VII European Conference on Psychological Assessment
VI Congreso Nacional de Evaluación Psicológica
EAPA- 2004

Málaga
Del 1 al 4 de Abril, 2004

Simposia Proposal

Methodological Innovations in Program Evaluation

Salvador Chacón Moscoso
Dpto. Psicología Experimental
Facultad de Psicología
Universidad de Sevilla
C/ Camilo José Cela s/n
41018. Sevilla. (Spain)
Teléfono: 95-4557812 Fax: 95-4551784
Correo electrónico: schacon@us.es

General abstract

It is assumed that intervention programs have to be evaluated rigorously. Nonetheless, we find important methodological weaknesses in program evaluation practise. We consider that trained people in methodology should communicate and spread models, assumptions, conditions and methodological limitations where evaluative processes are base on.

In order to flourish valid design, measurement and analysis that increase quality in program evaluation practise, we present different methodological innovation in program evaluation from a theoretical and a practical point of view. First, we will present two papers related to the application of causal analysis from Rubin's Model into program evaluation; by one hand, we will present a general description of threats to validity from causal analysis and by the other hand, we will present specifically the analysis of causal effects in the non-equivalent control group design. Second, we will introduce one of the most important topics in measurement applied into program evaluation: The use of Item Response Theory Models in Program Evaluation, and one of its possible applications into Computerized Adaptive Testing in the context of training program evaluation. Third, we will present another important issue in low intervention design evaluations: The efficacy in a program of low intensity through lag-log observational design.

Contributions:

Paper 1. Threats to validity from causal analysis

Salvador Chacón-Moscoso (University of Sevilla, Spain)
Rolf Steyer (University of Jena, Spain)
Jose López Ruiz (University of Sevilla, Spain)
Fco. Pablo Holgado Tello (UNED, Spain)
José Antonio Pérez-Gil (University of Seville, Spain)

Abstract

In program evaluation practice, there is not a systematic way to control treats to validity and its consequences in effect size estimations. In this intervention context, Campbell's approach has given a conceptual framework for evaluating main threats to the various kinds of validity. His original work emphasized concepts from philosophy of science and the practical issues confronting social researchers. Nonetheless, there has not been much effort to systematize this conceptual framework, for example, clarifying key concepts such as plausibility. This paper calls for an empirical analysis of validity threats using Rubin Causal Model (RCM) mainly focused on formal statistical criteria for inference. Based on Shadish, Cook and Campbell we will describe the conceptual bases of validity and main threats to the various kinds of validity. Then, mainly based on Rubin's work, we will move to analyze main concepts of statistical inference for causal effects in experiments and observational studies. The use of Rubin causal model can help us to identify strengths and limitations of specific plausible threats to validity from an emprirical point of view. The global objective of this paper is to translate conceptual problems of validity threats into possible applications of RCM. This work will have ulterior developments analyzing actual practice in program evaluation when obtaining effect size estimation in specific contexts with different intervention designs and validity threats.

Paper 2. The analysis of causal effects in the non-equivalent control group design.

Rolf Steyer (University of Jena, Germany)
Felix Flory(University of Jena, Germany)

Abstract

Traditionally, data from non-equivalent control group designs have been analysed by methods based more or less on plausibility considerations. In the last years, however, based on Rubin´s approach to causality, we can derive which analyses yield unbiased estimates of the treatment effect under which assumptions. We outline the crucial assumptions, show how to test them and present a general procedure, how to analyse data obtained in such a design when the goal is to estimate the average treatment effect. Furthermore, it shown that the techniques presented also serve to analyse data in the non-orthogonal analysis of variance.

Paper 3. The use of Item Response Theory Models in Program Evaluation

José Muñiz (University of Oviedo, Spain)

Abstract

When professionals and researchers are involved in program evaluation, very frequently they use psychological and educational tests, scales, questionnaires, and other evaluation instruments. The psychometric technology applied to develop and analyse the evaluation instruments used is very often based on classical test theory, ignoring the important advantages that the use of new psychometric technologies could add to their data. In this presentation, some of the new psychometric tools are presented, paying special attention to those innovations derived from the Item Response Theory (IRT) models. The advantages of these IRT models with respect to the classical approach are underlined, the foundations of the models presented, and some practical implications described. The presentation will be focused on six main innovations derived from the IRT models: a) The estimation of the test reliability (Items and test Information Functions), b) Differential Item Functioning (bias evaluation), c) Computerized Adaptive Testing, d) Item Banks, e) New Item formats, and f) Advances in item writing technology. Finally, the implications of these IRT innovations on validity will be analysed, and future perspectives discussed.

Key words. Item Response Theory. Information Function. Differential Item Functioning. Computerized Adaptive Testing.

Paper 4. Computerized Adaptive testing (CAT) in training program evaluation.

Fco. Pablo Holgado Tello (UNED, Spain)
Salvador Chacón Moscoso (University of Sevilla, Spain)
Juan Antonio Ruiz Rivas (Universidad de Sevilla, Spain)
Juan Carlos Suarez Falcón (UNED, Spain)
Mª Isabel Barbero García (UNED, Spain)
José Antonio Pérez-Gil (University of Seville, Spain)

Abstract

One of the main problems in training program evaluation is to define criteria to certify level of knowledge and/or ability of participants in a specific training program. This problem is a key point to assess effect size of the program as well as to assign participants to different complexity levels of the training program.

A single test administered to a group of examinees cannot provide the same precision of measurement to every subject. An efficient way to solve the problem is to use computerized adaptative tests (CAT) for certification. CAT can be used to give every examinee test that is adapted to the examinee's ability level (basic, medium or high). Based on the examinee's prior performance, items that are maximally informative about the examinee's ability level are administered. This procedure allows reducing number of items of the test and increasing ecological validity in the evaluation process.

In this paper, we describe the design and implementation process of a CAT applied to ofimatic contents in the Training Centre of University of Seville. The main objective is to certify participants' ofimatic ability level (basic, medium and high), and to obtain criteria in order to assign subjects to the appropriate level of the training intervention. We present the interfaces and the procedures developed for items presentation and the items calibration under different sample size in order to analyze accuracy and bias in obtained estimations.

We emphasize the utility of this procedure to certify participants' ability levels with reliability, validity and equity. Also CAT use increase the efficiency and resources optimization in evaluation.

Paper 5. Searching the Efficacy in a Program of Low Intensity through Lag-Log Observational Design

Eulàlia Arias Pujol (University Ramon jull, Spain)
M. Teresa Anguera Argilaga (University of Barcelona, Spain)

Abstract

A program of low intervention was implemented whose objective it was the improvement of the communicative performance in adolescents by means of the analysis of the conversation. The program is characterized as unobtrusive, and taking place in a context initially not known by the users, but later was habitual for them.

The users were adolescent with low academic performance and interactive problems. The number of sessions was thirty. The users group has been developed through a requests received in the corresponding section of a public hospital of Barcelona, and the waiting list allowed to select the components of "communication group" that they needed psychotherapeutic support in function of the established criteria.

It has used the observational methodology for data collection, and it has recorded the verbal and nonverbal behavior in the course of conversation between users and between users and therapist. Two category systems have been made, that later both were integrated in a dimension of communicative behavior. The analysis of the psychotherapeutic efficacy has been carried out by means of linear trend by the method of least squares, sequential analysis, and polar coordinates analysis.

The obtained results show an increase of the participation and the spontaneous behaviour of users throughout the implementation of the program, reduction of their inhibition, and appearance of new communicative forms.