is on the response of program stakeholders to an evaluation study and to the information provided by the evaluation. The three research questions are the following:

- i. What is the influence of program theory on the use of evaluation information?
- ii. Which factors have the greatest impact on the use of evaluation information? and
- iii. How do these factors interact with each other to affect use?

The answers to these questions should clarify the implications of adopting a theory-based approach in a particular evaluation study and how a theory-based study should be designed so as to optimise the utility and use of the information provided by the evaluation study.

#### THE MODEL OF EVALUATION INFORMATION USE

The model tested for this investigation includes five of the primary factors previously identified as facilitating the utilisation of evaluation information (Alkin, 1985; Cousins and Leithwood, 1986, 1993; Cummings, 1997; Leviton and Hughes, 1981; Hudson-Mabbs, 1993; Pawson and Tilley, 1997; Preskill and Torres, 1999). The five variables included in the model are *Program Theory*, *Evaluation Study Characteristics*, *Process Use*, *Commitment to Study (Pre)*, and *Commitment to Study (Post)*. While other factors have been identified as having an influence on an evaluation study and its outcomes (see Cousins and Leithwood, 1986, 1993; Hudson-Mabbs, 1993), the structural equation modelling process used here required that the Model be parsimonious, enabling the most efficient use of the information gathered. In addition, four separate outcome variables were observed: *Influence of Study Findings*, *Influence of Use of Program Theory in the*

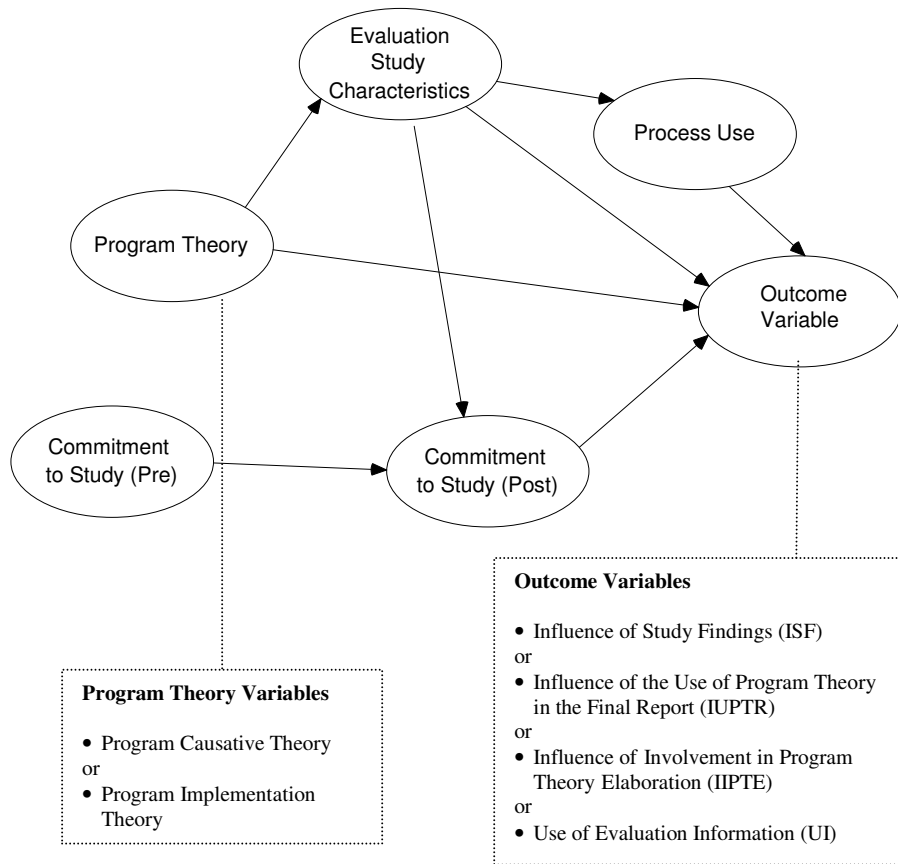
---

*The Influence of Program Theory-Based Evaluation on the Use of Evaluation Information: A Longitudinal Study*

*Sheri Hudson, Ralph G Straton and Richard Cummings*  
*Paper Presented at the Australasian Evaluation Society Conference, Canberra, 2009*

*Final Report, Influence of Involvement in Program Theory Elaboration, and Use of Evaluation Information* (see Figure 1 and Table 1).

**Figure 1: The Structural Equation Model**



## RESEARCH METHODS

The investigation focussed on three evaluation studies using a longitudinal case study design over the period of each evaluation. The cases were selected using six primary criteria:

- i. Extent of program theory use in the evaluation;
- ii. Duration of the evaluation study;
- iii. Stakeholder group identification;
- iv. Evaluation of an ongoing program;
- v. Program stage of development (ongoing program); and
- vi. Location of the evaluation consultants to the program.

**Table 1: The Eight Versions of the Model**

| <b>Versions Number</b> | <b>Outcome Variable</b>                                | <b>Latent Program Theory Predictor Variable</b> |
|------------------------|--|---|
| 1a                     | Influence of Study Findings                            | Causative Program Theory                        |
| 1b                     | Influence of Study Findings                            | Implementation Program Theory                   |
| 2a                     | Influence of Use of Program Theory in the Final Report | Causative Program Theory                        |
| 2b                     | Influence of Use of Program Theory in the Final Report | Implementation Program Theory                   |
| 3a                     | Influence of Involvement in Program Theory Elaboration | Causative Program Theory                        |
| 3b                     | Influence of Involvement in Program Theory Elaboration | Implementation Program Theory                   |
| 4a                     | Use of Evaluation Information                          | Causative Program Theory                        |
| 4b                     | Use of Evaluation Information                          | Implementation Program Theory                   |

The variations in theory use among the three evaluation studies were incorporated into the research design to allow for an analysis of the effects of theory use on the use of evaluation information. One evaluation study was selected for its strong use of program implementation theory with little attention being paid to causative theory, another for its strong use of program causative theory with little regard for implementation theory, and the third study for its apparent disregard of program theory. The three studies are briefly described below.

*Youth on Health Festival (YOHFest)*. This program was closely tied to its causative theory. It is a festival during which students from high schools throughout the State of Western Australia perform plays they have written and choreographed, perform a dance, or enter an art piece that they have created. The festival has been developed around the logic that peer education is an effective method of addressing youth health issues. Primary stakeholders of YOHFest include: the Lions Club (Mandurah, Western Australia); the South Metropolitan Health Unit; the Department of Education and Training; and Healthway. The YOHFest evaluation study was selected for its strong use of program causative theory.

*Pre-Apprenticeship Training Programs in Western Australia*. The program provided entry-level courses leading into trade-specific employment or further training. These

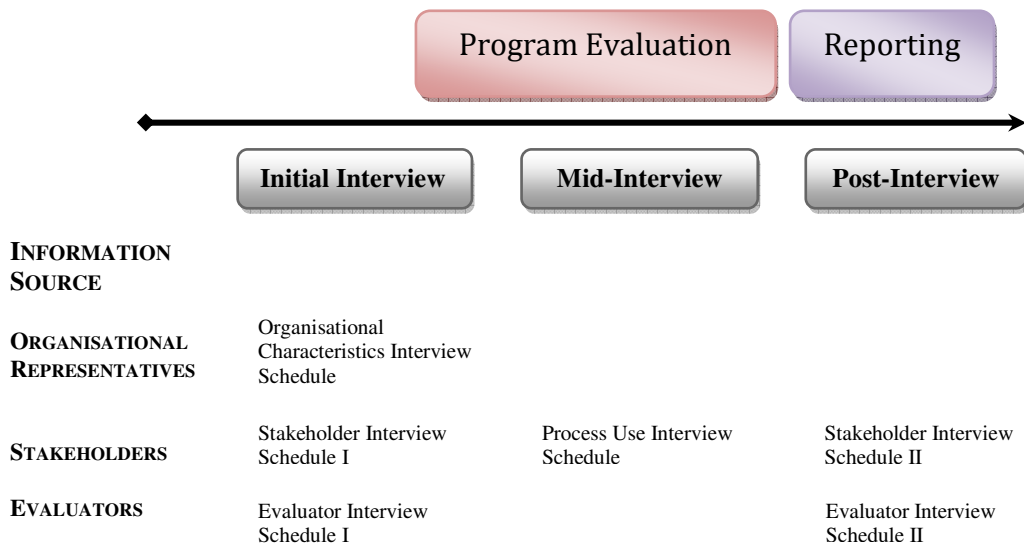
courses are primarily institution-based and include a work experience component. The primary stakeholders of the program were: the Department of Education and Training; Technical and Further Education Colleges; the State Training Board; the Training Accreditation Council; relevant industry groups; and the students. The evaluation study of the Pre-Apprenticeship Program was selected for its strong use of program implementation theory.

*Making Consistent Judgments Professional Learning (MCJ) Module.* The program aimed to make the Year 3 and Year 9 standards for Mathematics and English explicit and support teachers to implement a moderation process endorsed by the system, to enable the making of consistent judgments in relation to these standards. The primary stakeholders of the program are: the Department of Education and Training; the Curriculum Directorate; the Australian Government Quality Teacher Program; the Curriculum Council; the Western Australian Primary Principals' Association; the Western Australian Secondary Schools Association; the teachers of Western Australia; and students and parents (Guardians). The *MCJ Module* evaluation was selected for its apparent disregard of theory. In fact, it was later found to have made some use of program implementation theory and a limited use of program causative theory in the evaluation study.

The data for the investigation came from formal interviews with members of three key groups as well as some other additional sources. The interviews were conducted with the stakeholders of the three programs, the evaluators of each program and representatives of each of the organizations responsible for program delivery. In addition, given the complex nature of the use of the evaluation information, document review, observation and informal communication with key players were also used to collect qualitative information which provided valuable insight into the quantitative findings of the study.

For each study interviews were conducted at key points in the evaluation process, i.e. at the beginning, midway through the conduct of the study, and following the release of the final report (refer to Figure 2). Six interview schedules were developed for data collection focussing on the characteristics of the host organisation of each program, the views of stakeholders, the characteristics and views of the evaluators, and the process use of the information emerging from each study. All six interview schedules contained a mix of open and closed response items. The information gathered from the open response items was expected to add insight and aid in the interpretation of the quantitative data. All of the

rating scale items used a continuum from 1 ('Not at all') to 5 ('A great deal'). A total of 24 scales were developed to provide measures of the observed variables which were the indicants of the latent variables of the Model. Reliability analyses found that each scale had a Cronbach's Alpha  $\geq 0.70$ . Items with low inter-item correlations ( $r_{ii'} \leq 0.30$ ) or low item-total correlations ( $r_{it} \leq 0.30$ ) generally were deleted from its scale, although consideration was given to the contribution of the item to the range of concepts that are of focus in that scale.



**Figure 2: Study Procedure**

## FINDINGS OF THE STUDY

Two sets of model versions were tested and comparisons of model fit considered, with a view to gaining insight into the influence of the predictor variables on the outcome variations. One set included the Program Causative Theory variable and the four outcome variables, while the other set included the Program Implementation Theory variable and the four outcome variables. This strategy was expected to facilitate model fit in view of the small sample size of the investigation. Furthermore, in consideration of the simpler

models, it was expected that the influences of the predictor variables on the outcome variables, both direct and indirect, would be clearer.

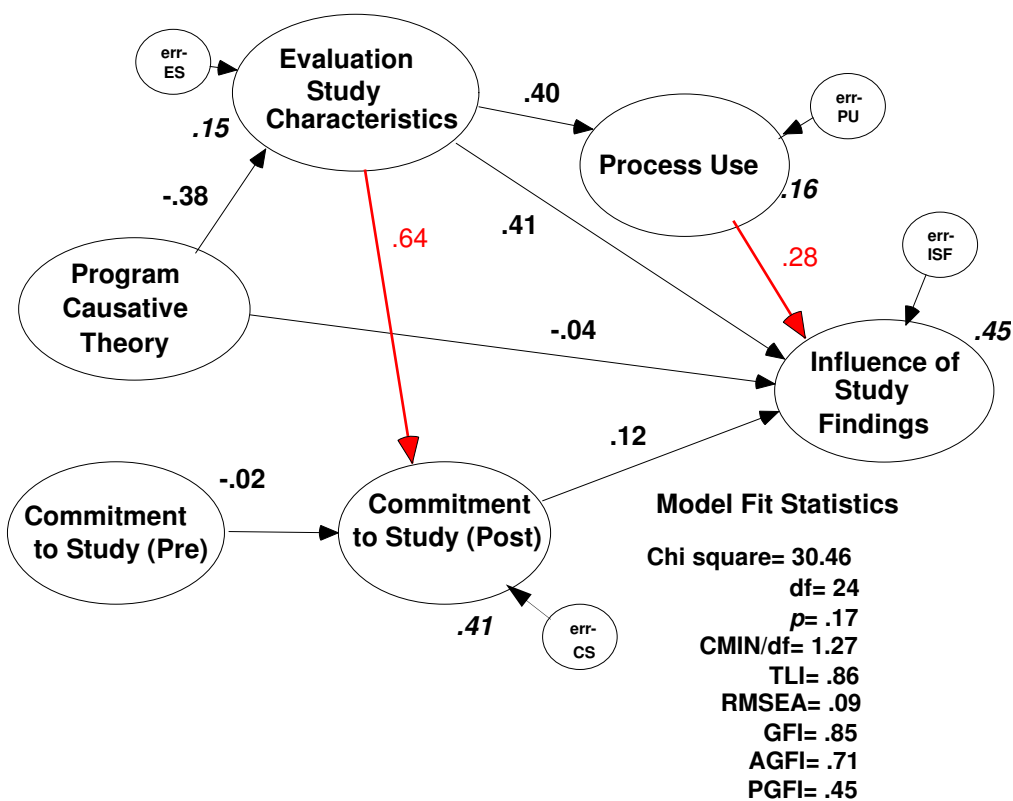
The fit indices selected for evaluation of the measurement model were also used to determine the fit of the structural model with the data from which it was estimated. In addition, the need to address the issue of parsimony in the structural model led to the inclusion of the Parsimony Goodness-of-Fit index (PGFI) in the evaluation of the model. To complement the PGFI measure, the Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI) measures were also included. The fit indices and the acceptable thresholds for each are represented in Attachment 1. Analysis of Moment Structures, or AMOS version 7.0 (Bryne, 2001), is the structural equation modelling computer package used to conduct the analysis of the model for this investigation.

Standardised indirect, direct and total effect size of each variable for each version of the model was considered with regards to the influence of the variables of the model<sup>1</sup>. Furthermore, standardised regression weights were used to determine the significance of relationships between variables in the models. With regards to regression weights, the 0.10 significance level threshold was adopted rather than the more common and rigorous 0.05 level for three reasons. First, the small sample size of the investigation may compromise the accuracy of the statistics the investigation yields. Second, the investigation rests on the investigation of only three case studies and the degrees of freedom available for the statistical analysis of the data are small. Finally, this investigation deals with a relatively unexplored area of research. One expected outcome of the present investigation is to identify areas with value for further research.

---

<sup>1</sup> Holmes-Smith (personal communication, July 24, 2007) suggests any effect size value below 0.2 be considered weak, a value between 0.2 and 0.3 be considered small, while values between 0.3 and 0.5 indicate a stronger effect values between 0.5 and 0.8 indicate a very strong effect.

A total of eight versions of the model were tested. The versions which include *Influence of Study Findings* (model version 1) as an outcome variable fit the data well. The general results for this model are presented in Figure 3. Model version 1a (*Program Causative Theory*) and model version 1b (*Program Implementation Theory*) each explained 45% of the variance in the outcome variable *Influence of Study Findings* and yielded very similar fit indice values. The model versions which include the *Influence of Use of Program Theory in the Final Report* (model version 2) outcome variable also fit the data well. Model version 2a (*Program Causative Theory*) and model version 2b (*Program Implementation Theory*) explained 27% and 28%, respectively, of the variance in the outcome variable. The remaining four versions of the model were found to not fit the data sufficiently well.



**Figure 3: Structural Model 1a**



## DISCUSSION OF FINDINGS

The three research questions which were the focus of this investigation are used here to summarise the conclusions of the investigation.

*What is the influence of program theory on the use of evaluation information?*

The findings indicate the Program Theory, defined as neither Program Implementation Theory nor Program Causative Theory, had no significant influence on the outcome variable. However, as the data was drawn from only three studies, the contextual issues relevant to each have significant opportunity to influence the findings. A consideration of the qualitative information gathered from stakeholders gives some insight into the findings, indicating the context of the programs and the evaluations, and not necessarily the use of program theory in the evaluations, has likely limited the effect of program theory use on the outcome variables.

When first interviewed, all YOHFest Program stakeholders involved in the program theory elaboration process were extremely positive regarding the gains they had made in terms of program insight by being involved in the process. Unfortunately, it seems other characteristics of the evaluation have limited the use of the YOHFest Program evaluation information. The stakeholders of the YOHFest Program evaluation were very unhappy with some of the characteristics of the evaluation study. They were unhappy that the final report did not contain the quantitative data they required to support their request for further funding of the program. All also had concerns regarding some of the data collection methods employed in the evaluation and were disheartened that the final report was delayed by nearly a year from the expected date of delivery. They all were, however, very happy with the program theory elaboration process conducted by the evaluators, which occurred early in the evaluation.

Similarly, stakeholders of the Making Consistent Judgments Program evaluation were very happy with the use of program theory in the evaluation. However, issues regarding

---

*The Influence of Program Theory-Based Evaluation on the Use of Evaluation Information: A Longitudinal Study*

*Sheri Hudson, Ralph G Straton and Richard Cummings  
Paper Presented at the Australasian Evaluation Society Conference, Canberra, 2009*

the nature of program implementation (the program was centrally driven from the head office of the Department of Education and Training) made actual use of the evaluation information difficult for those located in the district offices and schools. Furthermore, because the program was so strongly driven from the central office there was some disillusionment among stakeholders in the district offices and schools that their opinions had little weight with regards to both the evaluation and the program.

Finally, qualitative data gathered from stakeholders of the Pre-Apprenticeship Program evaluation indicate use of the evaluation information may have been limited by the Department of Education and Training's consideration of alternative avenues for training young people in Western Australia, and there was likelihood that the Pre-Apprenticeship Program would be discontinued. There was also some concern among stakeholders regarding the statistics included in the final report of the evaluation which led them to be cautious about relying on the information.

*Which factors have the greatest impact on the use of evaluation information?*

The *Evaluation Study Characteristics* variable appears to be the strongest predictor variable of the model versions with *Influence of Study Findings* as the outcome variable and with *Influence of the Use of Program Theory in the Final Report* as the outcome variable. In all model versions except 1a, the *Evaluation Study Characteristics* variable has indirect influence on the outcome variable of the model, either through *Process Use* or *Commitment to the Study (Post)*. In all instances, the strength of this indirect effect is strong. The *Influence of Study Findings* variable is significantly affected by the *Process Use* variable in model versions 1a and 1b, indicating that involvement in the evaluation process has a positive influence on the use of the study findings.

*How do these factors interact with each other to affect use?*

Regression estimates of model version 1a indicate the *Program Causative Theory* variable has no significant direct effect on any other variable. However, the regression estimate

values of model version 2a indicate the *Program Causative Theory* variable has a small negative effect on the *Evaluation Study Characteristics* variable. Furthermore, regression estimate values indicate that the *Program Implementation Theory* variable has a significant, but small, positive effect on the *Evaluation Study Characteristics* variable in all models of the investigation which were found to fit the data.

Model version 1a indicates a significant link between the *Evaluation Study Characteristics* variable and the *Commitment to Study (Post)* variable, the *Process Use* variable and the *Influence of Study Findings*. In most of the model versions which fit the data, the link between the *Evaluation Study Characteristics* variable and the *Commitment to Study (Post)* variable is significant.

## CONCLUSION

This investigation examined the influence of the use of program theory in an evaluation study on the use of the evaluation and its information. Using a longitudinal design, three evaluation studies with varying strengths of program theory use were the focus of the investigation. Although there were a small number of studies, a number of implications for evaluation practice were evident.

The investigation results indicate that the use of program theory in the undertaking of an evaluation study has no direct impact on the use of the evaluation's information. Alternatively, program theory use in the evaluation was found to have a significant indirect impact on the use of evaluation information through the stakeholder's perceptions of the characteristics of the evaluation study. These influences, if they exist, are apparently overwhelmed by other factors such as the characteristics of the evaluation study and issues related to organisational use of evaluation information.

This investigation has was not able to confirm claims by Bickman (1987), Chen (2004), and Weiss (1998) that a more valuable and useful study will result if an evaluation is based on program theory. It does, however, highlight the need for evaluation practice to be subjected to empirical tests to validate the practice. Good quality empirical research will

enable evaluation practitioners to have greater confidence that they are maximising the likelihood the information provided by their evaluations will be used effectively.

#### REFERENCE LIST

- Alkin, M. and Associates (1985). *A Guide for Evaluation Decision-Makers*. Beverly Hills, California: Sage.
- Arbuckle, J. 1995-2005. AMOS<sup>tm</sup> 6.0 User's Guide. AMOS Development Corporation: Spring House, PA.
- Bickman, L. (1987). The importance of program theory. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18. San Francisco: Jossey-Bass.
- Byrne, B. (2001). *Structural Equation Modelling with AMOS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum.
- Chen, H. (1990b). *Theory-Driven Evaluations*. Newbury Park, CA: Sage.
- Chen, H. (2004). *Practical Program Evaluation, Assessing and Improving Planning, Implementation, and Effectiveness*. London: Sage.
- Cousins, J. and Leithwood, K. (1986). Current empirical research on evaluation utilisation. *Review of Educational Research*, 56, 331-364.
- Cousins, J. and Leithwood, K. (1993). Enhancing knowledge utilisation as a strategy for school improvement. *Knowledge: Creation, Diffusion and Utilisation*, 14, 305-333.
- Cummings, R. (1997). *The Influence of Stakeholder Involvement in Evaluation Studies on the Use of Evaluation Information: A Longitudinal Study*. Unpublished Doctor of Philosophy dissertation, Murdoch University, Western Australia.
- Holmes-Smith, P. (July 24, 2007). Personal communication.
- Hudson-Mabbs, S. (1993). *Influences on the use of evaluation information: Three case studies*. Unpublished Master of Education dissertation, Murdoch University, Western Australia.
- Leviton, L. C. and Hughes, E. (1981). Research on the utilisation of evaluations: A review and synthesis. *Evaluation Review*, 5(4), 525-548.
- Pawson, R. and Tilley, N. (1996). What's crucial in evaluation research: A reply to Bennett. *British Journal of Criminology*, 36, 574-578.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- Preskill, H. and Torres, R. (1999). *Evaluative Inquiry for Learning in Organisations*. Thousand Oaks: Sage.
- Rogers, P., Hacsı, T., Petrosino, A., and Huebner, T. (2000). Program Theory in Evaluation: Challenges and Opportunities. *New Directions for Evaluation*, 87, San Francisco, CA: Jossey-Bass.

---

*The Influence of Program Theory-Based Evaluation on the Use of Evaluation Information: A Longitudinal Study*

Sheri Hudson, Ralph G Straton and Richard Cummings  
Paper Presented at the Australasian Evaluation Society Conference, Canberra, 2009

Schereir, M. (1987). Program theory and implementation theory: Implications for Evaluators. In L. Bickman (Ed.), *New Directions for Program Evaluation*, 33, 5-18.

Weiss, C. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review*, 21, 501-524.

Weiss, C. (1998). *Evaluation: Methods for Studying Programs and Policies*. (2<sup>nd</sup> Ed). Englewood Cliffs, N.J.: Prentice Hall.

**Attachment 1: Fit Indices Adopted to Determine Measurement and Structural Model Fit**

| FIT INDICES  | THRESHOLD VALUES USED TO DETERMINE MODEL FIT |
|--|--|
| $\chi^2$   | >1 but < 3                                   |
| $p^1$  | 0.10 level of significance                   |
| CMIN/df <sup>2</sup>   | >1 but <2                                    |
| RMSEA <sup>3</sup>   | < 0.10                                       |
| Tucker-Lewis Index (TLI) <sup>4</sup>  | > 0 but <1                                   |
| GFI  | > 0 but <1                                   |
| AGFI   | > 0 but <1                                   |
| PGFI   | > 0 but <1                                   |
| Regression Weights   | 0.1 level of significance                    |
| Notes <sup>1</sup> A probability value for Chi Square test smaller than 0.05 indicates the null hypothesis that the model fits the data should be rejected.<br><sup>2</sup> Holmes-Smith (2001) suggests CMIN/df values should be greater than 1 but less than 2. However, he considers values between 2 and 3 are to be indicative of reasonable fit while values less than 1 indicate over fit.<br><sup>3</sup> The more generous upper threshold (0.1) of Browne and Cudeck (1989) has been used in the evaluation of model fit.<br><sup>4</sup> Generally, adjunct fit indexes (TLI) range between 0 and 1.0. A value of 0.90 is widely considered to be the threshold an index must exceed before a model is considered to be consistent with the observed data from which it was estimated (Hoyle, 1995; Byrne, 1995). |  |