

# **The Program Evaluation Standards in International Settings**

Edited by Craig Russon

Papers by

Nicole Bacmann  
Wolfgang Beywl  
Saviour Chircop  
Soojung Jang  
Charles Landert  
Prachee Mukherjee  
Nick Smith  
Sandy Taut  
Thomas Widmer

*The Evaluation Center  
Occasional Papers Series*

May 1, 2000

# **The Program Evaluation Standards in International Settings**

Edited by Craig Russon

Papers by

Nicole Bacmann  
Wolfgang Beywl  
Saviour Chircop  
Soojung Jang  
Charles Landert  
Prachee Mukherjee  
Nick Smith  
Sandy Taut  
Thomas Widmer

*The Evaluation Center  
Occasional Papers Series*

May 1, 2000

**The editor acknowledges the valuable assistance of Brian Carnell, Karen Russon, Ragnar Schierholz, Sandy Taut, and Sally Veeder in the preparation of this manuscript.**

## Table of Contents

|   |     |
|---|-----|
| Foreword .....  | 1   |
| Cross-Cultural Transferability of <i>The Program Evaluation Standards</i> by Sandy Taut .....   | 5   |
| Considerations on the Development of Culturally Relevant Evaluation Standards by Nick L. Smith, Saviour Chircop and Prachee Mukherjee ..... | 29  |
| The Appropriateness of Joint Committee Standards in Non-Western Settings : A Case Study of South Korea by Soojung Jang .....                | 41  |
| Standards for Evaluation: On the Way to Guiding Principles in German Evaluation by Wolfgang Beywl .....                                     | 60  |
| Evaluating Evaluations: Does the Swiss Practice Live up to <i>The Program Evaluation Standards</i> ? by Thomas Widmer .....                 | 67  |
| Evaluation Standards Recommended by the Swiss Evaluation Society (SEVAL) by Thomas Widmer, Charles Landert, and Nicole Bacmann .....        | 81  |
| Annex: A Summary of the <i>The Program Evaluation Standards</i> .....   | 103 |

## Foreword

On February 18-20, 2000, the W. K. Kellogg Foundation-funded residency meeting of regional and national evaluation organizations took place in Barbados, West Indies. Several organizations were represented at the meeting: African Evaluation Association, American Evaluation Association, Asociacion Centroamericana de Evaluacion, Associazione Italiana de Valutazione, Australasian Evaluation Society, Canadian Evaluation Society, European Evaluation Society, Israeli Association for Program Evaluation (IAPE), Kenyan Evaluation Association, La Societe Francaise de l'Evaluation (SFE), Malaysian Evaluation Society, Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean (PREVAL), Reseau Ruandais de Suivi et Evaluation (RRSE), Sri Lanka Evaluation Association (SLEvA), and the United Kingdom Evaluation Society. In addition, representatives from the W. K. Kellogg Foundation, The University of the West Indies, the Caribbean Development Bank, and the United Nations Capital Development Fund were present.

One of the issues around which a lively debate ensued was *The Program Evaluation Standards* (The Joint Committee on Standards for Educational Evaluation, 1994). The Joint Committee has always maintained that *The Standards* are uniquely American and may not be appropriate for use in other countries without adaptation. Several national evaluation organizations from developing countries are anxious to undertake projects to create standards suitable for use in their own cultural contexts as a means to guide practice in their countries.

Some of the evaluation organizations from developed countries were less enthusiastic. They view standards as a barrier to innovation. They think that the word *standards* implies imposition of best practices. (Developing countries are not married to the term standards.) Furthermore, they think that no model can apply to all evaluations. The participants at the Barbados meeting postponed resolution of the issue by saying that it was not appropriate to make a determination regarding this matter at the international level. Rather, this decision should be made by each country.

It is in the context of this debate that The Evaluation Center is pleased to print the seventeenth volume in the Occasional Papers Series entitled, *The Program Evaluation Standards in International Settings*. The volume contains six outstanding papers. The first paper is entitled "Cross-Cultural Transferability of *The Program Evaluation Standards*" and was written by Sandy Taut, a Fulbright scholar at The Evaluation Center. In this paper, Taut examines the theory behind the use of *The Standards* in international settings. She states that *The Standards* are values based. Taut then proceeds to examine several value dimensions such as individualism-collectivism, hierarchy vs. egalitarianism, etc. Finally, she applies the previously discussed value dimensions to each of the 30 *Program Evaluation Standards*.

The second paper, entitled "Considerations on the Development of Culturally Relevant Evaluation Standards," was written by Nick Smith, Saviour Chircop, and Prachee Mukherjee.

This article originally appeared in *Studies in Educational Evaluation* in 1993. (Many thanks to Elsevier Science for permission to reprint.) It is one of the earliest attempts to look at the application of *The Standards* in international settings—one referenced by most of the other papers in the monograph. Within the contexts of Malta and India, the authors explored the extent to which *The Standards* could be used as the basis for developing comparable, culturally relevant standards for indigenous use and the ways they could be modified for use by U.S. evaluators in cross-cultural settings.

The third paper is entitled “The Appropriateness of Joint Committee Standards in Non-Western Settings: A Case Study of South Korea” by Soojung Jang. This is the paper for which Jang won the student paper competition at the 1999 conference of the American Evaluation Association. Jang conducted a workshop on the standards for Korean graduate students and administered a questionnaire to workshop participants. The study aimed to establish a framework of advice that can be applied to Korean evaluation projects and to guide western evaluators who wish to develop culturally effective techniques.

The fourth paper, entitled “Standards for Evaluation: On the Way to Guiding Principles in German Evaluation”, was written by Wolfgang Beywl, who translated *The Program Evaluation Standards* into German. This paper was originally published in a 1999 volume of the German Evaluation Society’s newsletter. It’s purpose was to initiate development of a unique set of German evaluation standards.

The fifth paper is entitled “Evaluating Evaluations: Does Swiss Practice Live Up to *The Program Evaluation Standards*?” by Thomas Widmer. This is a paper that Widmer presented at the historic International Evaluation Conference that was held in 1995 in Vancouver, British Columbia. In the paper, Widmer presents the results of a metaevaluation of 15 Swiss programs that he conducted using *The Standards*. Recommendations are also included for adapting *The Program Evaluation Standards* to the Swiss context. This paper foreshadowed the important efforts carried out by the Swiss that are detailed in the last paper.

The sixth, and last paper, is entitled “Evaluation Standards Recommended by the Swiss Evaluation Society (SEVAL)” by Thomas Widmer, Charles Landert, and Nicole Bacmann. In this paper discusses the evaluation standards developed by SEVAL. These standards are in part based on *The Program Evaluation Standards*. This is an historic first. These are the first evaluation standards to be developed in a country outside of the USA. The experience of the Swiss may be a model for other countries to follow.

Analysis of the findings of the three empirically based papers (see Table 1) shows a substantial amount of agreement among Smith et. al. and Jang on the difficulty of applying the Propriety Standards in Malta, India, and Korea. Widmer found less difficulty with the Propriety standards in Germany. Perhaps the sense of propriety in a Western country such as Germany is closer to that found in the USA than is the sense of propriety found in non-Western countries such as Malta or India.

Hopefully, this volume will contribute to the understanding of applications of *The Program Evaluation Standards* in international settings. Evaluators outside the USA should not hesitate to use the Standards they feel are appropriate. Someday these evaluators may wish to undertake the task of developing their own culturally relevant standards. Should they wish to do this, the process followed by the Joint Committee for Program Evaluation Standards (Russon, 1999) may be a helpful guide.

Table 1: Comparison of Findings From Three Empirical Papers

| <b>Standard</b>                         | <b>Smith</b> | <b>Jang</b> | <b>Widmer</b> |
|---|--------------|-------------|---------------|
| U1 Stakeholder Identification           |              | X           |               |
| U2 Evaluator Credibility                |              |             |               |
| U3 Information Scope and Selection      |              |             |               |
| U4 Values Identification                |              |             |               |
| U5 Report Clarity                       |              |             |               |
| U6 Report Timeliness and Dissemination  |              |             |               |
| U7 Evaluation Impact                    |              |             |               |
| F1 Practical Procedures                 |              |             |               |
| F2 Political Viability                  |              |             |               |
| F3 Cost Effectiveness                   |              |             | X             |
| P1 Service Orientation                  |              | X           | X             |
| P2 Formal Agreements                    | X            |             |               |
| P3 Rights of Human Subjects             | X            | X           |               |
| P4 Human Interactions                   | X            |             |               |
| P5 Complete and Fair Assessment         |              |             | X             |
| P6 Disclosure of Findings               | X            | X           |               |
| P7 Conflict of Interest                 | X            | X           |               |
| P8 Fiscal Responsibility                |              |             |               |
| A1 Program Documentation                |              |             |               |
| A2 Context Analysis                     |              |             |               |
| A3 Described Purposes and Procedures    |              |             |               |
| A4 Defensible Information Sources       |              | X           | X             |
| A5 Valid Information                    |              |             | X             |
| A6 Reliable Information                 |              |             |               |
| A7 Systematic Information               |              |             |               |
| A8 Analysis of Quantitative Information |              |             |               |
| A9 Analysis of Qualitative Information  |              |             |               |
| A10 Justified Conclusions               |              |             |               |
| A11 Impartial Reporting                 |              |             |               |
| A12 Metaevaluation                      |              | X           |               |



**Cross-Cultural Transferability  
of *The Program Evaluation Standards***

*“ . . . there is no such thing as human nature independent of culture” (Geertz,  
1973)*

Sandy Taut  
Visiting Scholar  
The Evaluation Center

Abstract

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 Utility and Propriety Standards, in particular, 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.

## Introduction

The purpose of this paper is to analyze the cross-cultural applicability of *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994). A special emphasis is placed on the Standards' implicit cultural values because these values make the transfer of the Standards to other value contexts (i.e., cultures) difficult. In cross-cultural psychological literature, the concept of *standards* is used to define *values*, e.g., as Smith and Schwartz (1997) put it, "values serve as standards to guide the selection or evaluation of behavior, people, and events" (in: Berry, Segall & Kagitçibasi (ed.), 1997, Vol. 3, p. 80). According to Peoples and Bailey's (1994) seminal anthropological textbook, *values* are defined as "providing the ultimate standards that people believe must be upheld under practically all circumstances" (p. 28). In addition, Smith and Schwartz (1997; in: Berry et al., Vol. 3, p. 79) state that "[...] value priorities prevalent in a society are a key element, perhaps the most central, in its culture, and the value priorities of individuals represent central goals that relate to all aspects of behavior." These researchers further note that "as standards, cultural value priorities also influence how organizational performance is evaluated – for instance, in terms of productivity, social responsibility, innovativeness, or support for the existing power structure" (p. 83).

The first assumption derived from these statements is that *The Program Evaluation Standards* are values-based. The second assumption, supported by a large amount of research on cultural values (e.g., see Berry et al., 1997, Vol. 3, chapter 3), is that values differ across cultures: "cultures can be characterized by their systems of value priorities" (p. 80). Members of one cultural group share many value-forming experiences and therefore come to accept similar values.

The discrepancy in values is not only apparent between the North American and other cultures around the world, but also, for example, within the United States. There, one can identify some cultural subgroups whose values deviate from "mainstream America," e.g., Native Americans (see Miller, 1997; in: Berry et al., 1997, Vol. 1, p. 105).

In contrast to values, *norms* are "shared ideals (rules) about how people ought to act in certain situations, or [...] toward particular other people" (Peoples & Bailey, 1994, p. 28). Norms are based on values. In some of the Standards (e.g., Formal Agreements), the more concrete norms are reflected in lieu of the higher-level values.

When analyzing the Program Evaluation Standards it seems important to refer to the Joint Committee's definition of *standards* and to address the Joint Committee's conclusions regarding their level of generality (across cultures and even within North America). The Joint Committee (1994) defines a *standard* as "a principle mutually agreed to by people engaged in a professional practice, that, if met, will enhance the quality and practice of that professional practice, for example, evaluation" (p. 2). The Joint Committee acknowledges "that standards are not all equally applicable in all evaluations" (p. 2), that they are "guiding principles, not mechanical

rules” (p. 8) and, implicitly, that they are based on North American values: “They [the Standards] define contemporary ideas of what principles should guide and govern evaluation efforts” (pp. 17-18). In a section on guidelines for application of the Standards, the third step is to “clarify the context of the Program Evaluation” (p. 11), the cultural context unquestioningly being of major importance. Stufflebeam (1986), Chair of the Joint Committee from its birth in 1975 until 1988, notes that “their [the Standards’] utility is limited, especially outside the United States” (p. 18). Evaluators have called for more research on their transferability (e.g., Nevo, 1984, p. 386). The following paper is an investigation along this dimension.

With evaluation recently becoming a stronger professional field in many countries, Standards for professional conduct also become an important issue. The process of developing Standards is effortful and cost-intensive (see Joint Committee, 1994, pp. xvi-xviii; Stufflebeam, 1986, pp. 2). Therefore, it is only logical that evaluators from different countries turn to the Program Evaluation Standards as a model for their own work. The Standards have been translated into German, after Beywl (1996, 1999) concluded that – with minor adaptations – they are transferable to the German cultural context. The Swiss Evaluation Society developed Standards closely following the North American model (see SEVAL-Standards, Schweizerische Evaluationsgesellschaft, 1999). However, the Schweizerische Evaluationsgesellschaft did not specifically address the degree of transferability of the Standards. Neither did the Spanish translation of the Standards (Villa, 1998). During a 1998 UNICEF Evaluation Workshop in Nairobi, Kenya, a focus group discussion dealt with the applicability of the Standards in African cultures (Russon & Patel, 1999, p. 4). Of the nine African evaluators, the majority found the Standards to be transferable if some modifications were undertaken. In their eyes, the Standards are not laden with values incompatible with African values. During the discussion, the participants suggested adaptations of twelve Standards (see Russon & Patel, 1999, pp. 5). A minority held that African Evaluators should develop their own Standards. The Standards have also been used in Israel (see Nevo, 1984; Lewy, 1984), Brazil (see Rodrigues de Oliveira & Gomes-Elliot, 1983), and Sweden (see Marklund, 1984). These authors conclude that the importation of the Standards requires at least some adaptation to fit the host culture but in few countries has the transferability issue been addressed in detail. Consequently, an in-depth analysis of the Standards’ general cross-cultural transferability is crucial and overdue.

The present paper does not investigate the likelihood that various cultures meet the Standards based on their cultural idiosyncrasies. Instead, the values the Standards are based on will be examined. It will be shown that in certain cultures these values may limit the applicability of the Standards as guiding principles for evaluators (see Beywl, 1999; Smith et al., 1993; Jang, 2000). In this context, it is important to differentiate between indigenous Standards emerging from within different cultures, cross-cultural Standards intended for evaluators working in similar cultures (or with multicultural stakeholders), and ubiquitous Standards intended to fit all cultures. The latter Standards might need to be highly general or vague to fit all possible contexts, which likely results in less specific direction and therefore less utility of these Standards. As pointed out above, Standards are per se values-based and therefore culture-specific. Consequently, it seems that ‘covering up’ their origin by adding statements that stress generality can only work at the expense of the Standards’ effectiveness as guiding principles.

There is evidence to support the Joint Committee's recommendations that care has to be taken when applying the Standards as *goals* for professional conduct outside the North American ('mainstream') cultural context.

The cross-cultural psychological and anthropological literature revealed values reflected in the Program Evaluation Standards. Researchers from these fields examine cultural value dimensions and a number of other concepts as a way to describe and differentiate cultures. The cultural dimensions and concepts used for the analysis of the Standards will be explained in more detail in the following chapter.

### A Summary of Relevant Cultural Value Dimensions

Cross-cultural psychologists commonly distinguish seven culture areas, namely, West European, Anglo, East European, Islamic, East Asian, Latin American, Japanese (Berry et al., Vol. 3, p. 105). This distinction obviously displays a simplified view of the true diversity and number of cultures all over the world.

As mentioned earlier, cross-cultural literature uses cultural value dimensions to categorize differences between societies. For example, one of the most widely used and researched categories is the *Individualism-Collectivism* dimension. This cultural dimension has also been employed by Smith et al. (1993, p. 12) to explain difficulties in the applicability of the Standards in India and Malta. Some additional common dimensions include *Egalitarianism-Hierarchy* (or *Power Distance*), *Conservatism-Autonomy* and *Mastery-Harmony*. Empirical studies show that these dimensions are valid and reliable (see Berry et al., 1997, Vol. 3). This paper will briefly describe these dimensions, followed by a description of other relevant cultural concepts used in this analysis.

#### Individualism-Collectivism

Philosophic trends beginning with Descartes, followed by British empiricism (Hobbes, Smith), social Darwinism (Spencer), phenomenology and existentialism (Kierkegaard, Husserl, Heidegger, Sartre) made individualism the hallmark of European social history since the beginning of the 16<sup>th</sup> century (see Berry et al., 1997, Vol. 3, p. 4). In contrast, eastern religions and philosophies such as Confucianism, Taoism, Buddhism, Hinduism and Shintoism can be seen as roots of collectivism in these cultural settings. Early monotheistic religions arising in the Middle East (Judaism, Christianity, Islam) also emphasize collective loyalties. However, the European Reformation in Christianity (16<sup>th</sup> century) later introduced individual responsibility (see Berry et al., 1997, Vol. 3, p. 5). Kagitcibasi (1997, in: Berry et al., Vol. 3, p. 5) states that "all societies must deal with tensions between collectivism and individualism, [...] there is some of both everywhere. Nevertheless, differences in emphasis appear to be real."

Hofstede (1980) provided a more recent appraisal of the collectivist-individualist distinction. He conducted a survey on work-related values with 117,000 IBM-employees from 40 (later expanded to 50) countries. A factor analysis revealed four value dimensions:

‘Uncertainty Avoidance,’ ‘Masculinity,’ ‘Power Distance’ and ‘Individualism-Collectivism.’ Hofstede’s (1980, pp. 260f) original definitions of the those last two concepts read as follows: “Individualism stands for a society in which the ties between individuals are loose; everyone is expected to look after himself or herself and his or her immediate family only”; “Collectivism stands for a society in which people from birth onwards are integrated into strong, cohesive ingroups, which throughout people’s lifetime continue to protect them in exchange for unquestioning loyalty.”

Triandis’ research (1990, p. 113) elaborated on Hofstede’s work. Triandis notes that the defining attributes of individualism are “distance from ingroups, emotional detachment, and competition,” while he characterizes collectivism as stressing “family integrity and solidarity.” Based on empirical findings, Schwartz (1990, 1994a,b) refined the individualism-collectivism dimension with two higher order dimensions. At the cultural level, these dimensions are ‘Autonomy vs. Conservatism’ and ‘Hierarchy vs. Egalitarianism.’ Elaboration on the behavioral differences based on the collectivism-individualism dimension and their relation to the analysis of the Standards will be provided in a later section of this paper.

Similar to Hofstede (1980), Schwartz (1994a,b) conducted a value survey including 44,000 respondents (using teachers and students) from 56 countries. After statistical analyses of the data he proposed the following three cultural level value dimensions.

#### Conservatism-Autonomy

The *Conservatism-Autonomy* dimension addresses the relations between individual and group. At the conservatism pole, people are strongly embedded in a collectivity – similar to the definition of collectivism. Schwartz adds the notion of “social order, respect for tradition, [...] and self-discipline” (Berry et al., 1997, Vol. 3, p. 99) to this concept. People in highly autonomous cultures “find meaning in [...] their] own uniqueness”; they are encouraged to express their own preferences, feeling, motives etc. (Berry et al., Vol. 3, p. 99).

#### Hierarchy vs. Egalitarianism

With this dimension, Schwartz (1994a) describes how cultures differ in their way to motivate people to display responsible social behavior. In highly hierarchical societies (e.g., Japan), a system of ascribed roles assures socially responsible behavior. People learn to comply with the demands attached to their roles and to see an unequal distribution of power, roles and resources as legitimate. In contrast, highly egalitarian cultures (e.g., Sweden) contain individuals who are equals and who are socialized to voluntarily cooperate with others (Berry et al., 1997, Vol. 3, p. 100).

#### Mastery vs. Harmony

Schwartz’s (1994a) third dimension deals with the role people play in the natural and social world. In high Mastery cultures, people seek to control and change the natural and social

world and exploit it in order to further personal or group interests. Values include ambition, success, and competence. High Harmony cultures accept the world as it is and try to preserve it rather than to change or to exploit it (Berry et al., 1997, Vol. 3, p. 100).

### Uncertainty Avoidance

Hofstede (1980, p. 110) describes this dimension based on three indicators, “rule orientation, employment stability and stress level.” In an extension of Hofstede’s research, Bond (1987) found some evidence that undermines the universality of this cultural value (Berry et al., 1997, Vol. 3, p. 99), but at the same time his study replicated Hofstede’s other three dimensions. However, other researchers have continued to use the *uncertainty avoidance* dimension.

Other cross-cultural psychological concepts used here to describe differences between cultures are:

### Direct vs. Indirect Communication

Generally, direct communication predominates in individualistic cultures whereas collectivist societies communicate indirectly (Gudykunst & Ting-Toomey, 1988). Self-disclosure (giving personal information to others they don’t know) is associated with direct communication styles. A crucial variable in collectivist cultures is whether communication occurs with members of the ingroup or with members of an outgroup; communication will be more direct (more self-disclosure) among ingroup members compared to communication with outgroup members. In contrast, there are no differences in self-disclosure between ingroup and outgroup in individualistic cultures (Gudykunst et al., 1992).

### High-Context vs. Low-Context

Closely related to the above mentioned communication issue is the distinction between high-context and low-context cultures. Members of collectivist cultures emphasize the importance of context information in explaining others’ behavior more than people in individualistic cultures. This also makes adapting to the context an important part of high-context communication patterns prevalent in collectivist cultures. In addition, Uncertainty Avoidance (UA, see above) influences the degree to which communication follows rules and rituals (high UA means highly ritualistic), especially for outgroup communication/interaction. Lastly, Power Distance (PD) also determines interaction/communication styles; high PD means low levels of egalitarian cooperation between people with different amounts of power. Since outgroup members are outsiders, they tend to have less power than insiders (see Berry et al., 1997, Vol. 3, p. 142).

### Seniority

‘Seniority’ generally refers to the level of one’s social standing based on one’s age. In cross-cultural organizational psychology, the concept of ‘seniority’ is commonly used to describe promotion systems. In contrast to individualistic societies, organizations in collectivist cultures base promotion on employees’ loyalty, seniority, and in-group membership (Berry et al., 1997, Vol. 3, p. 379).

Finally, one needs to address the fact that cultures are constantly changing over time. There are two ways in which *cultural change* comes about: *innovation* (internally generated change) and *diffusion* (adopting traits from another society). People are receptive of new traits when these enhance their economic and social well-being. In addition, political and economic domination, as well as environmental and demographic changes can all introduce cultural changes (see Peoples & Bailey, 1995, p. 105). In an increasingly interdependent world (with a global economy, immigration flows and world-wide communication), cultural change will likely happen more quickly. Since values are reflective of cultural changes, it is also likely that there will be value changes in the form of adaptations to the changing environment. Yet, every cultural setting is unique, thus making values a continuously valid way to differentiate cultures (Berry et al., 1997, Vol. 3, p. 110).

### The Role of Professional Standards Across Cultures

The concept of professionalism in evaluation differs across cultures. As one component of this concept, professional standards for enhancing the quality of evaluators’ performance are of varying importance in different societies. Smith et al. (1993) compare Malta, India and the United States along these lines. In the U.S., professional groups historically regulate themselves through professional standards and compliance with self-generated laws and principles has long existed. In contrast, “[...]in Malta and India, the quality of one’s evaluation work is more likely to be judged on political than on technical or professional grounds. Professional standards would play a much weaker role in regulating evaluation practice [...]” (p. 8). The authors conclude that “the development and use of indigenous professional standards are unlikely, unwelcome, and probably unuseful” (p. 9). In contrast, for example, Widmer (1999) observes for the Swiss context that the set of Standards proposed by the Swiss Evaluation Society (SEVAL) are used more and more as part of regulations, manuals, contracts, and decrees. He posits that, in the German-speaking countries, the Standards serve as a means to build a “*professionell orientiertes Selbstverständnis* [professionally oriented self-perception]” within the evaluation community (p. 8). Beywl (1999) theorizes that the Standards will influence the field of evaluation in Germany almost from its birth while in the U.S. the development of the Standards marked the arrival of a certain level of maturity of the field. He advises the German Evaluation Society (DeGEval) to endorse an adapted version of the Standards as principles for proper conduct. Likewise, Russon & Patel (1999) report from a focus group discussion on the Program Evaluation Standards that the nine participating African evaluators “reached a rapid consensus that there was a need for some set of criteria of quality” in Africa (p. 4).

From a more general point of view, we could explain these cultural differences in the role of professional standards using cultural dimensions and concepts. For example, one could argue that high-conservatism societies tend to not use Standards. In those cultures, the emphasis on the maintenance of the status quo keeps people from engaging in behavior that might disrupt the group solidarity (in this case the group of evaluators). The same may be true for collectivist cultures. In those cultures, a “socially-oriented achievement motivation” (Berry et al., 1997, Vol. 3, p. 25) might prevent individual accountability on the basis of Standards to be valued. In cultures with strong ingroups, “group achievement rather than individual expertise is valued” (Rohlen, 1979, in: Berry et al., 1997, Vol. 3, p. 380). In countries with “Confucian work ethics” (The Chinese Culture Connection, 1987, in: Berry et al., 1997, Vol. 3, p. 373), work is “a moral duty to the collectivity.”

In addition, for example, in China (Chen & Uttal, 1988, in: Berry et al., 1997, Vol. 3, p. 139), “effort is regarded as more important than ability in explaining academic outcomes than in the United States.” Extrapolated to the professional context, accountability based on Standards of proper conduct seems difficult. Also, in certain cultures, the motivation of people to perform is lower than in others. For example, an underlying assumption of Organizational Development is that the organizational members have the motivation ““to make [...] a higher level of contribution to the attainment of organizational goals [...]’ [French & Bell, 1990, p. 44] Probably this assumption does not hold true in many developing countries” (in: Berry et al., 1997, Vol. 3, p. 394).

One should also take into account current and future developments in Evaluation across the world. As more indigenous evaluators are being trained in western settings or by western evaluators, the Standards will probably become more widespread even in cultures where their use would not come about without such foreign influence. Additional dissemination might, of course, come from those western evaluators engaged in evaluations in foreign cultural settings.

Even though, in some cultures, evaluators do not and will not rely on Standards for *judging* professional performance, the Standards can still provide helpful *ideas* on important aspects of what good evaluations should look like. However, even these ideas need to be relevant to the specific cultural backgrounds and values, making an analysis of their cross-cultural transferability a sensible task.

#### A Values-Based Analysis of the Standards’ Cross-Cultural Transferability

For examining the Standards’ transferability to other cultures, it is important to analyze the Evaluation Standards regarding the North American values they reflect, as represented by the Joint Committee members. Values have a varying degree of generality across cultures. Like values, we consider some of the Standards as being universal. These Standards (e.g., Practical Procedures) need no further discussion from a cross-cultural point of view. Yet other Standards (e.g. Rights of Human Subjects) are unique to North American culture and lack transferability to other cultures.



What values or value dimensions are reflected in the Standards and how do these values differ across cultures? Which are valid Standards for evaluators in all cultures, and which Standards pertain to goals of conduct that do not apply in cultures different from the North American culture?

In examining the Standards, the complexity of the analysis was restricted by omitting to discuss the interdependencies among the Standards. By definition, focusing on one Standard will sometimes mean neglecting another (*trade-off*, see Stufflebeam, 1986, pp. 10; Joint Committee, 1994, p. 9); some Standards form an entity in the evaluation process. In addition, as mentioned above, the Joint Committee (1994) stresses that the Standards are not rigid principles. The analysis presented in this paper can neither account for all possible situations and cases in which the Standards are used, nor the different ways in which they are applied by different evaluators. Their negotiability ameliorates their transferability to other cultural contexts. However, their negotiability does not compensate for their reflection of values specific to their origin. In addition, an analysis dealing with cultural values is necessarily oriented to 'mainstream' values. Societal values are always arranged along a continuum, some people being closer to one end, some closer to the other; but the majority will be found in the middle. The interpretation and use of the Standards also occurs along a continuum which is affected by the evaluator's expertise and values and the current context of the evaluation (within a culture). But some degree of simplification is necessary to ensure the clarity of the analysis. We concede that the following analysis is based on subjective interpretation of the Standards as they appear in the second edition, with all the above mentioned restrictions.

A final remark is dedicated to the 'Guidelines' and 'Common Errors' sections in the Standards (Joint Committee, 1994). We are aware that they were designed as procedural suggestions based on the experiences of North American evaluators – they are not intended as general principles applicable in any situation (even less than the Standards themselves). Nevertheless, some of the observations made when reading these sections through the 'cross-cultural lens' are included since these remarks might help the reader in understanding the value dimensions that were identified and might also heighten the reader's awareness regarding cultural influences during the evaluation process.

### Utility Standards

|  |
|--|
| <p><b>U1 Stakeholder Identification</b>      Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.</p> |
|--|

The 'Stakeholder' concept is of central importance throughout the Evaluation Standards (it is part of almost all Standards), and it is most explicit in the Utility Standard **U1, Stakeholder Identification**. This concept is closely tied to the *Hierarchy-Egalitarianism* cultural dimension. For example, the inclusion of all significant stakeholders in the evaluation process is a desirable and legitimate goal in Western democratic societies, like the United States, that have an egalitarian understanding of society. In other cultures, people view the unequal distribution of

power, roles and resources as desirable and legitimate. People are socialized to comply with their roles and the obligations related to each role. Society often penalizes people for failing to comply with the obligations imparted by their social status (Berry et al., 1997, Vol. 3, p. 79). Thus, paradoxically, what the Standards list as *Common Errors* may serve as *Guidelines* in high hierarchy cultures, e.g., when stating: "Assuming that persons in leadership or decision-making roles are the only, or most important, stakeholders." (Joint Committee, 1994, p. 26. For example, for South Korea, Jang (2000, p. 10) points out that "the project focus is likely to be determined by the specific client group that supports and controls the project." Throughout her analysis of the applicability of the Standards in South Korean society, Jang stresses the importance of the focus on the top client group (see Jang, 2000, pp. 10) and she concludes that "evaluators [...] need to meet only the immediate stakeholder's expectations" (p. 24).

|   |
|---|
| <p><b>U2 Evaluator Credibility</b>                      The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.</p> |
|---|

The second Utility Standard *U2, Evaluator Credibility*, also touches upon culturally relevant value dimensions. As pointed out before by Smith et al. (1994, p. 7), the concept of a *professional* differs across cultures (e.g., between Malta, India, and the U.S.). They state that "in the U.S., a professional is someone who [...] has acquired a certain expertise [which] is officially acknowledged through certification and licensure [...]," whereas in Malta, "the notion of professional is much looser [...]." In India, "'Professionalism' is not a common word and to the orthodox or older generation it connotes a lack of love for the job [...]. Professionalism has nothing to do with the level of one's expertise" (p. 7). Consequently, the notion of 'competence' as specified in the Program Evaluation Standards (see Joint Committee, 1994, p. 31) is not a prerequisite for an evaluator's credibility across all cultures. The characteristics that cause clients to judge the professional evaluator as *competent* and *trustworthy* differ across cultures. For example, cross-cultural organizational psychology uses the principle of *seniority* to describe differences in authority held by members of an organization (Berry et al., 1997, Vol. 3, p. 379). According to this principle, older people are per se attributed a higher degree of competence in certain cultures, regardless of other factors also influencing competence, e.g., level of education (see Jang 2000, p. 26). The social status of a person (often positively correlated with age) also affects the attribution of competence and trustworthiness in certain cultures, contrary to a more achievement-oriented, individualistic approach to professional worthiness. The latter is conveyed in the Standards when stating that "evaluators are credible to the extent that they exhibit the training, technical competence, substantive knowledge, experience, integrity, public relations skills, and other characteristics considered necessary by clients [...]" (Joint Committee, 1994, p. 31). For example, Jang (2000, p. 11) notes that the well-known "name" of the evaluator is the basis for being considered competent by South Korean clients. As another example, an Indian evaluator (see Hendricks & Conner, 1995, p. 83) describes his culture as one in which "personalized relationships and connections matter more than public interest," so that "a nexus

develops where evaluation is a means to bestow favors to each other irrespective of their levels of competence.”

|  |
|--|
| <p><b>U3 Information Scope and Selection</b>      Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interest of clients and other specified stakeholders.</p> |
|--|

Standard *U3, Information Scope and Selection*, mentions multiple stakeholders, "each of whom should have an opportunity to provide input to the evaluation design and have a claim for access to the results" (Joint Committee, 1994, p. 37). As pointed out before, compliance with this Standard is not necessarily desirable in a hierarchically structured society. The client and the evaluator might not find it important to extend involvement to other than the single most influential stakeholder group because other groups may not or should not influence the evaluation process and results (see Jang, 2000, p. 10). In addition, group-orientation and loyalty toward authorities may forbid certain stakeholder groups to voice their opinions, contrary to the democratic approach generally evident in the Standards (Joint Committee, 1994). Also, as Jang observes, "Koreans [...] are not as experienced as Americans in expressing their personal opinions" (2000, p. 18). The author attributes this phenomenon to "the cultural norm of indirection" (p. 18). In addition, there is a high tendency to perceive group consensus (*false group consensus effect*) in collectivist cultures which further limits the information scope of the evaluation in these cultures (see Berry et al., 1997, Vol. 3, chapter 1, p. 58; Jang, 2000, p. 18, for an example from South Korea). Contrary to individualistic societies (e.g., Australia) where ingroup members are expected to act as individuals even if it means going against the ingroup, in collectivist cultures (e.g., Indonesia, see Noesjirwan, 1978) the members of the ingroup should adapt to the group, ameliorating its unity (see Berry et al., 1997, Vol. 3, p. 143).

|  |
|--|
| <p><b>U4 Values Identification</b>      The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.</p> |
|--|

Standard *U4, Values Identification*, is an example of a culturally sensitive Standard. It calls for the acknowledgment and description of contextually different value bases for interpretation of obtained information. "Value is the root term in evaluation; and valuing [...] is the fundamental task in evaluations" (Joint Committee, 1994, p. 43). Values differ across individuals as well as across cultures (see above); therefore "it seems doubtful that any particular prescription for arriving at value judgments is consistently the best in all evaluative contexts" (Joint Committee, 1994, p. 43).

**U5 Report Clarity** Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

The fifth Utility Standard *U5, Report Clarity*, generally seems culturally universal but the concept of *clarity* touches upon a cross-culturally relevant communication style issue. The Standard states that “[...] clarity refers to explicit and unencumbered narrative, illustrations, and descriptions” (Joint Committee, 1994, p. 49). This might be contraindicated in some cultures where indirect communication is the socially accepted way to exchange ideas and to make statements. Barnlund (1975) contends that “the greater the cultural homogeneity [collectivism], the greater the meaning conveyed in a single word, the more that can be implied rather than stated” (in: Berry et al., 1997, Vol. 3, p. 141). Jang (2000, p. 22) notes that in South Korea, public argumentation is “actively shunned because engaging in it means to stand out, risk public disagreement, and lose favor [...].” In addition, context information will have a different impact on the interpretations of given information. In collectivist cultures, the greater interdependence orients people toward greater sensitivity to context information and more meaning is drawn from it. In contrast, individualistic cultures, which are regarded as *low-context* cultures, place a greater emphasis on dispositional attributions (Berry et al., 1997, Vol. 3, chapter 1, p. 26).

**U6 Report Timeliness and Dissemination** Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

Standard *U6, Report Timeliness and Dissemination*, calls for report dissemination to “all intended users” (Joint Committee, 1994, p. 53). Reaching all stakeholders is not a goal if the client expects absolute authority over the evaluation and sole reception of evaluation findings. Jang (2000, p. 24f) argues that in South Korea, the responsibility of the evaluator is limited to “reporting to the top client” and that the top client decides about “the extent to which evaluation results are shared.” Consequently, the practice of “failing to recognize stakeholders who do not have spokespersons” (Joint Committee, 1994, p. 55, labeled as a *Common Error*) might be acceptable conduct in nonegalitarian cultures. Another culturally varying concept is that of timeliness. It is not a universal value and might therefore be excluded as part of proper conduct for evaluators in certain cultures. For example, “Africans have different concept of time than Western audiences” (Russon & Patel, 1999, p. 6; also see Jang, 2000, p. 25; Smith et al., 1994, p. 11).

|                             |   |
|-----------------------------|---|
| <b>U7 Evaluation Impact</b> | Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased. |
|-----------------------------|---|

Regarding Standard *U7, Evaluation Impact*, it is important to stress that the evaluator should encourage follow-through by stakeholders/clients in any culture—as long as hierarchical and collectivist constraints do not prevent the evaluator from playing such a role. For example, Jang (2000, p. 25) states that evaluators in South Korea should not “take responsibility for [enhancing] the use of the evaluation results”; this is solely the clients’ business. Lewy (1984, p. 377) reports about Israel that “utilization of study results are mediated through an academic filter operated by the chief scientist” who serves as a mediator between academic community and educational administration. Therefore, Israeli evaluators will not be active in promoting follow-through to the degree demanded by Standard *U7*.

### Feasibility Standards

|                                |  |
|--------------------------------|--|
| <b>F1 Practical Procedures</b> | The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained. |
|--------------------------------|--|

|                               |   |
|-------------------------------|---|
| <b>F2 Political Viability</b> | The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted. |
|-------------------------------|---|

|                              |  |
|------------------------------|--|
| <b>F3 Cost Effectiveness</b> | The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified. |
|------------------------------|--|

In general, the Feasibility Standards seem more culturally universal than the Utility Standards. For example, Standards *F1, Practical Procedures*, and *F3, Cost Effectiveness*, seem culturally universal. However, *F2, Political Viability*, displays an egalitarian approach to political viability when it demands “[...] fair and equitable acknowledgment of the pressures and actions applied by various interest groups with a stake in the evaluation” (Joint Committee, 1994, p. 71). See *U1* for further elaboration on the Stakeholder concept. Contrary to the Standards’ recommendation to avoid “giving the appearance -by attending to one stakeholder group more than another- that the evaluation is biased in favor of one group” (Joint Committee,

1994, p. 72), in hierarchical cultures it might be crucial for a successful evaluation to attend to one (the most powerful, not necessarily the most important) stakeholder group more than another. Generally, in more conservative (as opposed to autonomous) cultures, it is desirable to refrain from actions that might disrupt the solidarity of the group or traditional order (see Schwartz 1994a, in: Berry et al., 1997, Vol. 3, p. 99). In these cultures, political influences are likely to be stronger than in more autonomous societies like the United States where people seek to express their opinions (see Jang, 2000, p. 10, for South Korean example). In India, for example, according to Smith et al. (1994, p. 3), “the quality of one’s evaluation work is more likely to be judged on political than on technical or professional grounds.”

### Propriety Standards

Miller and colleagues (1990, in: Berry et al., 1997, Vol. 3, p. 24) argue that “[...] conceptions of morality vary across cultures: A justice-based notion of morality that emphasizes autonomy of the individual and individual rights [...] is found predominantly in individualistic cultural contexts.” This notion of morality (or propriety) relies upon impartial rules and standards that are equally applicable to everybody. In contrast, in collectivist cultures “[...] morality constitutes a social practice that should be understood in terms of a duty-based interpersonal code.” Lewy (1984, p. 379) also points out the contextual dependency of ethical categories. Stufflebeam (1986, p. 4) notes that the Propriety Standards are “particularly American.” Russon & Patel (1999, p. 8) quote their African focus group participants consenting that the Propriety Standards portray “the greatest challenges” for transferability.

There is also a close connection between the Propriety Standards and the *Hierarchy-Egalitarianism* cultural dimension; egalitarian cultures stress the moral equality of individuals (Schwartz 1994a, in: Berry et al., 1997, Vol. 3, p. 63) while hierarchical societies emphasize the legitimacy of inequality of individuals and groups. Jang (2000, p. 24) notes that in collectivist/hierarchical societies, ethical problems (as seen through Western eyes) might be “rationalized as being in the group’s best interest.” In addition, there are high-context cultures which stress situational (context) information when judging behavior. In contrast, low-context cultures, like the United States, refer more often to dispositional information, based on the assumption that traits have generality over time and situations (see Berry et al., 1997, Vol. 3, p. 63). For all the reasons mentioned above, it generally seems difficult to transfer the Propriety Standards to collectivist/hierarchical/high-context cultures.

#### **P1 Service Orientation**

Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

Specifically, Standard *P1, Service Orientation*, calls for serving “program participants, community, and society” (Joint Committee, 1994, p. 83). In societies where there are large differences between individual/group status and goals (and where these differences are considered legitimate), this range of service delivery seems to be out of reach for the evaluator. Consumer rights, mirroring service orientation and reflected by legislation, also vary across

cultures. North America is especially known all over the world for its service orientation. However, Standard P1 can be considered a universal goal for evaluators, even if not common in many cultures yet.

|                                    |  |
|------------------------------------|--|
| <p><b>P2 Formal Agreements</b></p> | <p>Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.</p> |
|------------------------------------|--|

Smith et al. (1994, p. 10) point out the cultural differences in the way agreements are handled. Formal agreements, especially as elaborated by Standard *P2, Formal Agreements*, are not common practice in many cultures. For example, in India the power of formal agreements is more limited than in the United States (see Smith et al., 1994, p. 10). Likewise, Jang (2000, p. 18) observes that, for South Korea, memorandums and contracts in business settings are less common than in Western cultures. A contract might actually deteriorate the client-evaluator relationship because the client could perceive it as a sign of distrust toward him. This practice might also be connected to communication style issues, i.e., the preference for indirect communication might coincide with a preference for informal agreements.

|   |  |
|---|--|
| <p><b>P3 Rights of Human Subjects</b></p> | <p>Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.</p> |
|---|--|

Standard *P3, Rights of Human Subjects*, is clearly based on U.S.-American legislation. As Chapman & Boothroyd (1988, p. 41) point out, American human rights laws “would never permit many practices in the United States that are common in evaluation studies in developing countries.” Smith et al. (1994, p. 11) also mention cross-culturally different ideas of what harms individual rights. For example, in Malta and India, reputation and mental health “are not possessed by the individual in the U.S. sense.” Differences in human rights legislation are also noted by Hendricks & Conner (1995, p. 85). African evaluators noted in a discussion of this Standard’s transferability that, contrary to many Western cultures, in Africa “the rights of the individual are perhaps more often considered subordinate to the greater good of the group” (p. 9).

|                                     |   |
|-------------------------------------|---|
| <p><b>P4 Human Interactions</b></p> | <p>Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.</p> |
|-------------------------------------|---|

*Standard P4, Human Interactions*, states cross-culturally relevant goals for evaluators although the actual professional behavior might differ considerably between different cultural settings, due to varying desirable interpersonal interaction patterns (for examples toward elders and high status persons, see Smith et al., 1994, p. 11).

**P5 Complete and Fair Assessment** The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

*Standard P5, Complete and Fair Assessment*, examining a program to find both strengths and weaknesses, seems a goal for evaluators everywhere, despite multiple factors that might get in the way of this Standard, for example, pressures exerted by powerful stakeholders to stress either strengths or weaknesses. These pressures could be said to have more impact on evaluators in hierarchical and collectivist societies (also see analysis of Standard P7, 'Conflict of Interest').

**P6 Disclosure of Findings** The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

*Standard P6, Disclosure of Findings*, displays the *right to know* which is legally granted to United States citizens as part of the *Bill of Rights*. As an example of how values provide ultimate standards, Peoples & Bailey (1997, p. 28) mention "the American emphasis on certain rights of individuals as embodied in the Bill of Rights in the Constitution [...] Freedom of the press, of speech, and of religion are ultimate standards [...]." In cultures other than the United States, these issues do not play such a vital role in an individual's rights (see Beywl, 1996). For example, Hendricks & Conner (1995, p. 85) point out that Australia has a Freedom of Information Act but "it would be difficult (even under this act) to allow all relevant stakeholders to have access to evaluative information, and ...actively disseminate that information to stakeholders." In addition, the "code of directness, openness, and completeness" (Joint Committee, 1994, p. 109) might not be in accordance with the preference (and superiority) of indirect communication and the existence of hierarchical structures in certain cultures. Jang (2000, p. 15) states that in South Korea, "the task of 'fully' informing research subjects who are unfamiliar with the purposes and procedures of the study is not the first concern." Likewise, absence of such disclosure might not, as claimed in the Standards (p. 109), threaten the evaluation's credibility. To the contrary, the absence of disclosure may preserve the evaluation's credibility in some cultures. Smith et al. (1993, p. 10) offer an example of the acceptable absence of disclosure when they refer to both Maltese and Indian cultures. For these cultures, the possibility of "placing someone in jeopardy" (e.g., by negative evaluation findings) is never communicated in writing, thus limiting the disclosure of findings.



**P7 Conflict of Interest**

Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

Standard *P7, Conflict of Interest*, is another Standard reflecting North American values. There are different cultural standards for dealing with conflicts. Group-orientation in collectivist cultures results in conflict avoidance or non-adversarial conflict resolution within groups (see Berry et al., 1997, Vol. 3, chapter 1, p. 26). In those cultures, it might be counterproductive to openly address an existing conflict. In some cultures, direct communication is generally avoided (see Jang, 2000, p. 18). Another cross-culturally relevant problem with this Standard is the higher interdependency of people in collectivist societies. Personal friendships or professional relationships play a much more crucial role in these cultures than in individualistic societies such as the U.S. As Jang (2000, p. 12) describes South Korean culture, “professionalism is defined mainly by an individual’s loyalty toward the group he or she serves.” In addition to the priority of group interest, Jang mentions the great impact of authority figures on opinion formation (p. 22). According to Smith et al. (1993, p. 10), Indian evaluators are “specifically chosen for a point of view consistent with the official position” and this would *not* be considered a conflict of interest situation there. Standard P7 can therefore be considered unsuitable in certain cultural contexts.

**P8 Fiscal Responsibility**

The evaluator’s allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Finally, Standard *P8, Fiscal Responsibility*, is seen as cross-culturally applicable.

Accuracy Standards

Across cultures, different degrees of importance are allocated to accuracy as part of an evaluation (see Nevo, 1984, p. 385 on the high priority of accuracy in Israeli evaluations). Data-based judgements might be less credible than personal opinions from authorities in certain high-hierarchy cultures (e.g., not meeting the notion of accuracy that the Standards display). The concept of ‘Uncertainty Avoidance’ might provide some cross-cultural psychological explanation for differences in importance of reliability. For example, Hofstede (1980) talks about Uncertainty Avoidance in terms of how comfortable people are with uncertainty or ambiguity in their lives, or else, how much they value security and reliability. Jang (2000, p. 16) notes that although South Koreans “believe western systematic methodology is very useful in analyzing information” (she also mentions the considerable U.S. influence within the academic community in South Korea), it is necessary to involve South Korean staff to ensure receiving reliable, valid data as the Standards call for. This quote implies that there is a “western” methodology that invariably needs adaptation when applied in non-western contexts (see Jang, 2000, for advice on South Korea).

In this context, research as a *communicative process* becomes highly relevant. Evaluators need to pay attention to cultural differences in the nature of knowledge (as a group rather than an individual notion in collectivist societies), the right to opinions (only group leaders and elders as being legitimate opinion holders in collectivist societies), the expectations toward the construction of interviews and surveys, ingroup versus outgroup dichotomies, sex and age factors (see Berry et al., 1997, Vol. 1, chapter 8, for further elaboration).

It is beyond the scope of this paper to deal with issues of cross-cultural applicability of research methodology in more depth. Regardless of the *difficulties* evaluators face when trying to comply with the Accuracy Standards, they seem to mostly contain *goals* for evaluators acceptable across cultures. However, further investigation needs to address (1) what is considered accurate information in different cultures (i.e., what is considered ‘valid,’ ‘reliable,’ and ‘systematic’) and (2) how emphasis on either qualitative or quantitative methodological approaches relates to cultural values.

**A2 Context Analysis** The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

Besides these general, admittedly sometimes speculative concerns, the Accuracy Standards as goals for proper conduct seem to be cross-culturally transferable. Nevertheless, a few relevant comments will be shared. Standard *A2, Context Analysis*, creates problems in those cultural settings where some context information, e.g., on the political and organizational context of the program being evaluated, is hard to obtain and to share because of hierarchical concerns and strong group ties. On the other hand, high-context cultures which place emphasis on situational (contextual) information for attributional and judgmental processes will especially value Context Analysis as part of an evaluation (Berry et al., 1997, Vol. 3, chapter 3, p. 53).

**A11 Impartial Reporting** Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

In addition, Standard *A11, Impartial Reporting*, contains non-universal cultural values; the analyses given for Standards U5, P6 and P7 apply to this Standard.

Overview of a Values-Based Analysis of the Program Evaluation Standards

| Cultural Value Dimensions |                                       |                                     |                                      |
|---------------------------|---------------------------------------|-------------------------------------|--------------------------------------|
|                           | <i>Individualism vs. Collectivism</i> | <i>Hierarchy vs. Egalitarianism</i> | <i>Other dimensions and concepts</i> |
| Standard U1               |                                       | X                                   |                                      |
| Standard U2               |                                       |                                     | X                                    |
| Standard U3               | X                                     | X                                   |                                      |
| Standard U5               | X                                     |                                     | X                                    |
| Standard U6               |                                       | X                                   | X                                    |
| Standard U7               | X                                     | X                                   |                                      |
| Standard F2               |                                       | X                                   | X                                    |
| Standard P2               |                                       |                                     | X                                    |
| Standard P3               | X                                     |                                     |                                      |
| Standard P6               | X                                     | X                                   | X                                    |
| Standard P7               | X                                     | X                                   | X                                    |
| Standard A11              | X                                     | X                                   | X                                    |

## Conclusion

Generally, it seems that what is useful and ethical differs to a greater extent across cultures than what is feasible and accurate. In accordance with Jang's (2000, p. 28) and Smith et al.'s (1994, p. 12) findings, the author found Utility and Propriety Standards to be much more cross-culturally problematic than Feasibility (except F2) and Accuracy (except A11) Standards. However, a more in-depth analysis of the cross-cultural understanding of the concepts contained in the Accuracy Standards (e.g., validity, reliability) seems necessary to make a more reliable judgement concerning this dimension of the Standards. Marklund (1984, p. 357) also identifies the Utility Standards as being non-transferable. Contrary to our findings, he regards the Propriety Standards (along with the Accuracy Standards) as being "quite okay," referring to them as "concerning technical and not political demands for evaluation." This statement is not consistent with the Joint Committee's definition of Propriety Standards. According to the Committee, the Propriety Standards are intended to ensure "that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved [...]" (Joint Committee, 1994, p. 81). Thus, it becomes clear that Propriety Standard issues are highly dependent on both political and cultural influences. Russon & Patel (1999) arrive at the same conclusion in their account of a focus group discussion with African evaluators (p. 8). Nevertheless, they resolve that their participants considered most of the Standards to be "relevant, useful and applicable" in Africa (p. 12); based on the discussion findings, twelve Standards were extended or modified for the African context. Beywl (1999, p. 255) concludes in his German translation of the Standards that these principles can be transferred to cultures comparable to the USA but not to societies with greater cultural distance to the USA (also see Smith et al., 1993, p. 12 for a similar conclusion).

Finally, some recommendations are presented for those foreign evaluators who see the use of a set of Standards as beneficial for their professional conduct (see Lewy 1984, p. 381) and do not benefit from an already existing adaptation of the Standards. First, evaluators from cultures foreign to the United States should describe their societies with regard to the cultural dimensions used in this article. In addition to their own as well as their colleagues' perceptions and consultation with cultural experts, non-US evaluators may use empirical findings by Hofstede (1980), Trompenaars (1993), Schwartz (1994a, b) and Smith et al. (1996) to guide their analysis. The preceding analysis along with the chart provided above may (1) heighten evaluators' sensitivity toward Standards transferability issues and (2) encourage non-US evaluators' efforts to critically and systematically analyze the Standards. As a result of the analysis, the Standards with poor 'goodness of fit' within a cultural context may have to be abandoned or adapted to fit the new context. Professional evaluation associations now existing in many countries can provide a forum for this challenging task (see SEVAL-Standards, Schweizerische Evaluations-Gesellschaft, 1999).

## References

- Berry, J.W., Segall, M.H. & Kagitçibasi, C. (ed.) (1997). *Handbook of Cross-Cultural Psychology*. Vol. 1 & 3. Needham Heights, NC: Allyn & Bacon.
- Beywl, W. (1999). *Handbuch der Evaluationsstandards* [Handbook of Evaluation Standards]. Opladen: Leske + Budrich.
- Beywl, W. (1996). Anerkannte Standards und Leitprinzipien der amerikanischen Evaluation. In Heiner, M. (Hrsg.), *Qualitätsentwicklung durch Evaluation* [Quality improvement through evaluation]. Freiburg im Breisgau: Lambertus.
- Beywl, W. (1999). Standards für Evaluation. Auf dem Weg zu Regelwerken in der deutschsprachigen Evaluation [Standards for evaluation. Toward regulations in German-speaking evaluation]. *DeGEval News*, 2(1), 1-4.
- Chapman, D.W. & Boothroyd, R.A. (1988). Evaluation dilemmas: conducting evaluation studies in developing countries. *Evaluation and Program Planning*, 11, 37-42.
- Gudykunst, W. B. & Ting-Toomey, S. (1988). *Culture and interpersonal communication*. Beverly Hills, CA: Sage.
- Gudykunst, W. B., Gao, G., Schmidt, K.L., Nishida, T., Bond, M.H., Leung, K., Wang, G. & Barraclough, R.A. (1992). The influence of individualism-collectivism on communication in ingroup and outgroup relationships. *Journal of Cross-Cultural Psychology*, 23, 196-213.
- Hendricks, M. & Conner, R.F. (1995). International Perspectives on the Guiding Principles. *New Directions for Program Evaluation*, 66, 77-90.
- Hofstede, G. (1980). *Culture's consequences. International differences in work-related values*. Beverly Hills, CA: Sage.
- Jang, S. (2000). *The appropriateness of the Joint Committee Standards in non-western settings: A case study of South Korea*. Manuscript submitted for publication.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*. Thousand Oaks, CA: Sage.
- Lewy, A. (1984). Evaluation Standards: Comments from Israel. *Evaluation and Program Planning*, 7, 377-381.
- Marklund, S. (1984). Applicability of Standards for Evaluations of Educational Programs, Projects, and Materials in an international setting. *Evaluation and Program Planning*, 7, 355-362.

Nevo, D. (1984). Applying the Evaluation Standards in a different social context. *Evaluation and Program Planning*, 7, 383-386.

Peoples, J. & Bailey, G. (1994). *Humanity*. Saint Paul: West Publishing Company.

Rodrigues de Oliveira, T. & Gomes-Elliot, L. (1983). *Evaluation Standards: Reorganization for the Brazilian Educational Reality*. Paper Presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Russon, C. & Patel, M. (1999). *Appropriateness of the Program Evaluation Standards for use in African cultures*. Manuscript submitted for publication.

Schwartz, S. H. (1990). Individualism-collectivism: Critique and proposed refinements. *Journal of Cross-Cultural Psychology*, 21, 139-157.

Schwartz, S. H. (1994a). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S-C. & G. Yoon (Eds.), *Individualism and collectivism: Theory, method and applications* (pp. 85-119). Thousand Oaks, CA: Sage.

Schwartz, S.H. (1994b). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, 50, 19-45.

Smith, N.L., Chircop, S. & Mukherjee, P. (1993). Considerations on the development of culturally relevant evaluation standards. *Studies in Educational Evaluation*, 19, 3-13.

Smith, P. B., Dugan, S. & Trompenaars, F. (1996). National cultures and the values of organizational employees: A dimensional analysis across 43 nations. *Journal of Cross-Cultural Psychology*, 27, 231-264.

Stufflebeam, D. L. (1986). *Standards of practice for evaluators*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Triandis, H.C. (1995). *Individualism & Collectivism*. Boulder, CO: Westview Press.

Trompenaars, F. (1993). *Riding the waves of culture*. London: Brealey.

Villa, A. (1998). *Estándares para la evaluación de programas* [Standards for the evaluation of programs]. Bilbao: Ediciones Mensajero.

Widmer, T. (1999). Kontext, Inhalt und Funktion der 'Standards für die Evaluation von Programmen'. In Mueller-Kohlenberg, H. (Ed.) (2000). *Evaluation und Qualitätsmanagement in der sozialen Arbeit* [Evaluation and quality management in social work]. Opladen: Leske + Budrich.

Widmer, T., Landert, C. & Bachmann, N. (1999). *Evaluations-Standards empfohlen von der Schweizerischen Evaluationsgesellschaft (SEVAL)* [Evaluation Standards recommended by the Swiss Evaluation Society]. März 1999. Genève: SEVAL.





## **Considerations on the Development of Culturally Relevant Evaluation Standards**

By Nick L. Smith, Saviour Chircop and Prachee Mukherjee

Reprinted from *Studies in Educational Evaluation*, Vol. 19(1), 1993, pp.3-13 with permission from Elsevier Science.

What standards should be invoked in establishing proper evaluation procedures when working in a different cultural context? Western evaluators working in cross-cultural settings have suggested that what constitutes appropriate evaluation behavior depends on the culture in which one is working. Smith (1991), for example, provides a detailed summary of cultural influences in evaluation, noting the different cultural norms about what constitutes incompetent, unethical, and criminal behavior.

For example, in the U.S., it is expected that one pay a small amount of extra money to someone providing a service if he or she is a waiter or cab driver, but not if he or she is a police officer or court clerk – the difference between a tip and a bribe. Other cultures have different standards ... Just what is ethical evaluator behavior and by whose standards do we judge? Will cross-cultural evaluators have the same problems as U.S. multi-national corporations who insist they have to bribe (tip?) Foreign bureaucrats as the culturally accepted way to conduct business, but are then subject to U.S. criminal prosecution? Conducting evaluations in culturally sensitive ways might cause serious problems for the evaluator upon returning home (Smith, 1991, p.17).

*The Standards for Evaluations of Educational Programs, Projects, and Materials* of the Joint Committee (1981) has been developed and widely distributed as a guide to the proper practice of educational evaluation in the U.S. To what extent are these standards applicable in non-U.S. or non-western settings for prescribing proper evaluation practice? U.S. evaluators Chapman and Boothroyd (1988) suggest that the combined centralized authority in ministry offices and the impact of U.S. AID influence

. . . opens doors to evaluative researchers that are usually unavailable in more industrialized societies where teachers are unionized and where student and family rights are given a high level of importance. Principles of privacy, the public's right to know, and the Hatch amendment would never permit many practices in the United States that are common in evaluation studies in developing countries (Chapman and Boothroyd 1988, p.41).

These practices might be seen as violations of the Joint Committee (1981) standards, or as culturally sensitive and adaptive evaluation procedures. Should an evaluator working in a cross-cultural setting conform to the mores of his or her own culture or to the mores of the culture in which the evaluation is being conducted?

In this article we present preliminary considerations on two questions related to these cross-cultural evaluation concerns: to what extent might the Joint Committee (1981) standards be used as a basis for developing comparable, culturally relevant standards for indigenous use, and in what ways might the Joint Committee (1981) standards be modified for use by U.S. evaluators in cross-cultural settings.

Initially we outlined a straightforward development process: examine the Joint Committee (1981) standards from a particular cultural perspective and (a) delete those standards that appear irrelevant or inappropriate, (b) modify any standards to conform with the new cultural norms, and (c) add any new standards that would be required, so that the final set would be a culturally relevant counterpart to the original U.S. standards. We selected two disparate cultures to use in contrast with the U.S. culture within which the Joint Committee (1981) standards were developed: the culture of a small, Mediterranean developing country—Malta, and the culture of a large, eastern, developed country—India. Two of the co-authors of this article are from Malta and India respectively, have studied and practiced evaluation in the U.S., and are intimately familiar with the Joint Committee (1981) standards. In order to understand better the nature of culturally relevant evaluation standards, then, we initially worked to develop versions of the standards that could be used by indigenous evaluators in Malta and India, and/or by U.S. evaluators working within the Maltese or Indian cultures.

We soon came to realize that this seemingly straightforward task actually involved a kind of “linguistic translation” strategy in which (a) a given U.S. standard had to be analyzed in terms of its underlying U.S. values, norms, laws, and context, (b) a translation to counterpart values, norms, laws, and context in the second culture had to be made, and (c) a “new” counterpart standard relevant to the second culture had to be envisioned based on these translated values, norms, etc. This strategy had the benefits of uncovering a deeper understanding of the Joint Committee (1981) standards as a reflection of American culture and liberal democracy, and of heightening our awareness of some of the cultural bases of differences in evaluation practice in cross-cultural settings.

The full development of independent collections of culturally relevant evaluation standards is, of course, a highly complicated task. Through our exploratory efforts using this “standards translation” process, however, we did achieve some important insights into the relationships among culture, standards, and evaluation practice. Before proceeding to the results of our analysis, it is important that the reader have some appreciation of the historical and evaluation contexts of the two non-U.S. cultures used in this study.

#### The Historical and Evaluation Contexts of Malta and India

Both Malta and India were once British colonies. Malta is a small country with a local identity which experienced a friendly transition to independence in 1964. India is a very large country with many sub-culture identities. Its independence was achieved in 1947 in a much more adversarial manner.

Malta is a small Mediterranean island (122 square miles) with a population of 350,000. It is one of the top countries within the category of “developing countries” as listed by the World Bank and International Monetary Fund. The Island has a history that dates back more than 5000 years, and has been dominated by ever group seeking control of the Mediterranean. The survival of its people has depended on their ability to operate simultaneously within European and North African cultural milieus. Due to its history of colonization, interpersonal relationships are highly important, kinship and friendship ties are very strong in Malta. The political position of the person one is speaking with is of considerable concern, and communication is most rapid along “friend-of-friends” patterns. To be effective, the evaluator must be able to make use of these friends-of-friends networks and to refrain from asking for information in a way that the informant could be identified.

Chircop (1987) describes the role of evaluation in Malta as follows:

Formalized evaluation in Malta has largely been instigated by foreign needs and demands. Within a colonial structure, the representatives of the powers that be were always required to send reports to their superiors. Evidence of these reports exists for 2000 years. The reports by the various commissions and governors sent to the British Crown are perhaps the most recent and preserved documents. When the Island became independent, such reports were conducted by foreigners for the purpose of aid agreements with national or international institutions. Evaluations have also been conducted by foreign industrial concerns. Thus one might say that formalized evaluation, though existent in Malta, is largely a foreign concern (Chircop, p.8).

With a population of over 700 million, India is the seventh largest country in the world, three-fifths the size of the U.S. It has a long history that dates back to 1500 B.C. and has been dominated by Islamic and British powers for more than 700 years. The Indian tradition is characterized by cultural heterogeneity and religious tolerance. As laid down in the Indian constitution, India is a secular democracy.

India’s political structure consists of a powerful central government; any state government which happens to be headed by an opposition party will have a most difficult time gaining approval for its projects. Although India’s constitution provide for a multi-party system, the Congress party has historically dominated all other parties lacking a national majority. In November 1989, the Congress party failed to win the national elections, only the second time since India’s independence in 1947.

The major problems of India are over-population, poverty, and unemployment. These problems have to be kept in mind in understanding what works and what does not work in India; what is acceptable and what is not. The “culture” of India changes and reacts to these problems, and it is not possible to understand India’s culture independently of the context these problems create.

As with Malta, evaluation in India is seen largely as an external concern.

All of the Major South Asian countries have established a central monitoring and evaluation agency (CMA) responsible for monitoring all foreign-aided (and largely nationally financed) projects. Although the mandate of most CMAs includes evaluation, much less attention is paid to evaluation . . . Project monitoring systems were established soon after each country gained independence, but the evaluation of the systems has been very controversial, and in each country . . . the present monitoring and evaluation system has been operating for less than three years (Sri Lanka since 1986, India since 1985, Bangladesh since 1984, Nepal since 1986, and Pakistan since 1987) (Bamberger, 1989, p.225).

Since 1985 in India, the Project Monitoring Division of the Ministry of Program Implementation has been responsible for monitoring all major projects, while the Planning Commission is charged with evaluating all social programs. From a study of public sector evaluation in India and other South Asian countries, Bamberger (1989) concludes that monitoring and evaluation are used much less in developing than in developed countries because they are perceived as being imposed by donor agencies and are intended to serve donor, not national, interests. Because of negative experiences with evaluation on national projects, “. . . there continues to be a widespread questioning of the practical utility of broad-based impact evaluations, and many project manager argue that evaluation should essentially be abandoned except as part of academically oriented research projects” (Bamberger, 1989, p.234).

#### Development of Indigenous Standards

Against the backdrop described above, we raised the questions of whether it would be useful and possible to construct sets of evaluation standards for use in internal, indigenous evaluations in Malta and India, standards perhaps modeled on the Joint Committee (1981) standards.

In spite of the dramatic cultural differences in Malta and India, they are likely to share a common position on the use of evaluation standards, a position dramatically different from that of the U.S. Evaluators both in Malta and India might be required to employ standards, for example, as a condition of international funding, and they might “adopt” the Joint Committee (1981) standards in response to such a requirement or as a result of evaluation training by westerners. However, the evaluators in these two countries would not necessarily develop or use anything similar to the Joint Committee (1981) standards if left free to do evaluations in their own way. We make this claim based on several observations about cultural differences with respect to the nature and type of evaluation practice, the nature of professionalism, the role of standards in regulating professional practice, what it means to do a good job, and the role of government and the force of law. We discuss each of these points briefly below.

The nature of evaluation practice is markedly different in the U.S. in contrast to Malta and India. In the U.S., the Joint Committee (1981) standards were developed by representatives of 12 professional groups involved in and concerned with the practice of evaluation. The standards were intended to guide the professional practice of a growing, local, state, and national enterprise in need of professional direction and control. In India, however, there are no independent evaluators. All evaluation is done by the government of its own programs for its own use. There have only been a handful of major educational evaluations in India since its independence. All these evaluations have been initiated and conducted by the Central Government. All-India Educational Surveys are conducted periodically and they collect data regarding gross enrollment, number of schools and their distribution, number of teachers, etc., for use by the Indian Planning Commission. As in Malta, most program evaluations are done for and typically by foreigners: local level evaluations are not done. In both India and Malta, therefore, evaluation is not widespread. There are very few indigenous evaluators, and there is, therefore, neither a need for indigenous evaluation standards nor anyone to develop or employ them.

One might suppose that if the need for evaluation were to increase in these two countries, there might develop a number of trained evaluators and perhaps a new social role for the professional evaluator. With an evaluation profession, then, might come the need for something like the Joint Committee (1981) standards to guide professional practice. But cultural differences run deeper, and the concept of a “professional” differs across the U.S., Malta, and India.

In the U.S., a professional is someone who through formal education and experience has experience has acquired a certain expertise, whose expertise is officially acknowledged through certification and licensure, and who practices as a member of a peer group which oversees proper professional conduct through the use of standards and codes of ethics.

The notion of professional in Malta is much looser than in the U.S. In Malta, a professional is one who knows what he or she is doing, is a “gentleman” in manners and behavior, exhibits a sense of responsibility and care, and does not attempt to trick others. Professionalism is also defined by what it is not: financial greed or personal interest. In India, the more orthodox or traditional organizations find the concept of “professionalism” distasteful. “Professionalism” is not a common word and to the orthodox or older generation it connotes a lack of love for the job, the image of a person who just works for the money and will leave as soon as the job is completed. Professionalism has nothing to do with the level of one’s expertise.

In private and multinational corporations in India, which have been growing in the recent past with more liberal government policies, management personnel who have studied western management techniques encourage the western “professional” mind-set. But in government jobs, being a “professional” is rarely rewarded. The Indian government generates a vast number of jobs for the educated unemployed who are without special skills or professional degrees. Unemployment and poverty are such overwhelming problems in India that government jobs are at a premium among the educated unemployed. Obtained such a job means security, since one very rarely loses a government job except as a result of criminal activity; it means a steady income; it means not being in competition with thousands of others who seek such jobs. The job

one is fortunate enough to hold may not match personal interests, but there are few other options. Under such conditions, it is very difficult for a system of independent, collegially-bound professionals to arise. Such a group would threaten the stability of the current social and economic structure. An individual who attempted to engage in western “professional” behavior would do so without a supporting social structure and at considerable personal risk.

Not only is the concept of professionalism very different in India than in the U.S., but so is the role of standards in defining proper conduct. In India, there are general standards or expectations of proper conduct across all professional areas, but the idea of different standards for each profession is generally absent (except for the military, law, and medicine who do have separate standards). Behavior is dictated by the mores of the culture, not by explicated standards related to specific jobs or professional roles. The principles that guide one in everyday life are simply applied to professional activities.

In the U.S., an evaluator might advance in his or her career through professional peer recognition of outstanding work, with the Joint Committee (1981) standards perhaps useful in judging that work. In Malta, being chosen by the government to do certain work might be a source of credibility. In being chosen for a government task, it might be assumed that one is the only person available, or that one is a person who inspires credibility and trust with political or professional counterparts. Thus credibility is more important than expertise. It might also be assumed that one is chosen due to having proven oneself on previous occasions. Proving oneself may be based more on political criteria than on academic or professional ones. In some instances, people who select work on the basis of principles rather than convenience are careful not to be even perceived as working for the powers that be. Other people may accept such jobs, but then make a visible effort to prove their fairness to the different political stakeholders. The primary focus is more a question of being impartial, and of being perceived as such, rather than a question of the quality or outcomes of the work in terms of academic standards.

The situation is very different in India where there is never an assumption that the one chosen is the only person available. Individuals with the right “connections” have a higher likelihood of being chosen—the current practice is to hire people “. . . by the principle of ‘who knows whom’” (Bahgwati, 1976, p.670). Even academic settings are highly political. Both faculty and students are often affiliated with political parties, especially in government-run institutions and using one’s political influence to advance oneself or to further one’s own interests is considered *de rigueur* unofficially. Access to higher officials is limited, junior officials or clerks answer queries and “forward” an individual’s case. Both information and access are limited unless one knows the “right” people. In India, many of the best jobs are in government service which does not reward independent, critical activities, and, without a support group of external professional peers as exists in the U.S., such behavior would clearly be foolhardy.

Therefore, in both Malta and India the quality of one’s evaluation work is more likely to be judged on political than on technical or professional grounds. Professional standards would play a much weaker role in regulating evaluation practice than in the U.S.

In all three countries (U.S., Malta, and India), evaluation has been primarily an initiative of the federal government. The role of government and law in Malta and India contrasts sharply with that in the U.S., however, especially with respect to allegiance, efficacy, and the force of the law. Understanding these differences helps further explain the probable limitations in the use of any set of formal standards to guide evaluation practice in Malta and India.

In the U.S., the government is “. . . of the people, by the people, and for the people.” It is, more or less, run by the people to further their own interests. It is expected to function effectively, which it does in many respects, and laws are to be obeyed. One may disagree with a law, but is expected to abide by it even while working to change it. Engaging in civil disobedience to make a political statement is considered a strong action in a country where laws are enforced as much by social expectation of compliance as by official sanctions.

India, and especially Malta, however, have long histories of invasion, occupation, and colonization, in which the government was run by foreigners for the benefit of foreign interests. At least traditionally, in the U.S., to “beat the system” by not paying all of one’s taxes or utility bills was seen as stealing from one’s neighbor. In Malta and India, however, such action has been a means of keeping from the “oppressor” what is rightly one’s own. It is not stealing when one reclaims one’s property. This latter rationale has also come to be employed more widely in the U.S. as the government and large corporations are increasingly seen as impersonal, exploitative, and interested in only their own profit. Rises in insurance fraud, tax cheating, and employee theft evidence such a public attitude. In Malta and India this point of view has a stronger, more pervasive influence due to historical conditions of external exploitation.

If one derives benefit from a foreign-dominated government, one might support it. If, however, it serves foreign interests, and is ineffective in meeting local needs, then it is unlikely to receive local support. There has been a long tradition in Malta and India of not supporting ineffective colonial governments which serve other interests. The laws created by a foreign occluding power for its own interest thus do not carry the same moral weight as laws created “. . . by the people and for the people.” After 4000 years of foreign occupation in Malta, it is now sometimes said, “if you are born Maltese, you are born with a bachelor’s degree in cheating and a master’s degree in cheating the government.”

Malta and India now have independence and elected governments, but much of their governmental structure is still based on the British system. Further, after years of seeing the government as the oppressor, modern leaders may continue to exploit citizens as they learned to do from the colonial governments, maintaining a system that still does not meet citizens needs or expectations. It is not surprising that citizen attitudes about government and the law may not have changed fundamentally since independence.

Thus professional evaluation standards are less likely to arise and to be useful in these two settings. In the U.S., power is vested in individuals and professional groups who then regulate themselves either through professional standards or a system of checks and balances. In Malta, power is vested largely in the state. Even if the university grants the degrees, it is the state which

grants warrants to practice to lawyers, architects, and teachers. Although the Maltese Medical Association is an exception—they fought a hard battle to keep the right to certify their own people—in most other instances it is the state that certifies practice and acts as a referee, coming onto the scene if needed to insure fairness and equity. In India, the use of professional evaluation standards would hurt vested interests and would likely soon be subverted into a mockery of the very ideals such standard are designed to uphold. For example, it is believed to be common practice in India for a manufacturer of a less than acceptable product to get that product certified by the Indian Standard Institute through payoffs to relevant officials. Given current Indian cultural norms, one might expect evaluation standards to be treated similarly.

In the U.S., a pluralistic culture of strong professional groups with a tradition of professional standards and a history of compliance with self-generated laws and guidelines, the development and use of professional evaluation standards is reasonable and likely to be useful. In other cultures, such as Malta and India, where a well-defined ethos determines behavior; where few individuals practice professional evaluation; where professional identity is not dominant, the use of standards is uncommon and explicit laws are historically the result of oppressive governments serving foreign interests; the development and use of indigenous professional evaluation standards are unlikely, unwelcome, and probably unuseful.

#### International Use of the Joint Committee (1981) Standards

Indigenous evaluation standards seem unlikely to be developed and used in such cultures as Malta and India as they are in the U.S. But what of the western evaluator who is conducting an evaluation in a country such as Malta or India: are there possible modifications to the Joint Committee (1981) standards that would guide one in conducting both a professionally adequate (by U.S. standards) and culturally sensitive evaluation? We discuss a few examples below to illustrate some of the ways in which individual standards would have to be modified to be more relevant to the cultures of Malta and India. Since there can be no universal, culturally-relevant standards, different modifications would have to be made for each separate culture.

#### C1 - Formal Obligation

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it (Joint Committee, 1981, p.63).

Formal agreements are comparatively rare in India. Litigation in civil courts is largely limited to property and tenant-landlord disputes. Matters like libel and alimony (divorce/separation is not common enough to be considered a problem) are not settled in court. Since litigation about contractual agreements is less common, the power of a formal agreement, which in the U.S. might then be enforced through the courts, is much less.



## C2 - Conflict of Interest

Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results (Joint Committee, 1981, p.63).

This standard would be less important in India, or at least would be construed very differently. Since evaluations are done at the federal level by individuals specifically chosen for a point of view consistent with the official position, it would be expected an accepted that the evaluator would have what, in a U.S. view, would be considered a conflict of interest. In India, the evaluator would not be considered to be in a conflict of interest situation. One would not particularly gain anything by employing an independent point of view. Since the political system is one of majority control, without the two party checks and balances of the U.S. system, independent viewpoints would be considered neither prudent nor necessarily laudable.

## C3 - Full and Frank Disclosure

Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation (Joint Committee, 1981, p.63).

In both Malta and India, one would simply never write anything that would jeopardize someone's job, except for political reasons, as, for example, if one were in a politically adversarial position. Otherwise, placing someone in jeopardy would be considered inappropriate behavior. Such information might be orally communicated, but never written. Thus a U.S. evaluator would need to modify the nature and content of written reports in ways likely to violate the original (U.S.) interpretation of this standard.

## C5 - Rights of Human Subjects

Evaluations should be designed and conducted, so that the rights and welfare of the human subjects are respected and protected (Joint Committee, 1981, p.64).

Although it is important to protect the rights of individuals in any culture, those rights differ from culture to culture. A U.S. evaluator would have to modify the application of this standard to fit the Maltese and Indian cultures. For example, what constitutes injury and redress in these cultures is different from the U.S. In the U.S., reputation and mental health are viewed as possessions or property of the individual, the damage of which can be translated into dollars so that redress can be made in financial terms. In Malta or India, litigation is common over land or personal property, but not over reputation or mental health, which are not possessed by the individual in the U.S. sense. Individuals in these two cultures do not take offense as readily and redress may be in more symbolic ways, such as public apology.

## C6 - Human Interactions

Evaluators should respect dignity and worth in their interactions with other persons associated with an evaluation (Joint Committee, 1981, p.64).

Special attention would need to be paid to this standard by U.S. evaluators in Malta and India where greater “respect” is accorded to elders and persons of high status than in the U.S. In those cultures, normal U.S. modes of interaction are likely to be seen as disrespectful.

Although these examples are all from the Propriety Standards, modifications would also be required of standards in the other categories. For example, the standard *B2 - Political Liability* takes on added meaning in cultures where the actual physical survival of the evaluator may depend on how his or her work is received. Also, the standard *A7 - Report Timeliness*, would have to be reinterpreted in the many places in the world which do not share the U.S. sense of time. In the U.S., it is the individual’s responsibility to be on time, to meet deadlines, etc. It is customary to apologize, for example, if one is late for an appointment. In Malta and India, however, time is not considered to be so rigid. It is assumed that one makes a reasonable effort to be on time and thus apologies for tardiness are neither expected nor sometimes even acknowledged. To insist on holding someone to an officially stated deadline is viewed as nit-picking and unreasonable.

The Joint Committee (1981) standards might, therefore, serve as a general starting point for guiding proper evaluation practice by U.S. evaluators in other cultures, but substantial modifications may be required in order to insure that culturally sensitive evaluation practice is accomplished. The existing Joint Committee standards are likely to be even less appropriate for indigenous evaluators.

## Conclusion

We have selected Malta and India as illustrative examples which have been useful in exploring general issues concerning the development of culturally relevant evaluation standards. Other lines of research support or observations about the difficulty of transferring the use of evaluation standards across cultural boundaries.

Kedia and Bhagat (1988) suggest that in the transfer of process or person-embodied technology (such as evaluation methods and standards in our case), from industrialized to developing nations, that societal culture will play a greater role than organizational or managerial structure in determining success of transfer. They discuss a number of cultural characteristics affecting technological transfer, including *individualism* (in which ties between individuals are loose and people are inner-directed), and *abstractive* (in which thinking is causal and much communication is through mass media) versus *associative* (in which thinking is highly contextual and much communication is face-to-face). The degree of match between cultures influences the ease of transfer. Since the U.S. is generally individualistic and abstractive, and India and Malta are both much more collectivist and associative, the transfer of evaluation

standards is likely to be more problematic between the U.S. and India or Malta, than between the U.S. and, say, Germany.

But as we have tried to point out, the use of the Joint Committee (1981) standards in non-U.S. settings is more than an issue of technology transfer. The concept of standards as employed in the U.S. is much less relevant within the Maltese and Indian traditional cultures. Although standards may be imposed (or “transferred”) from the outside, indigenous are unlikely to emerge.

We have raised several issues about the cultural relevance of the Joint Committee (1981) standards, indeed about any set of professional standards developed for use within a particular culture. We have suggested that in some cases they may be modified to guide U.S. evaluators in adhering to U.S. standards when conducting effective, culturally sensitive evaluations in non-U.S. settings. We have also argued that neither these standards, nor in fact any similar kinds of standards, are likely to either arise or be seen as useful in indigenous evaluation in cultures such as Malta and India. these observations lead us to caution U.S. evaluation instructors to more carefully discuss the limited application of the Joint Committee (1981) standards in international work, and to be especially careful in advocating the use of these standards to international students who will be conducting indigenous evaluations in their native countries. As Asamoah (1988) has pointed out, parochial, western views of evaluation have resulted in western-trained indigenous evaluators being more insensitive to local culture than even outsiders. We hope that the foregoing discussion will contribute to more thoughtful application of the Joint Committee (1981) standards in international contexts.

## References

- Asamoah, Y. (1988). Program evaluation in Africa. *Evaluation and Program Planning*, 11 (2), 169-177.
- Bamberger, M. (1989). The monitoring and evaluation of public sector programs in Asia. *Evaluation Review*, 13 (3), 223-242.
- Bhagwati, J. (1976). Education and income equality. In J. Bhagwati, *The Spirit of India*, volumes presented to Indira Gandhi by the Indira Gandhi Abhinandan Samiti-Bombay, Vol. 2. New York, NY: Asia Publishing House.
- Chapman, D.W., & Boothroyd, R.A. (1988). Evaluation dilemmas: Conducting evaluation studies in developing countries. *Evaluation and Program Planning*, 11 (1), 37-42.
- Chircop, S. (1987). *Evaluation approaches: Utility across cultures?* Paper presented at the American Evaluation Association annual meeting. Boston, MA. October, 1987.
- Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials*. New York, NY: McGraw-Hill.
- Kedia, B.L., & Bhagat, R.S. (1988). Cultural constraints on transfer of technology across nations: Implications for research in international and comparative management. *Academy of Management Review*, 13 (4), 559-571.
- Smith, N.L. (1991). The context of investigations in cross-cultural evaluations. *Studies in Educational Evaluation* 17 (1), 3-21.

## **The Appropriateness of Joint Committee Standards in Non-Western Settings : A Case Study of South Korea**

By Soojung Jang, Syracuse University

### Abstract

To examine the appropriateness of Joint Committee Standards in non-Western settings, a case study of South Korea was conducted. This study aimed to establish a framework of advice that can be applied to Korean evaluation projects and to guide Western evaluators who wish to develop culturally effective techniques. It presents descriptive statistics from questionnaire analysis on the Joint Committee Standards. Also, it examined a one-day workshop discussion regarding 13 Standards that may be appropriate or inappropriate in relation to Korean cultural factors. Many participants agreed that the Joint Committee Standards conflict with Korean values. The Feasibility and Accuracy standards are mostly acceptable. However, some of the Utility and Propriety standards—such as Service Orientation, Stakeholder Identification, Conflict of Interest, and Disclosure of Findings—are inappropriate.

### Introduction

In the past, western evaluators used their own standards to judge the quality and adequacy of Korean programs mostly in the economic situation. South Korea does not have her own specific nation-wide evaluation standards nor use Joint Committee Standards. Joint Committee Standards have rarely been used in Korea since the Korean notion of evaluation aims to make a summative judgment or compare her competitiveness with other countries in consideration of international standards that are often proposed by international organizations. As Smith, Chircop, and Mukherjee (1993) indicated in the case of India and Malta, Korea may be required to use standards as a condition of international funding or cooperation, but she would not use the Joint Committee Standards or anything similar, if allowed to do evaluation in her own way. Before entering into the project methods and results, I address general information about Korea which might be relevant in conducting evaluation projects.

### General Information about South Korea

#### Demographics, Language Education and Tradition

South Korea lies in Eastern Asia, the southern half of the Korean Peninsula bordering the Sea of Japan and the Yellow Sea. She is a small republic state with a population of 46,416,796 as of July 1998, composed almost entirely of a hundred percent homogeneous ethnic group. Their language is a unique blend of a distinct phonetic alphabet called "Hangul," traditional Chinese characters, a visual ideogram in a written mode, and English that is taught in junior and senior high schools. Koreans take great pride in the fact that their history dates back more than 5,000

years. Due to the scarcity of natural resources in Korea, the survival of its people has not been so easy. For this reason most Koreans are industrious and hardworking in order to survive in the competitive social environment. They pay much attention to the value of education and have great respect for highly educated people. Over the past few decades, Korean education has experienced an extraordinary economic growth and has served as the prime force of national progress. Surprisingly, the literacy rate in Korea is up to almost one hundred percent. Everybody wants to get a higher education. Kinship and college alumni ties are very strong and comprise a critical part of Korean social relationships. Like most East Asian countries, Korea has been influenced by both Confucianism and Buddhism, but somewhat more bound to Confucianism. Since Korea is a small country, a more manageable and homogeneous unit than sprawling China, it may have been more uniformly and fully permeated by Confucian ideas than China was. Seniority, one of the most important Korean traditions stemming from Confucianism, is still cherished, and there is a clear structure of authority in any kind of social group such as families, business corporations and schools. The greater respect is accorded to elders and authorities, like a father in the family, a top manager in a business corporation, or a principal in a school. Many people pay much attention to family values and prioritize group interests over individual rights. Many marriages are still arranged based on each part's agreement. Social honor and monetary wealth do not always go together. A person's reputation and mental health can not be redressed in financial terms.

### Economy and Politics

As one of the "Four Economic Dragons of East Asia," South Korea has achieved an incredible record of growth. Three decades ago its Gross Domestic Production (GDP) per capita was comparable with levels in the underdeveloped countries of Africa and Asia, but today its GDP per capita is eight times India's, 15 times North Korea's and already at the levels of lesser economies of the European Union. This success through the late 1980s was achieved by a system of close government control, and has continued through the first half of the 1990s. At the end of 1997, an international effort, spearheaded by the International Monetary Fund was underway to shore up reserves and stabilize the economy. Growth in 1998 was sharply cut. Long-term growth will depend on how successfully South Korea implements planned economic reforms that would bolster the financial sector, improve corporate management, and further open the economy to foreign participation. In the past, Korean industry mostly depended on agriculture, but now the farm portion of the total population is just 9.7% as of 1998 (Korean National Statistics Office, 1998). According to the Statistics Office of the Korean government, the Korean Informatization Index which is a standardized number to show how much and how broad information technology affects Korea for international comparison, is 742, or about one seventh of United States Index of 5350. But its rate of increase of 40.2% is higher than that of the United States' at 34% (Korean National Statistics Office, 1997).

Korea is a high technology society. The internet user population is about 28% of the whole population, and wireless and mobile communication coverage rate is 37.6% (Kim, 2000). She began to provide a free international phone call service to U.S. in 1999.

South Korea still has a strong central government. Even though it adopted an independent local administration system in 1994, critical decisions are still made by the President, a representative of the central government. A top-down decision making hierarchy is prevalent in many social organizations, which is one of vestiges of the Korean military regime. All healthy Korean men have the responsibility of military service for either a term of eighteen or twenty-six months. According to the United Nations in 1997, a reliable mailing and telephone system spreads to almost every household. The Human Development Index of South Korea is ranked 30th out of 174 countries.

### The Relationship of the College Entrance Exam to the Notion of Evaluation: Evaluation Anxiety

The nature of evaluation practice is markedly different in Korea when compared to the United States where a variety of evaluation projects are done by every sector of society: government, business corporations and all levels of education for dozens of purposes. Korean educational evaluation is centered on measurement and testing. The major function of evaluations in education is selection and placement. This notion affects the evaluation practice of the entire society.

The central government or business conglomerates mainly initiate critical evaluation projects seeking international experts. In many cases they deal with a kind of organizational change, in that they need western substantive knowledge and highly advanced techniques. Korean stakeholders are inclined to think that their social positions may be influenced by the project results, which are frequently shown a kind of organizational change resulting in structural replacement or expansion and cutbacks. Evaluating something itself means the judgment of value or merit of an object. But the notion of evaluation to Koreans goes beyond the technical meaning, that is, Koreans are concerned about the effects of large-scale projects conducted by western evaluators, and sometimes they feel intimidated simply by being evaluated. They do not like evaluation. They are inclined to think that they are being measured in a certain point, rather than being helped by an evaluation.

During the past forty years, Korean notion of evaluation is especially related to the fierce college entrance examination experience. Most Koreans go through a very painful kind of evaluation: nationwide college entrance exam after one's secondary education is finished. Academic achievements are evaluated through the exam's total score. Since Koreans believe that the social level of a college diploma determines much of his future socioeconomic status, they make every effort to enter prestigious colleges. Since so many people want to go to college, the admission rate of 36% in 1998 is fairly low when compared to the number of applicants. It causes a phenomenon known as 'exam hell.' College entrance, as determined by the exam scores, have influenced Korean society at large in the formation of eliticism, strong alumni units, or limited information scope, even though criticism of the excessive competitiveness is continuously brought to the attention of the entire nation.

This test-based selection procedure, which influences what kind of social position a person will have in the future, causes people to be overly concerned. Therefore, many people have come to have negative feelings towards being evaluated. They are frequently intimidated by the examination in particular, and by being evaluated in general. Aside from college entrance, high-level civil servants, who are believed to hold one of the most honorable jobs in Korea, are selected through very competitive, open examinations. So being successful in taking examinations is one way of getting a high social position in Korea. Many Koreans recall "testing" when they hear of evaluation. Accordingly, the exam anxieties are serious.

Evaluation, especially in the field of education, has significantly concentrated on the measurement of academic achievement. It is unlikely that "exam anxiety" will be relieved any time soon, even though educational reform is taking place at a systemic level. For example, college recruitment without testing, shifting primary and secondary school focus to pupil orientation, and self-leadership education that encourages creativity and character development (Lee, 1998). At the same time, there has been a growing reflection on the fact that education to date has concentrated on the accumulation of information, that is to say, acquiring technical knowledge while ignoring the nurturing of a sense of intellectual responsibility and development. Educational evaluation has centered on outcome based testing, not on the learning process. Its focus has been a summative testing, not a formative one. There have been very few educational evaluations that consider the multiple dimensions of persons and social programs in a systematic way beyond the simple evaluation of a paper and pencil exam. There exists neither indigenous evaluation standards nor persons to develop or employ them.

### Method

In October 1998, a workshop was conducted at Syracuse University in Syracuse, New York, to examine the appropriateness of Joint Committee Standards for a Korean setting. Participants involved in the workshop were 21 graduate students at Syracuse University who have ever been involved in an evaluation project at government, business or education settings. Eleven participants were high level Korean civil servants, and seven participants were corporate managers. The others were educators, soldiers, etc.

The workshop began with one-hour lecture by Soojung Jang about the Joint Committee Standards, 2nd edition. She provided all participants with a detailed description on the Joint Committee Standards. After the lecture, she distributed 5-likert scale questionnaires, and asked participants to fill out the questionnaires regarding the appropriateness of each standard for a Korean setting. Smith's (1991) cultural model was circulated to stimulate their ideas. They were also asked to take notes on the reason certain standards were appropriate or not in the open ended section of the questionnaires. After the individual judgment about each standard was completed, participants and Soojung Jang discussed what standards are appropriate or inappropriate, and why. She played the role of facilitator throughout the workshop. Some people may have changed their mark on the questionnaire during the discussion, but the correction was not encouraged. All workshops were recorded for the author's use and all audiotapes were



removed (It is a normal Korean custom to throw away any interview recording after the transcription).

Based on their working experience, participants expressed a broad range of views on several of the Standards discussed in this paper. During the discussion time, they often showed unanimous agreement or conflicting opinions. For example, almost all participants agreed with U2 Evaluator Credibility, but had different opinions about P8 Fiscal Responsibility depending on their fields (government, education, business, etc.). The participants were 21 graduate students, so rather than generalize the results, this paper focused more on what kind of cultural factors drive changes of the notion of Joint Committee Standards.

Based on the workshop discussion, I describe unique Korean factors which can be involved in changing evaluation designs, along with the typical evaluation procedures of deciding focus, collecting, analyzing, interpreting, reporting and managing information. The structure of this paper follows the typical steps of an evaluation project, and indicates what would be appropriate and inappropriate standards in each step. That is, what standards can be adopted or adapted in Korean setting?

The results of the workshop were represented in two-fold. First, the author showed a descriptive statistics from the questionnaire surveys. Then, the remainder of this paper synthesized the workshop discussion notes. This paper did not examine the appropriateness of 30 Joint Committee Standards. The first concern of this paper is not in making a decision whether certain Standard is appropriate or inappropriate in Korean culture. Many workshop participants agreed that some Standards could be understood in totally different ways rather than they were inappropriate. Therefore, the author more focused on Korean cultural factors rather than on the original intent of the Standards. Thirteen Standards that might be highly related to Korean cultural factors were examined based on workshop discussion notes among 30 Standards.

The purpose of this paper is (1) to try to suggest what standards may be appropriate or inappropriate, (2) to explain why they may be appropriate or inappropriate in consideration of some Korean cultural factors, and (3) to provide some useful cultural guidelines for western evaluators who will conduct evaluation projects in Korea.

## Results

### Descriptive Statistics of Questionnaire Analysis

5-Likert scale (1 Not appropriate at all; 2 Somewhat inappropriate; 3 Don't know; 4 Somewhat appropriate; Very appropriate) questionnaires were used for descriptive statistics. The number of total participants was 21. Under each of five scale, the number of participants who chose a scale was presented, and mean score, standard deviation, minimum and maximum scale were shown in Table 1. Inappropriate Standards that have mean score of less than 2 (scale value: somewhat inappropriate) were highlighted.

Table 1. Descriptive Statistics of Questionnaire Analysis

|     |                                      | 1  | 2 | 3 | 4 | 5  | M    | SD   | Min | Max |
|-----|--------------------------------------|----|---|---|---|----|------|------|-----|-----|
| U1  | Stakeholder Identification           | 15 | 4 | 1 | 1 | 0  | 1.42 | 0.81 | 1   | 4   |
| U2  | Evaluator Credibility                | 0  | 0 | 0 | 0 | 21 | 5.00 | 0    | 5   | 5   |
| U3  | Information Scope and Selection      | 0  | 2 | 2 | 3 | 17 | 4.66 | 0.79 | 2   | 5   |
| U4  | Value Identification                 | 4  | 5 | 2 | 8 | 2  | 2.95 | 1.35 | 1   | 5   |
| U5  | Report Clarity                       | 0  | 1 | 0 | 2 | 18 | 4.85 | 0.35 | 4   | 5   |
| U6  | Report Timeliness and Dissemination  | 0  | 2 | 2 | 7 | 10 | 4.19 | 0.98 | 2   | 5   |
| U7  | Evaluation Impact                    | 0  | 0 | 0 | 1 | 20 | 4.95 | 0.21 | 4   | 5   |
| F1  | Practical Procedures                 | 0  | 2 | 2 | 1 | 16 | 4.47 | 1.03 | 2   | 5   |
| F2  | Political Viability                  | 0  | 1 | 1 | 4 | 15 | 4.57 | 0.81 | 2   | 5   |
| F3  | Cost Effectiveness                   | 3  | 4 | 3 | 8 | 3  | 3.19 | 1.32 | 1   | 5   |
| P1  | Service Orientation                  | 12 | 4 | 2 | 2 | 1  | 1.85 | 1.23 | 1   | 5   |
| P2  | Formal Agreements                    | 8  | 7 | 3 | 2 | 1  | 2.09 | 1.17 | 1   | 5   |
| P3  | Rights of Human Subjects             | 14 | 3 | 3 | 0 | 1  | 1.61 | 1.07 | 1   | 5   |
| P4  | Human Interaction                    | 9  | 5 | 6 | 0 | 1  | 2.00 | 1.09 | 1   | 5   |
| P5  | Complete and Fair Assessment         | 5  | 5 | 2 | 8 | 1  | 2.71 | 1.27 | 1   | 4   |
| P6  | Disclosure of Findings               | 19 | 2 | 0 | 0 | 0  | 1.09 | 0.30 | 1   | 2   |
| P7  | Conflict of Interest                 | 18 | 2 | 1 | 0 | 0  | 1.14 | 0.35 | 1   | 2   |
| P8  | Fiscal Responsibility                | 7  | 5 | 2 | 4 | 3  | 2.57 | 1.50 | 1   | 5   |
| A1  | Program Documentation                | 3  | 4 | 2 | 7 | 5  | 3.33 | 1.42 | 1   | 5   |
| A2  | Context Analysis                     | 0  | 0 | 1 | 7 | 10 | 4.57 | 0.59 | 3   | 5   |
| A3  | Described Purposes and Procedures    | 0  | 0 | 1 | 2 | 18 | 4.85 | 0.35 | 4   | 5   |
| A4  | Defensible Information Sources       | 15 | 4 | 1 | 1 | 0  | 1.38 | 0.66 | 1   | 3   |
| A5  | Valid Information                    | 0  | 1 | 1 | 3 | 16 | 4.66 | 0.65 | 3   | 5   |
| A6  | Reliable Information                 | 0  | 1 | 0 | 2 | 18 | 4.90 | 0.30 | 4   | 5   |
| A7  | Systematic Information               | 1  | 2 | 1 | 2 | 15 | 4.14 | 1.42 | 1   | 5   |
| A8  | Analysis of Quantitative Information | 1  | 1 | 3 | 0 | 16 | 4.38 | 1.20 | 1   | 5   |
| A9  | Analysis of Qualitative Information  | 4  | 2 | 1 | 7 | 7  | 4.09 | 0.88 | 2   | 5   |
| A10 | Justified Conclusions                | 0  | 1 | 3 | 7 | 10 | 4.28 | 0.78 | 3   | 5   |
| A11 | Impartial Reporting                  | 0  | 0 | 1 | 2 | 18 | 4.80 | 0.51 | 3   | 5   |
| A12 | Meta-evaluation                      | 17 | 2 | 1 | 0 | 0  | 1.19 | 0.51 | 1   | 3   |

P1 Service Orientation. Korean evaluation projects may not serve the needs of the full range of target population. In most international evaluation projects, the central government or business conglomerates oversee them. Since other independent research firms and their fund raising activities are not popular in Korea, evaluation projects are usually sponsored by the government or business corporations' budgeting offices. The range of the audience is not very broad, with top executives and internal staff members among the client group. Immediate stakeholders include the people whose work or business structure is being evaluated. In the case of government projects, evaluation results reach the whole nation. It is unique for Koreans to deal with stakeholder groups in conducting any evaluation project. The ideal is to involve potential stakeholder groups and strive to ascertain and accommodate their information needs. However, since the nation's or the group's interest is frequently more important, the individual's or small stakeholder group's interest may not be significantly considered. Sometimes the latter may be ignored completely. Only a small part of the stakeholder group, the immediate stakeholder, is considered in the evaluation design. Accordingly, the project focus is likely to be determined by the specific client group that supports and controls the project.

F2 Political Viability. Even though Wholey (1981) made it clear that evaluation findings can rarely change the power structure, but can promote something. However, in Korea, the utility of project results can be more than that. Sometimes, the results of a project might be used for political purposes which were not the intention at all in the beginning stages. Also, the successful completion of an evaluation project can be thwarted in a political system where definition is more a political exercise than a technical one, where radical ideas about solutions occur less often than minor variations in the status quo and where adopting successful solutions is dependent on practical ability and professional discretion as much as on efficacy. Therefore, "being a responsible evaluator" depends on excellent communication with the decision-makers. They are, in most cases, the top client group.

A2 Context Analysis. It is hard for foreign evaluators to understand the sociopolitical context of evaluation research and the political effects of an evaluation project. In a complicated sociopolitical context, western evaluators should be sincere with clients who call for an evaluation project. As time goes by, they may be getting more familiar with the situational context a certain evaluation project is involved in, as well as the political relationships among different stakeholders and clients while conducting a project with Korean staff members. However, they should first focus on who the top client is, and what their basic values are, because the project will be conducted under the top clients' direction. This means that the project focus is in the hands of the top clients. They will be the ones who support the project performed by western evaluators in terms of human and physical resources. For example, evaluators should ask Korean clients to indicate any cultural and technical pitfalls, expectations, and potential problems before they start to work. While applying their expertise, they should continuously ask clients for assistance. The internal staff members who work with western evaluators represent a top client group, while at the same time play roles as mediators between evaluators and clients, and reporters with western evaluators. They may or may not be immediate stakeholders.

U2 Evaluator Credibility. In selecting an evaluation consulting group, Korean clients pay much attention to the importance of the "name value" of the evaluation team. They may believe that the name of evaluation group or evaluator itself guarantees the rationality of the evaluation project. Usually Korean client groups will seek western evaluators if they need any substantive knowledge and highly advanced techniques for "big organizational change". Accordingly, they have high hopes regarding western evaluators' project capability. They do not mind research design or schedule, changes which western evaluators suggest unless it reverses their top clients' goals. That is, since the evaluators are brought in to introduce changes, it can not be avoided. It is believed that new techniques, new forms of activity, and investment in evaluation specialists will widen and strengthen Korean managerial and professional evaluation practices. What western evaluators have to do is to reach the right project focus in an effective way. However, the success of an evaluation project depends upon whether or not top client groups are satisfied with it. It means whether the top client should finally acknowledge the evaluators credibility. To complete this goal, communication between western evaluators and the Korean staff is crucial. An evaluator's professional competence is one of responsibility for what was assigned by their clients, not indifference towards the client's focus. Professionals are expected to do their best with limited resources. In case a western evaluator is successful at getting Korean clients' credibility, the evaluator will be involved in another future project of the Korean clients'.

P8 Fiscal Responsibility. In the Korean sense, professionalism is defined mainly by an individual's loyalty towards the group he or she serves, care towards other staff members, as well as in-depth knowledge and equalized skills. Western researchers may think critical issues such as allocation and expenditure of resources may arise when clients' perspectives and objectives are vastly different from those of the target community. However, western evaluators should keep in mind that top clients control project. They need to first focus on what the clients want to know, report it without bias, and make useful recommendations. Needless to say, they must show fairness in the way they conduct the project. Even though the environment may not be ready to support a good evaluation project, Korean clients will expect western evaluators to make good recommendations for client groups as much as resources will allow. This is in contrast with the way professional evaluators evaluate a project.

Weiss (1998), during a recent interview in the Harvard Family Project newsletter, indicated that evaluators should not undertake a study if the conditions for a good evaluation are not present. Evaluation takes time, resources and skills wherever it is conducted. She argues that evaluators should not take on studies when they know they can not do a good job. Evaluators do not have to be passive in accepting whatever conditions their clients set. They can negotiate by explaining that the time is too short, the requisite dates are unavailable, appropriate comparisons are missing, the money is insufficient for the size of the task, or whatever the problem may be. This is responsive evaluation in a real sense. However, it is very hard to be responsive in Korea in this way. Western evaluators get contracts and negotiate the terms of those projects. Korean clients, however, including strict government clients, will not always be willing to listen to evaluators' arguments. For example, a Korean client group may insist that the cost of the evaluation project should not exceed its benefit or effectiveness, or that the decent (not necessarily the best) recommendation can be made as resources permit. It means that however

bad the situation is, there is a way western evaluators can provide a valuable service for Korean clients through good communication.

### Collecting and Analyzing Information: Work Closely with Korean Staff Members

F1 Practice Procedures. To keep disruption to a minimum, western evaluators are required to closely work with Korean staff members. The involvement of members of the target culture in all aspects of the design and implementation of evaluation research, who are knowledgeable about the norms, expectations, and experiences of the target community will resolve many communication issues. Patton (1990, p. 264) cautioned that the danger in using key informants is that their perspectives will be distorted and biased, thus giving an inaccurate picture of what is happening. However, the use of several key informants selected to represent diverse view points, concerns, and experiences will enable western evaluators to obtain a more comprehensive understanding of the phenomenon, and thus will increase the reliability of the information obtained.

When collecting data, it is necessary to team up with the internal members who are skilled members of the client group. Internal Korean staff members play the role of representatives of the client group. They will make the information scope and selection under the monitor of the top client group. It is unlikely that western evaluators will assist with the data gathering process, including interviewing and surveying with a translator who is frequently an internal Korean staff member. Usually the Korean staffs are in charge of data gathering. It is typical that the internal Korean staffs will receive the responses, summarize them and provide the resulting information for the western evaluators. Then the western evaluators will collect all the information in order to develop a tentative draft of the detailed evaluation design.

When western evaluators analyze information, it is expected that they will keep working closely with their Korean staff. When using methods and techniques, a major task is to adopt western research techniques, and then adapt them by adjusting to the Korean culture. This concern highlights the importance of understanding the target culture and relating the research and instrument design to the culture Korean staff members who are knowledgeable of western techniques and culture may be better able than western evaluators to anticipate and reduce potential biases. They can then modify research design, sampling procedures, instrumentation, and/or interview processes. The internal staff can make better judgments about what should be changed or improvised. Western evaluators can then confirm the reliability and validity of the information by asking the same research questions to several other Korean staff members.

When western evaluators are involved in collecting and analyzing information, the language can be resolved in a unique way. Language may get in the way of conducting an international evaluation project. Koreans think that they should communicate with foreigners in English, not in Korean. Needless to say, it is considered a favor to try to learn Korean, and this makes it easier to get close to native Koreans. However, at least in Korea, I would say that western evaluators do not have to worry about a language deficiency. While language is significant, Korean clients will not resent a foreigner's language deficiency. Koreans are

generous to foreigners who can not speak Korean very well. There must be a factor that mitigates the deficiency. If western evaluators do not know the language of the target community, they can be relatively free from any sociopolitical effect. Also, western evaluators can usually find English translators inside the Korean organization. It is highly likely that a staff member who is proficient in English is, at the same time, working as a key informant for western evaluators.

Also, as Pareek and Rao (World Bank, 1998, p. 137) indicated, informing subjects is the most serious ethical consideration in collecting and analyzing information. However, in Korea the task of 'fully' informing research subjects who are unfamiliar with the purposes and procedures of the study is not the first concern. If the researchers show their names and assigned position, briefly introduce the research purposes, indicate how the results can be used, and get permission via telephone or email, Koreans will cooperate. In those cases, western evaluators can get more information if they are accompanied by a Korean staff member. It is more important that the Korean staff members understand western research methodology, because they are the ones who organize the data collection process, including interviews and surveys with western evaluators. Korean clients will value the western research mindset, and the professional techniques that western evaluators bring, since they believe western systematic methodology is very useful in analyzing information. Among western countries, especially the U.S. has made a strong impact in Korean politics, economics and education since she got independence from Japan in 1945. Especially, the Korean academic society is highly U.S. centered.

In the past, cross-cultural projects sounded like they were emerging from ethnocentric beliefs. The projects historically exploited and marginalized Korean cultures in order to pursue, the academic, political or economic developmental objectives. Koreans have had some degree of hostility towards western evaluators in the past; however, they are now more open-minded to western projects.

Another aspect of cultural difference that may influence the data collection process is the target community's level of familiarity with western scientific beliefs and methods. The lack of familiarity with western research methods may complicate many aspects of the research project (World Bank, 1998). Unless Korean staff acknowledges and become familiar with the advantages of western research techniques which evaluators possess, information may become biased. For example, in the case of focus group interviews, it is recommended that the interviewees be composed of an almost homogeneous group in terms of socioeconomic status, level of education, sex and age. Koreans have a strong group consciousness, and different groups of people are inclined to have different kinds of restrictions regarding who may ask, what types of questions, of whom, and under what circumstances. Therefore, the evaluation team should consider with whom it is appropriate for the participant to discuss these topics. Consideration of where it is appropriate for the participants to discuss these topics and under what circumstances can be ascertained by the Korean staff members. For example, Korean staff members can suggest what is a good time to apply the Delphi technique because Koreans are inclined to reach consensus easily on a certain topic. Unless the Korean staff (the gatekeeper of the data) helps

western evaluators in maintaining the authenticity of information as well as its reliability and validity, it is hard to expect to collect accurate information. In addition, the follow-up treatment of participants for the data collection process could be done better by Korean staff. For example, they know that payments and incentives offered to research subjects in exchange for their participation, if necessary, should be put into a white envelope. This is one traditional courtesy by which Koreans express their gratitude, congratulations, and sympathy. Unless western evaluators adhere to these kinds of cultural norms, they may offend Korean participants.

Lastly, revision of instruments should be made in a culturally-specific way. Developing new or adapting existing instruments or protocols that meet the needs of the study and are sensitive to the unique characteristics of Korean society is required. Whenever possible it is a good idea to get feedback from members of the community after designing and modifying instruments. Through close collaboration with Korean staff members, sampling and instrument bias are more easily accessible.

A8 Analysis of Quantitative Information. Koreans are very sensitive at numbers. News release reports contain statistical data. Much social research conducts both quantitative and qualitative data analysis, but prefer statistical analysis in reporting a project outcome. Through news release, statistics inform people of evaluation results.

Interpreting Information: Korean Communication Style Should Be Considered: Indirection  
Situational Inference; Complementary Way of Thinking

P2: Formal Agreements. Public affairs were documented formally, but in daily life Koreans do everything without a formal agreement. They treat many things informally. Koreans depend less on memorandums or contracts in business settings than western people. Evaluators must take cultural norms into account not only when designing processes, but also when interpreting the responses, because style of interaction and conversation vary across cultures. It is difficult for western evaluators to get authentic and rich information from Koreans and interpret it without understanding their ethnic communication characteristics, indirection, situational inference and complementary ways of thinking. That is the reason western evaluators need the help of Korean staff as key informants throughout the entire data collection process.

Western evaluators may see that there are few Koreans who answer any question clearly and openly. This does not mean Koreans are shy or calm, rather, it is because they are not as experienced as Americans in expressing their personal opinions. Many Koreans are reluctant to speak to strangers, even though they may guarantee confidentiality. Therefore, it is better to start a conversation in a roundabout way instead of immediately directing the conversation towards data collection.

The Korean method of communication is indirect. The cultural norm of indirection is in particular contrasts to western people's assumptions and values surrounding their ideas of confrontation. Most Koreans do not like to reveal their feelings to others, much less, those to whom they do not know. They simply use a small number of facial expressions to convey their

feelings. This partly stems from the historical influence of Confucianism. Korean ancestors held the philosophy that the serious person, stern or strict, is better than a smiling (Lee, H., 1998). They look like they are having trouble smiling at strangers. Foreigners say that most Koreans' facial expressions are flat and dull and they are poor at verbally expressing their opinions. For example, a western evaluator might feel offended in working with Korean staff members because they rarely say "thank you," "excuse me," or "I'm sorry," even if they bump into others. Sometimes, paradoxically, Koreans can not express their feelings because they feel very sorry. This is not because most Koreans are shy or node, but because they value indirection rather than expressing themselves openly.

Generally speaking, Koreans are not good at verbal communication; they are better trained to communicate by the principle of "the edge of eyes." In Korean family, work place, and society in general, they have a cultural trait that nonverbal communications based on insights are richer than verbal ones based on words. The following example represents a total understanding of the situation and the person without relying on verbal communications.

It starts raining on a cloudy day. A woman is pressing clothes, while breast feeding her baby. Her mother-in-law, who is in the next room says, "Grandson, do you want me to play with you?" Here, her words are not directed toward the grandson who is sucking the breasts, but toward her daughter-in-law, delivering a message that she needs to remove the laundry from a clothesline because it is raining. The daughter-in-law immediately grasps, by the insight, the real intention of her mother-in-law and removes the laundry . . . From the lesser family to the greater state, Korean people communicate too much by insight. If the insight is smoothly shared among the people, as we saw in the above example, there are no troubles; but if not, they face some serious conflicts and even hatred. (Lee, 1977, pp. 198-99).

This kind of cultural norm may dictate that respondents answer questions indirectly, thereby complicating the interpretation of responses. Western evaluators need to keep in mind that Koreans may be constrained by their inability to articulate their own expressions, feelings, or actions. More often than not evaluators may hear ambiguous responses which are difficult to interpret, such as "it's hard to say," "I can't tell right this moment," or "please, ask my friend" which means "please, don't ask one." This is especially true if evaluators want to get valid and reliable information on some "bad feeling experiences" from Koreans, or fallout experience at school. In such cases, western evaluators must listen to Korean staff members' advice. For example, in asking about the fallout experience in high school, it is better to present open-ended questions and allow enough time to write their experiences down than to present specific items. Evaluators must determine what will be hard to interpret, and look for alternative ways to get more frank responses.

Patton (1990, p. 339) cautions that researchers can not simply presume that they have the right to ask intrusive questions, and many topics may be taboo. Interestingly some socially implausible questions in western culture to strangers such as "how old are you?" (in this case Koreans usually ask college class) or "are you married?" are often accepted without serious



conflict. Koreans do not share western norms of the acceptability of discussing oneself and one's experiences in public, even though it is hard for western evaluators to consider this taboo. This also comes from the tradition of Confucius, the complementary way of thinking. Briefly, Confucianism could be expressed by the terms "moderation and balance". Confucian thinking never goes to extremes. When a complementary way of response exists such as "I don't know," Koreans will usually prefer it to direct agreement or disagreement. They are reluctant to respond it is strongly yes or no to an interviewer even though they may have an opinion. Koreans think that a balanced response that minimizes the loss of each alternative is the best. They avoid the imprudent decision based on a single piece of information.

Choi (1998) contended that East Asians' correspondence bias is lower than that of westerners. He indicated that Koreans were highly sensitive to situational constraints on behavior; they could easily explain it and mistakenly felt that they could have predicted it. There is substantial evidence concerning the process of dispositional versus situation inference that contributes more to cultural differences. Koreans tend to believe that personal dispositions of individuals are actually shaped by the surrounding context. For example, Choi and Markus 1996 asked Korean and American college students to explain a murder case. Korean students described the particular social contexts that presumably had given rise to some dispositions. For example, "He was violent because he had a hobby of shooting," or "He became ambitious because he had grown up in a small town." Remarkably, these kinds of situation explanations are rare for American students. Work by Cha and Nam (1985) made a similar point. Their Korean respondents, unlike American respondents, made substantial use of consensus information when making attribution about a person's behavior. Their research results show that Korean respondents were more contextually sensitive than the Americans predicted. Compared to Americans, Koreans predicted that more people would engage in the given behavior when the situation facilitated that behavior. Similarly, they predicted that fewer people would engage in the behavior when the situation inhibited the behavior. Miller (1995) found that such cultural differences may appear gradually through different kinds of socialization.

Since Koreans live in a world fundamentally defined by social relationships, they are expected to know a great deal about the people around them, including their beliefs and attitudes. Not only is public argumentation considered unnecessary, but it is actively shunned because engaging in it means to stand out, risk public disagreement, and lose favor with fellow members of society (Becker, 1986; Galtung, 1981). In relation to this point, Koreans complementary way of thinking makes people less vulnerable to social influence. I suspect that Koreans' acute sensitivity to situational influence on behavior may enable them to avoid, paradoxically, undesirable situational influences on their own behavior. It means they want to be in balance with people around them, and not to go too far. So when a western evaluator makes a survey item or interview agenda, it is recommended that they exclude extremely strongly worded questions. Also, western evaluators should keep in mind that they sometimes need to interpret Koreans' response according to the particular situation.

Reporting Information: Client-Centered or Sponsor-Centered Reporting

P7 Conflict of Interest. Conflicts of interest happen in every society, and each society has its own culturally-specific solution. The Korean way of conflict resolution is, in most cases, based on what is best for the group to which the individuals belong. A nation's interest is placed before everything. This phenomenon, the priority of group interest, is common in most organizations of Korean society. Koreans live in a culture where an authority figure, for example a father in a family, an immediate manager in a business team, or a teacher in a class, can exercise considerable control over individuals. Public opinion formation by a stakeholder group is not so prevalent compared to western societies in which individuals have opinions about most issues, and have the right to speak out. Many foreigners notice that Koreans do not seem to demand their rights as individuals. Instead, they unquestioningly rely upon, and accept the right of the hierarchy to make the best decision for them. This is because Koreans think that the small, personal interest should sometimes be ignored for the bigger group interest.

For example, several years ago the Korean government evaluated the current city traffic system to learn the best way to maximize public interest. Outside evaluators composed of various specialists from each sector, such as human resources, environmental law, and transportation, did surveys and interviewed stakeholders throughout Seoul and its suburban area. As a result of the evaluation's conclusions, everyone was prohibited from operating their automobiles once every ten days. This meant that each person could drive a car nine out of ten days, and that these days rotated so that there would be about a ten percent reduction in traffic every day. Most Koreans willingly obeyed the regulation and it finally worked, so the evaluation project was successful. It is the Korean sensibility that the individuals' sacrifice ends up being good for the entire nation's interest. One newspaper article below shows the basic function of a Korean evaluation project.

The Office of National Budget said Wednesday that it will form an evaluation team for 13 government-invested corporations and launch yearly evaluations of their performances. Evaluation results will be reported to the President and National Assembly by June 20. If the results indicate poor performance, then a recommendation that the head of the firm be replaced will be issued (Chosunilbo, December 10, 1999)

U1 Stakeholder Identification. The range of a stakeholder, when compared to a western range, is quite limited in Korea. It is rare that broad groups of stakeholders are considered with clients in an evaluation project. According to Scriven's thesaurus (1991), a stakeholder in a program is defined as one who has substantial ego, credibility, power, future, or other capital invested in the program, and thus can be held to be to some degree at risk. The Korean concept of a stakeholder is quite different. Some Korean stakeholders in Scriven's sense have no credibility or power even though they are susceptible to program changes because a policy of sacrifice of an individual stakeholder's interest is still dominant. Ethical reasons can be rationalized as being in the group's best interest. It is more likely that western evaluators need to meet only the immediate stakeholders' expectation, which is usually the clients'. Korean clients may say their interests identify with the stakeholders', stressing these two groups as one unit.

Accordingly, it may be impossible for western evaluators to have broad contact with the stakeholder group, so they are too removed to respond faithfully to them. So western evaluators do not have the responsibility to make evaluation results available in the language of all stakeholders through oral explanation or translation of their report. Korean clients do not always ensure that the full set of evaluation findings, even along with pertinent limitations, are made accessible to all stakeholders. In many evaluation projects, the audience is the top client group. It looks as though certain individuals are perceived as having a legitimate right to make a decision. Western evaluators come to know that there are quite a few questions the staff members can not answer even though they may know the answer. The responsibility of western evaluators is limited to reporting to the top client, and making good recommendations. Afterwards, top clients can make decisions as to how the results will be shared throughout the organization. It means that the top client group decides the extent to which evaluation results are shared. Sometimes Korean clients do not believe the evaluation results a western evaluator. This is because, in most cases, the 3 outside evaluators have not fully interacted with clients, and have failed to make good recommendation closely related with the Korean situation that needs to be evaluated. Western evaluators do not take responsibility for the use of evaluation results, and they can recommend how the results can be used. The authority of the client group will be in charge of their use.

#### Managing Time and Personnel: Impatience and Seniority

U6 Report Timeliness and Dissemination. In managing projects with a Korean staff, there are several useful pieces of advice regarding Koreans' characteristics: impatience and seniority. There is a Korean proverb that says you should "fry beans like lightning." Koreans are shown doing business very quickly with foreigners. Western evaluators may think that Korean clients are demanding and the staff impatient. Koreans have a disgraceful phrase, "Korean time," which means Koreans are not punctual. In recent days, however, it is viewed as a form of laziness to fail to keep a promise on time, especially when doing business. Koreans will try to keep the project on track as scheduled, and it is an acknowledged practice to push the staff to meet an officially stated deadline. It is considered tardiness if a timeline is not met. However, Koreans are known to not have a rigid sense of time between friends. So in cases where western evaluators construct a focus group which is composed of the same level of person, they may or may not be on time; however, if there is an authority figure among a certain group, the other group members will try to be there before him.

Due to very limited natural resources, Koreans know very well that unless they work hard, they can not survive in society. Western evaluators are likely to think that Koreans are hard-workers and 'hasty,' while working with them. They might want to ask Koreans, "Why are you in such a hurry all the time?" This kind of national trait might sound contradictory to the Confucius tradition I mentioned of "moderation and balance." However, Koreans make haste in order to keep balance in their lives, that is, to keep pace with others. Should they fall behind, it might influence the span of the project, that is, the timeline may be very tight. It is hard to establish a realistic timeline in which dates for reports to clients are specified. Western evaluators may always feel the pressure even though they work on Saturday mornings.

U4 Values Identification. The evaluation project should not be seen as a means of judging the program with foreigner's values. It should be seen as a process of helpful confrontation, a search for appropriate information suitable for Korean culture. In managing project resources, including time, personnel and budgets as I indicated, communication is critical. It means that communication is meaningful only if it is truly a dialogue. Since Koreans are susceptible to western advanced research techniques, western evaluators might try to export their past successes and achievements to Korea without attempting to reach the minds, tap the energies, and lift the hearts of people. Western evaluators provide Korean clients with very accurate and rich information. However, unless the information is truly communicated with the Korean clients, it might be of no use. Western evaluators need to assist in sharing in a process that gives Korean clients resources to help themselves. That is the essence of responsible evaluation consultation.

Also, issues of seniority can not be missed in explaining Korean behavioral norms. They bow to elders and speak in a more polite mode. Many Koreans feel bad about disagreeing with older persons. It is awkward and uncomfortable to evaluate a person who is older than the evaluator himself, whatever the evaluation purpose is. Salaries and benefits was decided upon how long a person has worked for the organization, and not just based on the quality of achievement. Seniority in many organizations is collapsing since nationwide organizational changes occurred at the end of 1997, which was caused by the Asian financial crisis. All persons are evaluated on the basis of their job capability. However, that does not necessarily mean that seniority does not exist. Koreans still pay much attention to the class number, the order in family structure, and hierarchy in the job position.

#### Metaevaluation: A Different Notion of Metaevaluation

A12 Metaevaluation. Concerning meta-evaluation, Koreans have a low level of evaluation appreciation. They do not think that the evaluation project itself should necessarily be evaluated. They believe re-evaluating something means that the results of former evaluations were unsatisfactory. Meta-evaluations are rarely done in public. In many cases, since evaluation projects intend to result in the modification or cancellation of an ineffective system, the evaluators appear as those who can exercise power over many people who are in the system. Accordingly, the purpose of meta-evaluation, if it is done, is to show the unreliability of the former project in an effort to try to minimize the mistakes. A major function of meta-evaluation is to confirm or correct the results of prior evaluations, unless that evaluation truly reflects clientbased worth and is an optimistic paper project; in terms of their point of view. Meta-evaluation will be conducted by someone else, an external evaluator, and not by those who designed and carried it through initially. However, even if formal meta-evaluation is not done, the top clients will evaluate an evaluation project by the following. Were the findings of the evaluation project clear, and were their methodology appropriate? Were the conclusions western evaluators made justified by the valid and reliable data? Then the top clients will make a decision regarding how the evaluation results are disseminated and used.

## Conclusion

In a nutshell, workshop participants agreed that the western evaluator's professional competence is one of responsibility, especially for their top clients' (decision-makers') project focus. Western evaluators are expected to work closely with Korean staff members, who are key informants while collecting and analyzing data. Also, the evaluators must keep in mind Koreans' communication characteristics: indirection, situational inference, complementary ways of thinking, and behavioral norms such as impatience and seniority. Western evaluation standards must be developed intelligently, thoughtfully revised, and creatively undertaken in accordance with these kinds of cultural norms. Western evaluation projects will succeed if they convey, through good communication, the conviction to Korean clients that any change recommended in their evaluation project will widen and strengthen the program.

While the need for standards was assumed, there was considerable discussion over the types of standards that should be changed among workshop participants. Many participants felt that the Joint Committee Standards did conflict with Korean values. Feasibility, Accuracy standards are mostly acceptable. However, Utility and Propriety standards such as Service Orientation, Formal Agreements., Rights of Human Subject and so on, are inappropriate. So they should be changed to more culturally appropriate ones.

There appeared to be two positions. The first position was that it is acceptable to adopt an international standard with sufficient sensitivity to Korean context. The second position is that Korea need to develop her own standards to get more fruitful evaluation outcome and to promote her credibility in international society.

This paper is intended to establish a framework of advice that can be applied to Korean evaluation projects and to guide western evaluators who wish to develop effective techniques for doing projects in Korea in a meaningful. The cultural factors mentioned in this paper are the challenges western evaluators may face in conducting evaluation projects in Korea, a small but economically promising country in East Asia. The author has tried to capture the richness of workshop participants' comments. But their opinions expressed in this paper cannot be taken as an official position of South Korea and are not even those of any individual participants. All errors are the author's.

## References

- Becker, C.B.(1986). Reasons for the lack of argumentation and debate in the far East. International Journal of Intercultural Relations, 10, 75-92.
- Cha, J., & Nam, K. D. (1985). A test of Kelley's cube theory of attribution: A cross cultural application of McArthur's study. Korean Social Science Journal 12, 151-180.
- Choi, I. (1999). Causal attribution across cultures: Variation and universality. Psychology Bulletin, 125, 47-63.
- Galtung, J. (1981). Structure, culture and intellectual style: An essay comparing Saxon, Teutonic, Gallic & Nipponic approaches. Social Science Information, 6, 817-856.
- Korean National Statistical Office (1998). Statistics based on graphics. [WWW documents. URL <http://wrrwww.nso.go.kr/grgph/k-grgp.htm>
- Korean National Statistical Office (1997). Comparison of international informatization index. [WWW document]. URL [http:// forum.nca.or.kr/index/welcome.html](http://forum.nca.or.kr/index/welcome.html).
- Kim, M. (2000). Korean mobile telephone ratio beat Japan. Hankook-ilbo . [WWW documents. URL [http://dailysports.co.kr/11\\_8/199905/hl851138.htm](http://dailysports.co.kr/11_8/199905/hl851138.htm).
- Lee, G. T. (1977). Korean mental structure. Seoul, South Korea: Moon-Ri-Sa.
- Lee, H. (1998, December 22). Koreans practice smiling. New York Times, B-1.
- Lee, Y. (1998, October 12). Systemic educational reform. Chosunilbo, p. 12.
- Miller, J.G. (1987). Cultural influences on the development of conceptual differentiation in person description. British Journal of Developmental Psychology, 5, 309-319.
- Pareek, U., & Rao, T. (1998). Cross-cultural surveys and interviewing. In H.Triandis & W. Lambert (Eds), Handbook of Cross-cultural psychology, 2. Boston: Jossey-Bass.
- Patton, M. (1990). Qualitative evaluation and research methods. Newbury Park, CA: Sage.
- Shadish, W.R., Cook, T.D., & Leviton, L.C. (1991). Foundations of program evaluation. Newbury Park, CA: Sage.
- Smith, N.L. (1991). The context of investigations in cross-cultural evaluations. Studies in Educational Evaluation, 17, 3-21.

Smith, N.L., Chircop, S., & Mukherjee, P. (1993). Considerations on the development of culturally relevant evaluation standards. Studies in Educational Evaluation, 19(1), 3-13.

Weiss, H. (1998, Spring). Interview with Dr. Carol Weiss. Evaluation Exchange, 1, 2-3.

Wholey, J.S. (1981). Using evaluation to improve program performance. In R.A. Levine, M.A. Solomon, G.M. Hellstern, & H. Wollman (Eds.), Evaluation research and practice: Comparative and international perspectives (pp.92-106). Beverly Hills, CA: Sage.

## **Standards for Evaluation: On the Way to Guiding Principles in German Evaluation**

Wolfgang Beywl  
Board member of the  
German Evaluation Society

Translation:  
First draft by Ragnar Schierholz;  
Final draft by Sandy Taut  
in cooperation with the author

This article was written as part of the effort of the evaluation standards workgroup of the DeGEval (German Evaluation Society). The purpose of the article is to initiate the development of the first draft of evaluation standards by the DeGEval. The author welcomes comments from readers. This is not an official document by the DeGEval; the author claims sole responsibility for the contents. Special thanks to Thomas Widmer of the University of Zürich for sharing materials and important preparatory work.

### **I. Introduction to Evaluation**

*The radical “value” is a reminder that evaluation is a procedure that deals with values in a fair and comprehensible way. To determine the value of something is a fundamental component of every evaluation.*

The goal of evaluation is to assess complex subjects such as governmental policies or social and educational programs and to contribute to their improvement. For this purpose, the evaluator collects descriptive data about the evaluand and uses the empirical evidence to determine its strengths and weaknesses according to predefined criteria. During the process, the evaluator goes through several steps: clarification of purposes and uses of the evaluation and contracting come first; followed by planning of content, logistics and finances; data collection and analysis; interpretation of the results, conclusions and recommendations; and finally presentation of findings. These concrete steps, as well as broader tasks like staffing and training of the evaluation team, continued cost monitoring and conflict management, are to be carried out in a way that ensures that the final product is a high quality evaluation. Evaluations need to prove their quality on all four “classical” dimensions: planning (Konzept), input (Struktur), implementation (Prozess), findings (Ergebnis).

Values and interests play a major role in every evaluation. Who is chosen to do the evaluation? How do the evaluator’s methodological competence, knowledge of the specific content area, age, gender, ethnicity, and communicative competence affect the evaluation? Is it important for there to be a high degree of correspondence between the evaluator’s value system and that of the client? (2) How is the evaluand determined? Is everything subject to evaluation (risk) or are certain elements of the program or groups of persons excluded/protected (security)? (3) Which



data collection methods are preferred; ‘hard’, distanced, quantitative methods, sometimes without stakeholder participation; or ‘soft’, communicative, qualitative methods? (4) Most obviously value-based: who defines the assessment criteria upon which the interpretation of the data, as well as the recommendations for decision-making, are based? How is the choice of criteria legitimized? What role do economic and political means of power, as opposed to democratic ways of negotiation, play in the selection process?

## **II. The Necessity of Standards for Evaluation**

*Evaluation, more than any other professional discipline, needs criteria to judge the value of its achievements.*

Since evaluation claims to deal with values and assessments in a professional, methodologically sound way, there is a need for standards that are useful and comprehensible to both evaluation experts and non-experts. These standards can serve as the basis for the assessment of evaluation practice itself (metaevaluation). Evaluation in the German-speaking countries is currently in a pivotal situation:

- For decades, evaluation had an undefined status: it appeared as a special case of applied social research; it was associated with different synonyms (e.g., performance monitoring); it was diffuse in its professional core; and, its relationship to similar concepts such as program research, controlling, quality management or organizational development was ill defined.
- New concepts of political control that emerged in Europe and on local levels, during the 1990s, triggered the mandatory evaluation of large scale infrastructure programs, schools, and children and youth services.

So far, the scope of evaluation in Germany and Austria has been limited. There are, for instance, no university degree programs, no professional journals, there is no basic text book and no job description for an evaluator. Methodological, design-related, and ethical issues in evaluation have hardly been discussed.

Private and public institutions are both willing and obliged to do more evaluations. How are they to decide to which institution to assign the evaluation? Which minimal requirements does a state-of-the-art evaluation have to meet? How can a client, a program manager, or a program participant, determine whether an evaluation is of high quality? Given the ongoing restructuring of public and private service delivery, these questions become more and more pressing.

In North America and Australia evaluation, as a discipline, has grown over decades and standards were introduced in a state of professional maturity. In German-speaking countries, the chronological order will be inverse. The standard-setting process marks the starting point and will probably strongly influence the professional development.

The advantage of the inverse order is that already existing standards put out by various international organizations and those developed by German professional associations within the last ten years, can be used as a sound basis for the formulation of novel evaluation standards.

### III. A Comparison of Standards

An analysis of existing sets of standards will highlight different options for developing a set of German evaluation standards:

#### 1. Quality of What?

*Who or what is actually subject to quality demands? Is it the evaluation or the person conducting the evaluation?*

The most well-known set of standards describing the quality features of evaluations, *The Program Evaluation Standards*, were developed by the Joint Committee on Standards for Educational Evaluation (1994, 2nd edition) (JCS Standards). The 240-page book contains the 30 standards categorized by four central attributes: utility, propriety, feasibility, and accuracy. Every standard gives one or two sentences of instructions on how to achieve a high level of quality for each of the four dimensions. Every standard is described and explained in detail. There are 'Guidelines' for each standard which, if being followed, increase the probability of achieving quality as defined by the standard. In addition, 'Common Errors' are listed. Finally, examples from school, university, continuing education and social services contexts illustrate the application of each standard and its guidelines.

The most important set of norms and standards focusing on evaluators as a professional group are the *Guiding Principles for Evaluators* (1995), endorsed by the world's largest evaluation association, the American Evaluation Association, counting about 2,500 members (AEA Principles). The five principles introduced in the eight-page document advocate (1) systematic inquiry, (2) competence, (3) integrity/honesty, (4) respect for people, (5) responsibility for general and public welfare.

There is another, similar set of standards, the *Guidelines for the Ethical Conduct of Evaluations* (1998) by the Australasian Evaluation Society (AES Guidelines). These principles not only address evaluators, they are also geared toward persons who commission evaluations or those who do research, teach or publish on the subject. The three main chapters correspond to the steps of an evaluation: contracting and planning, implementation, and reporting - including a total of 22 guidelines.

## 2. Standard Setting by Whom?

*Which organization or which panel defines the set of standards or principles?*

In the case of the AEA Principles and the AES Guidelines, and similarly the *Guidelines for Ethical Conduct* by the Canadian Evaluation Society, standards were defined by professional associations whose members are mainly evaluators. Typically, standards that refer to persons as being subject to quality demands are developed by scientific or professional associations. Occasionally, such professional associations also define sets of standards addressing scientific methods and processes, e.g. the standards of the Schweizerische Evaluationsgesellschaft (Swiss Evaluation Society).

Ministries and other governmental organizations also set standards. *The Guiding Principles for Program Evaluation in Ontario Health Units*, published by the Ontario Ministry of Health and the Ottawa-Carleton Health Department in 1997, provide an example. The 11 guidelines, each consisting of 2 to 5 instructions, put an emphasis on evaluation being integrated in program planning. Such an approach requires a clear description of the program and its goals as well as explicitly defined decision needs.

Another example—on an international level—are the *Best Practice Guidelines*, developed by the Public Management Service at the OECD in 1998 (PUMA Guidelines). They primarily address persons who are responsible for evaluation in a political context (governmental organizations, politicians, top-level civil servants). The document clearly focuses on the summative function of evaluation (to provide a reliable basis for decision-making and resource allocation).

In contrast, the Joint Committee on Standards for Educational Evaluation is comprised of a mixture of professional associations and representatives of public institutions. The committee currently connects 16 organizations, among them AEA, American Educational Research Association, and the American Psychological Association, the Council of Chief State Schools Officers, and several associations of professional educators.

## 3. Sanctions?

*Do the standards suggest an individual commitment on a voluntary basis or do they implicitly or explicitly include external sanctions for noncompliance?*

Most of the investigated standards do not suggest any explicit sanctions. The JCS standards, for example, are meant to give guidance but are in no way associated with accreditation or a means to impose sanctions. However, there is a first report of the JCS standards being used in a civil lawsuit. In this case, a client refused to fully pay for an evaluation because of its allegedly bad quality. The potential for similar indirect mechanisms of sanctioning is contained in the standards published by public agencies. These standards will most likely serve as a major reference for the assessment of evaluation services which in turn will be used for deciding upon future evaluation assignments.

A more direct way to impose sanctions is provided by professional associations that convene a tribunal in cases of violation of professional principles, e.g. the common ethics code published in 1993 by the German Sociology Society and the Professional Association of German Sociologists (DGS/BDS-Codex). The *Ethische Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V.* (Ethical Guidelines of the German Society for Psychology and the Professional Association of German Psychologists), published in 1998, can have far-reaching consequences. Because the professional association has a strong impact on its members' job opportunities, exclusion can have serious economic consequences for the affected professional.

#### 4. Extent?

*How detailed and extensive should a set of standards be? How should it be structured?*

The length of the investigated sets of standards varies between 3 pages (Ontario Guidelines) and about 240 pages (JCS standards). The majority of documents contains about 4-12 pages. The JCS standards are also available in a short version, covering just the names of the 30 standards and their 3-sentence descriptions. The short version makes them easy to review and apply.

#### IV. Recommendations

It is time to create a set of evaluation guiding principles for use in German-speaking countries which is concise, ethically appropriate, and systematic. However, resources for achieving this goal are scarce. A clear definition of goals and suggestions for an efficient procedure are needed.

The German Evaluation Society has the legitimate authority to develop guidelines focusing on the work of evaluation professionals. However, target groups should also include clients and other important factions in the advancement of the evaluation profession. The AEA Principles as well as the CES and AES Guidelines can be used as important references. Standards originating from other disciplines in German-speaking countries, e.g. the DGS-BDS-Codex, can help determining the ethical, legal, and cultural characteristics of Germany (and Austria), thereby facilitating the adaptation of the foreign sets of evaluation standards.

The German Evaluation Society should actively promote the development of standards for evaluation. This task should be achieved in cooperation with other professional associations, including those of primary evaluation client groups, e.g. Deutsche Städte-und Gemeindetag (German board of local governments), Verein für öffentliche und private Fürsorge (Association for public and private welfare) and Wissenschaftsrat (Science council). Currently, it remains unclear who should start such a resource-intensive initiative. The standard-setting processes resulting in the JCS as well as the SEVAL standards can serve as models.

For the time being, it seems sensible to participate in the development of "sectoral" standards, e.g. in the fields of development assistance, education, health or youth services policy, as long as

such initiatives are supported by representatives from these fields. The Ontario Principles are a good example for such a concise set of standards.

Evaluation is by definition an interdisciplinary and intersectoral field of endeavor. In view of the ongoing European integration, an international perspective becomes more and more important. This should have an impact on the development of standards—concerning their content as well as the design process. Development will require the cooperation between different national evaluation societies.

*Originally published in German in the DeGEval News No 1, April 1999  
DeGEval e.V.  
c/o G. Tissen  
Lorenweg 41  
D-53347 Alfster  
Germany*

#### **Publications on the subject**

- Beywl, W.: Anerkannte “Standards und Leitprinzipien de amerikanischen Evaluation (“Accepted Standards and Guidelines in the American Evaluation”); in Heiner, Maja (Publisher), Qualitätsentwicklung durch Evaluation (“Achieving Quality through Evaluation”), Freiburg 1996, pp. 85-107
- Joint Committee on Standards for Education Evaluation (Publisher): Hanbuch de Evaluationsstandards (“Handbook of Standards for Evaluation”), caq. 310 pages,. 68 DM Leske + Budrich. 2<sup>nd</sup> ed. 2000, ISBN 3-8100-1977-1



## **Evaluating Evaluations: Does the Swiss Practice Live up to the ‘Program Evaluation Standards’?**

Thomas Widmer  
Institute of Political Science  
University of Zurich

### Abstract

In a meta-evaluation of fifteen Swiss evaluation studies, *The Program Evaluation Standards* have been applied as assessment criteria. The studies cover different policy fields (environmental, industrial, energy, social policy, and foreign policy). The analysis is based on written materials and qualitative interviews with both stakeholders and evaluators. The article assesses the relationship between *The Program Evaluation Standards* and current Swiss evaluation practice. It also highlights the factors that often prevent evaluators from undertaking thorough and provocative examinations in a small, consensus-oriented democracy without a long tradition in evaluation research and practice. Finally, the paper reflects on whether *The Program Evaluation Standards* are appropriate for a different sociocultural, political and policy context.

Paper presented at the 1. International Evaluation Conference, Vancouver, Canada, 1-5 November, 1995

## 1. The Swiss Political System and Evaluation

Evaluation, defined as a systematic inquiry assessing the outcomes of political programs, depends very heavily on its political context. The Swiss political system has some quite special features (compare Linder, 1994):

First, Switzerland has a three level government. Not only the Confederation, but also the 26 cantons (federal states) as well as the 3000 communities and municipalities have their own financial, legal and political resources. The power of government is therefore shared among the authorities on all three levels. “Most of the federal legislation is implemented by the cantons.” (Klöti, 1988: 194). The cantons have kept full autonomy in a limited number of areas (i.e. education, health, churches, taxes and political self-organization). Power sharing among the federal, cantonal, and municipal governments is very important. The organizing principle of this three-level system is subsidiarity. This means that the higher level has the power to intervene only in cases where the lower level is not capable of doing so successfully.

The second important characteristic of Switzerland is its small size (see Klöti, 1987: 2-4). Switzerland has approximately 6,700,000 inhabitants, which equals an average of 260,000 individuals per canton and 2,300 persons per municipality.

Third, Switzerland is characterized by a semi-direct democracy. Direct democracy has an important influence on the political system, namely on the decision rule: The dominant decision rule is consensus. On the federal level, the four political parties that have been represented in the federal government’s executive body in the same proportion since 1959 represent more than seventy per cent of the electorate.

The fourth aspect is the relative secrecy with which political decision-making takes place in Switzerland. According to the system of consensus, the principle of cooperativeness, called “Kollegialitätsprinzip” resp. “principe de collégialité”, prevents the individual members of the government from deviating from the official course of the body. In addition, the executive bodies of the federal government and almost all the cantons and communities do not apply the principle of non-restricted review of their decision-making process (“Öffentlichkeitsprinzip” resp. “principe de la publicité”).

In Switzerland, evaluation, and social science in general, has a relatively short history (see Bussmann, 1996: 307-308; Horber-Papazian and Thévoz, 1990). Until the mid-eighties, few in Switzerland talked about evaluation of the public sector. Since that time, new initiatives and changes have taken place. In 1987, the Federal Department of Justice and Police established a working group for legislative evaluations. This working group succeeded in contracting evaluation studies and further research projects on the use of systematic inquiry in Swiss legislation. In 1991, the group published its final report (AGEVAL, 1991). The report describes experiences in the evaluation process and contains some recommendations for practice and utilization of evaluation in Switzerland. In addition, the Swiss National Science Foundation started a national research program called *The Effectiveness of Public Programs* in 1989 (see



Bussmann 1996). Under the auspices of this program, a wide variety of evaluation studies was conducted to check the performance of different evaluation designs, methods, and techniques. The research program will be completed this year (see Bussmann, 1995; Bussmann et al. 1996). The meta-evaluative research project presented in the next chapter was a part of this program.

On the federal level, and to a lesser extent on the cantonal and local level, the use of evaluation has made some progress. First, the Federal Parliament has established a parliamentary administration control office whose duty it is to monitor administrative activities. This office acts only on the orders of the permanent commissions of the National Council and the Council of States. Since the new office was established, the parliament has used evaluation as a common instrument of administrative control. Second, the executive power now has its own office working on evaluation projects. The Federal Council's administration control office is in charge of monitoring administrative actions regarding effectiveness as well as efficiency.

Besides these two services, the government audit office has broadened the scope of its activities. The government audit office is increasingly involved in evaluative questions. In general, it is now more common to conduct evaluation studies in other branches of the central government. In the energy policy field, for example, many evaluations were carried out investigating the effects of the program called *Energy 2000*. This program promotes energy saving and the use of renewable energy sources. The evaluation studies analyze different program measures. The goals of the program were used as evaluative criteria. This is only one example of the widespread activities in evaluation that have taken place in the last few years.

In spite of the developments mentioned above, Switzerland still belongs to the countries lacking a strong evaluation community. Only a small group of social scientists is active in the field. There is little evaluation training in the social science curricula at the universities. There are no serious training programs at all for professional evaluators. A professional organization called Swiss Evaluation Society (SEVAL) has existed since 1995. Up to the present, the Society has been quite small and its activities have been rather limited. Switzerland still has to take many steps to become a country with a well-established evaluation tradition.

## 2. The Meta-evaluation: Research Questions and Design

The research project presented in this chapter was conducted from 1993 to 1995. It is called *Meta-evaluation—Criteria to Evaluate Evaluations* (see Widmer, 1996a&b). The central research question is divided into two parts:

- Does the Swiss practice live up to *The Program Evaluation Standards*?
- Are *The Program Evaluation Standards* appropriate for evaluating non-educational evaluations within the Swiss context?

Fifteen evaluation studies were selected from different policy fields in an effort to answer these questions. In a thorough analysis, each study was evaluated using an adapted version of *The*

*Program Evaluation Standards*, originally developed by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1981). Since the project started in 1993, it was not possible to use the recently published, second edition of *The Standards* (Joint Committee, 1994) as a reference in the empirical work. However, the 1994 edition was used to draw conclusions from the analyses. To facilitate the use of *The Standards*, they were adapted to Swiss conditions. This adaptation process included a translation into German and a slight adjustment to the non-educational and Swiss evaluation fields.

The resulting thirty criteria were used to evaluate each of the fifteen evaluations in an individual case study. The case studies are partially descriptive and partially evaluative. The descriptive part included information about the evaluand, the evaluative question, the evaluation design, the client, the evaluators, and a thick description of the evaluation process. The evaluative part contained a thorough assessment of the studies using the thirty evaluation criteria.

The alternative criteria proposed by Egon G. Guba and Yvonna S. Lincoln (1989: 236-243) were applied to those evaluation studies which did not follow a traditional scientific paradigm. These alternative criteria are more appropriate to evaluation studies that have an alternative paradigmatic orientation (e.g., naturalistic evaluation approaches). Both kinds of criteria were defined as idealistic criteria and were, therefore, hard to fulfill in reality: difficult circumstances confront every evaluation process (time and financial constraints, political pressure, poor data and so on). Therefore, in each case study, a short impressionistic (and subjective) judgment follows the assessment along the evaluative criteria. Furthermore, the case studies contain documentation which includes lists of the written sources consulted and of the interviewees.

The sample, consisting of the fifteen case studies, was drawn in two steps. A list of 110 evaluation studies conducted since 1990 constituted the universe for the selection of a sample of the first ten studies. The sampling procedure followed a set of theoretical dimensions. Among others, the following dimensions were considered: the governmental level, the institutional structure of the policy area, the financial, personnel, and time resources available to the evaluation, the methods applied, and the evaluator's professional and academic background. In this manner, the following ten evaluation studies were selected:

- Environmental policy and technological development, a study analyzing the effects of environmental regulation for heating systems on technological innovations in heating engineering (Balthasar and Knöpfel, 1994).
- *Innovation processes in enterprises*, evaluating the effects of public programs promoting economic development on innovation processes (Bierter and Binder, 1993).
- *Wage equality principle for homeworkers*, examining the effectiveness of a public program mandating equal wages for labor in factories and homework (Gerheuser and Schmid, 1993).

- *Housing program of the Federation*, an evaluation investigating the effects of a program to increase the owner-tenant ratio in the Swiss population (Schulz et al., 1993).
- *Community centers in Zurich*, analyzing the input-output efficiency of an organization that runs institutions open to the public promoting social life in the municipal districts of Zurich (Näf et al., 1993).
- *Clearing away the snow on the mountain passes in the canton of Uri*, a study searching for more efficient solutions to the problem of snow removal on the mountain passes in the canton of Uri (Sommer and Suter, 1993).
- *Fees for garbage disposal*, investigating whether the introduction of fees reduces the amount of garbage in Swiss municipalities (Haering et al., 1990).
- *The introduction of consumption-dependent charges for heating costs and the promotion of renewable energy sources*, examining the implementation and the effects of the two programs in three Swiss cantons (Linder et al., 1990).
- *Federal office for the environment, forests and landscapes*, an organizational evaluation assessing the organizational structure and the effects of the activities carried out by this federal office (Knoepfel et al., 1991).
- *The information campaign of the Federation concerning the European Economic Area-referendum*, a study dealing with the campaign of the federal government to promote the knowledge about the referendum on a Swiss EEA membership (Longchamp, 1993).

A further set of five studies, the second step in the sampling procedure, was chosen according to an additional research mandate. This mandate deals with the performance of evaluation studies under severe financial and time constraints (Widmer et al., 1996). The five studies were:

- *Investment aid in rural structural engineering in the canton of Zurich*, a study analyzing the effects of a cantonal program on structural change in agriculture (Schnyder, 1993).
- *Procedure of open bidding in the canton of Berne*, an investigation of the effects induced by the legislation on open bidding for contracts for public works and the possibilities for improvement (Suter and Walter, 1993).
- *Prevention of natural disasters in the canton of Wallis*, a project evaluating the measures taken to prevent damage by avalanches and landslides (Wuilloud, 1994).
- *Promotion of economic development in the canton of Neuchâtel*, an evaluation studying the diverse programs for promotion of regional economic development (Forster and Jeanrenaud, 1993).

- *Swiss design competition*, analyzing the economic and cultural impacts of a design competition called ‘Design Preis Schweiz’ (Thierstein, 1994).

The fifteen case studies, each dealing with one specific evaluation study, were based on written materials and qualitative interviews with both stakeholders and evaluators. The written materials consist of a wide variety of documents containing information about the evaluation. In addition to the evaluation report and its earlier draft(s), such materials include all correspondence related to the project, research designs, intermediate reports, the minutes, the evaluation contract (and if appropriate, contract revisions of it), reactions to the evaluation in the media or in other arenas. Interviewees were selected pragmatically. In all cases, the evaluator(s) and the client(s) were interviewed. Additional interviews were conducted with the most relevant stakeholders in the policy field under investigation.

The case studies were submitted to the evaluators, and if appropriate to the clients also, for their reaction to the arguments. On the basis of these statements, the case studies were revised. The evaluators were given the opportunity to explain any differing points of view in a formal statement, which was published along with the case studies (Widmer 1996a&b). The formulation of the case studies was, therefore, an interactive process with a strong negotiation component. In drawing the final conclusions, the case studies were compared with each other according to the evaluative criteria. The results of the meta-evaluation, presented in the next two chapters, answer the central research questions mentioned earlier.

### **3. Consequences for Evaluation Practice**

The case studies revealed that the differences among the fifteen evaluation studies were considerable. One project was, in effect, not an evaluation, but an inventory of the program accomplishments (study number 13). In this case, the researchers refused to conduct a full case study and restricted themselves to a description of the project. Two other studies (number 5 and number 10) were labeled “evaluations” but were not evaluations at all in terms of the requirements imposed by the standards. They both should be called advisory studies rather than evaluations (compare Stanfield and Smith, 1984). The first conclusion follows from the above-mentioned observations: the term “evaluation” should be used carefully. The examples also demonstrate that evaluation in Switzerland is not yet a well-established profession.

Despite the fact that the meta-evaluation showed the evaluations were successful in terms of utility, feasibility and propriety, it also revealed several structural weaknesses. Table 1 shows the criteria with a positive or a negative assessment compared to the average.

Table 1: Criteria with positive or negative assessments (compare Joint Committee, 1981)

| <i>Positive assessments</i>          | <i>Negative assessments</i>            |
|--------------------------------------|--|
| A.2. Evaluator Credibility           | D.3. Described Purposes and Procedures |
| A.3. Information Scope and Selection | D.5. Valid Measurement                 |
| A.6. Report Dissemination            | D.6. Reliable Measurement              |
| A.7. Report Timeliness               | D.10. Justified Conclusions            |
| B.1. Practical Procedures            |  |
| C.1. Formal Obligation               |  |
| C.4. Public's Right to Know          |  |
| C.5. Rights of Human Subjects        |  |

This survey highlights the discrepancy among the four criteria groups. On the left-hand side of the table, the reader will find the criteria of utility (A; four out of eight criteria), feasibility (B; one out of three criteria) and propriety (C; three out of eight criteria). On the right-hand side, in contrast, the reader will find the criteria dealing only with accuracy (D; four out of eleven). A comparable study on evaluation practice in the Netherlands came up with similar findings (see Algemene Rekenkamer, 1996).

The question, then, is how to eliminate the weaknesses detected in the accuracy domain without neglecting the strengths of practice in the other three areas. For this reason, it makes sense to take a closer look at the points attracting critique. Many of the evaluation studies analyzed did not report the procedures applied in the evaluation precisely enough. This is the reason why the evaluation process is not transparent enough to allow the reader's comprehension. This general weakness is more acute in the evaluations (or parts of them) employing qualitative data gathering or data analysis approaches. The explanation for the considerable qualitative-quantitative gap in procedural description has to do with the differing current levels of methodological development the two approaches have reached. The qualitative approaches, long neglected by mainstream evaluation (as in social science), have to make up for lost time. In addition, the inconsistencies between the requirements of a practical and a scientific report are apparent. The evaluators, criticized in this respect, often pointed out this dilemma. Evaluation has the obligation to combine the two requirements. Otherwise, evaluation risks losing its practical relevance or will not be considered for the serious peer-review that will ensure high-quality systematic inquiry. The two sides have to be considered in a balanced manner, without neglecting one or the other.

In addition to the developments in the Swiss evaluation domain described at the beginning of this paper, other steps should be taken toward professionalization. The authors suggest implementing plans for a strengthening of the Swiss Evaluation Society (SEVAL). Taking into account Switzerland's small size, it would perhaps be useful to consider stronger cooperation with the newly founded European Evaluation Society (EES). With the help of a strong professional organization it would be possible to establish an educational program dedicated to evaluation. Another tool in improving the professional self-image of evaluators in Switzerland would be the formulation of guiding principles (compare Shadish et al., 1995) or standards (see chapter 4). In the authors' opinion, it would be useful to conduct further meta-evaluations as a suitable instrument in quality assessment as well as in quality management.

#### 4. Some Remarks about the Relationship between the Standards and Swiss Practice

The implementation of the translated *Program Evaluation Standards* was very fruitful. In general, there appears to be no reason to deny the applicability of the criteria to the Swiss situation. The complete coverage of all relevant topics is especially useful. The criteria set proved to be very accurate. In spite of this rather positive assessment, there are also some points of critique to mention:

1. It is awkward to use such a high number (30) of criteria. The authors believe it is important to try to reduce the number of criteria.
2. The set of criteria does not fulfill the requirements usually imposed on a social science typology. The criteria have many links among themselves; therefore, they are not independent of each other. In addition, many aspects are covered by more than one criterion. Also, not all the criteria are located on the same analytical level. This could be the reason why "chain judgment" occurs. Furthermore, the postulates formulated in the criteria conflict with each other. This leads to severe problems. As a consequence, it is nearly impossible to fulfill all the criteria if they are strictly applied. This is not a satisfactory situation for the evaluators responsible for the evaluation under investigation. But it represents the evaluator's dilemma very well.

The authors investigated the relationship among the thirty criteria in the empirical case studies with a network analysis. Looking at the newly-published revised edition of *The Program Evaluation Standards* (Joint Committee, 1994) in the light of this empirical results, the following suggestions are made regarding the criteria:

- Criterion U.4. "Valuational Interpretation" should be deleted from the list because of its strong interdependence with other criteria (especially to A.3. "Described Purposes and Procedures", A.10. "Justified Conclusions", and U.5. "Report Clarity") and because of the high number of criteria.
- The authors fully agree with the Joint Committee's combining U.7. "Report Timeliness" and U.6. "Report Dissemination" into one standard.

- Contrary to the Joint Committee's decision, the authors would like to keep to the older version of criterion F.3. "Cost Effectiveness." The new formulation is less clear than the older one and we think the new version tends to overemphasize the justification of the resources expended. Compared to the revised version, more attention should be paid to the provision of valuable information.
- The revision of standard P.4. "Public's Right to Know" to the new entry "Service Orientation" leads to problems in application in a non-organizational context. The authors suggest keeping the older version.
- In the Swiss situation, the change of P.7. "Balanced Reporting" to "Complete and Fair Assessment" seems to be of little use. While the broadening of the scope of the standard from "presentation" to "examination and recording" makes sense, the authors agree with Michael Q. Patton's critique on the impossibility of totally achieving completeness (Patton, 1994: 197).
- The authors fully approve of the modifications proposed by the Joint Committee to criteria A.1. "Object Identification". The title of the new version comes closer to describing the content of the standard and the formulation of the standard has gained in clarity.
- The two criteria dealing with validity and reliability, as the authors empirical research and methodological reflections have shown, are strongly interdependent. It would be a worthwhile improvement to combine the two into one standard. In this way, it would be possible to give a far-reaching statement on the quality of measurement.
- The new edition of the standards has an entirely new entry called "Metaevaluation". The authors judge this to be a sound addendum and disagree with the opinion expressed by Joe Hansen (1994: 559-560). First of all, Gene V. Glass calls a substantive combination of different studies a *meta-analysis* and not a *meta-evaluation* (see for example Smith/Glass, 1981). Secondly, there is a difference between an evaluation and an audit (compare Schwandt and Halpern, 1988). The third point is that metaevaluation's semantic meaning is an evaluation of an evaluation (compare the analogous formation of concepts as meta-communication, meta-theory and so on). Fourth, the term metaevaluation has often been used in the way proposed by the Joint Committee (see for example Scriven, 1969; Stufflebeam, 1978; Smith, 1981; Schwandt and Halpern, 1988; Shadish, 1992; Gallegos, 1994; and Georghiou, 1995).
- In this empirical metaevaluation, the authors used the parallel criteria suggested by Guba and Lincoln (1989: 236-243). Depending on the paradigmatic orientation of the evaluation study under investigation, the original or the revised version was applied. Thus, there were two different criteria sets available and the authors were in the position to select the appropriate one. This procedure made empirical analysis more sensitive to the approaches the evaluations followed. Evidently, this dualism is not able to catch the diversity of

epistemologies. On the other hand, it also restricts the comparability of the case studies elaborated with the different criteria sets. The dilemma between flexibility and sensitivity on the one hand, and comparability on the other, is a subject for further discussion.

The suggestions are not meant as a critique of the work done by the Joint Committee. They are intended as input to the ongoing discussion of context sensitive evaluation standard formation (compare Marklund, 1984; Raven, 1984; Nevo, 1984; Smith et al., 1993; Hendricks and Conner, 1995). As is probably the case other countries, Switzerland needs a set of criteria defining good practice in evaluation which takes its specific situation into account. *The Program Evaluation Standards* are definitely a helpful starting point for the Swiss discussion.



### References

- Arbeitsgruppe 'Gesetzesevaluation' (AGEVAL) (1991) *Die Wirkungen staatlichen Handelns besser ermitteln: Probleme, Möglichkeiten, Vorschläge. Schlussbericht an das Eidgenössische Justiz- und Polizeidepartement. Oktober 1991*. Berne: EJPD/EDMZ.
- Balthasar, A. & Knöpfel, C. (1994) *Umweltpolitik und technische Entwicklung. Eine politikwissenschaftliche Evaluation am Beispiel der Heizungen*. Basle: Helbing & Lichtenhahn.
- Bierter, W. & Binder, H. (1993) *Unternehmerische Innovationsprozesse. Die Wirksamkeit staatlicher Förderungsmaßnahmen - eine Evaluation*. Chur/Zurich: Rüegger.
- Bussmann, W. (1995) *Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen. Ein Leitfaden*. Chur: Rüegger.
- Bussmann, W. (1996) Democracy and Evaluation's Contribution to Negotiation, Empowerment and Information: Some Findings From Swiss Democratic Experience. *Evaluation* 2(3): 307-319.
- Bussmann, W., Klöti, U. & Knoepfel, P. (1996) *Handbuch. Methoden der Evaluationsforschung*. Basle: Helbing & Lichtenhahn. (forthcoming)
- Forster, S. & Jeanrenaud, C. (1993) *La promotion économique neuchâteloise. Bilan et perspectives. Novembre 1993*. Neuchâtel: IRER, Université de Neuchâtel.
- Gallegos, A. (1994) Meta-evaluation of School Evaluation Models. *Studies in Educational Evaluation* 20: 41-54.
- Georghiou, L. (1995) Assessing the Framework Programmes. A Meta-evaluation. *Evaluation* 1(2): 171-188.
- Gerheuser, F. & Schmid, A. (1993) *Bedingt wirksam: Das Lohngleichheitspostulat des Heimarbeitsgesetzes. Serie 'Schlussberichte' NFP 27*. Berne: NFP 27.
- Haering, B. et al. (1990) *Evaluation Kehrachtsackgebühr. Expertenbericht von Infrac. November 1990*. Berne: BUWAL.
- Hansen, J. (1994) The Joint Committee on Standards for Educational Evaluations. (1994). The Program Evaluation Standards: How to Assess Evaluations of Educational Programs, Thousand Oaks, CA: Sage. Book Review', *Educational and Psychological Measurement* 54(2): 550-567.
- Hendricks, M. & Conner, R. (1995) International Perspectives on the Guiding Principles, in W.R. Shadish, D.L. Newman, M.A. Scheirer and C. Wye, Christopher (eds.) *Guiding Principles*

for Evaluators. *New Directions for Program Evaluation*, No. 66, Summer 1995, pp. 77-90. San Francisco: Jossey-Bass Publishers.

Horber-Papazian, K. & Thèvoz, L. (1990) Switzerland: Moving towards Evaluation, in: R.C. Rist (ed.) *Program Evaluation and the Management of Government*, pp. 133-143. New Brunswick: Transaction Publishers.

Joint Committee on Standards for Educational Evaluation, ed. (1981) *Standards for Evaluations of Educational Programs, Projects, and Materials*. New York: McGraw-Hill Book Company.

Joint Committee on Standards for Educational Evaluation, ed. (1994) *The Program Evaluation Standards. How to Assess Evaluations of Educational Programs. 2nd Edition*. Thousand Oaks: Sage.

Klöti, U. (1987) Swiss Democracy. Exception or Model? Zurich. Unpublished manuscript.

Klöti, U (1988) Switzerland, in D.C. Rowat (ed.) *Public Administration in Developed Democracies. A Comparative Study*, pp. 189-203. New York: Marcel Dekker.

Knoepfel, P. et al. (1991) *Evaluation des BUWAL. Expertenbericht zur Evaluation der Luftreinhaltung, des ländlichen Gewässerschutzes und der UVP des Bundes*. Berne: EDMZ.

Linder, W (1994) *Swiss Democracy. Possible Solutions to Conflict in Multicultural Societies*. New York: St. Martin's Press.

Linder, W. et al. (1990) *Verbrauchsabhängige Heizkostenabrechnung, Förderung erneuerbarer Energien. Evaluation von Prozess und Wirkungen in drei Kantonen. März 1990*. Berne: FSP.

Longchamp, C. (1993) *Evaluierung der bundeseigenen EWR-Informationskampagne. Schlussbericht*. Berne: GfS.

Marklund, S. (1984) Applicability of Standards for Evaluations of Educational Programs, Projects, and Materials in an International Setting. *Evaluation and Program Planning 7*: 355-362.

Näf, A. et al. (1993) *Evaluation der Zürcher Gemeinschaftszentren (ZGZ). Vom Januar bis Mai 1993. Im Auftrag des Sozialamtes der Stadt Zürich*. Berne: Price Waterhouse.

Nevo, D. (1984) Applying the Evaluation Standards in a Different Social Context. *Evaluation and Program Planning 7*: 383-386.

Patton, M. (1994) The Program Evaluation Standards: How to Assess Evaluations of Educational Programs, by The Joint Committee on Standards for Educational Evaluations, Newbury Park, CA: Sage, (In Press). Book Review', *Evaluation Practice 15*(2): 193-199.

Raven, J. (1984) Some Limitations of the Standards. *Evaluation and Program Planning* 7: 363-370.

Schnyder, S. (1993) Über die Verteilung und Wirkung der Investitionshilfen im landwirtschaftlichen Hochbau des Kantons Zürich (zwischen 1970 und 1990). November 1993. Zurich. Unpublished manuscript.

Schulz, H, Christoph, M. and Jörg, H (1993) *Wohneigentumsförderung durch den Bund. Die Wirksamkeit des Wohnbau- und Eigentumsförderungsgesetzes (WEG). Schriftenreihe Wohnungswesen, Band 55.* Berne: BWO/EDMZ.

Schwandt, T. & Halpern, E. (1988) *Linking Auditing and Metaevaluation. Enhancing Quality in Applied Research.* Newbury Park: Sage.

Scriven, M. (1969) An Introduction to Meta-evaluation. *Educational Product Report* 2(5): 36-38.

Shadish, W. (1992) Theory-Driven Metaevaluation, in H.-T. Chen and P. H. Rossi (eds.) *Using Theory to Improve Program and Policy Evaluations*, pp. 29-45. New York: Greenwood Press.

Shadish, W., Newman, D., Scheirer, M., & Wye, C. eds. (1995) *Guiding Principles for Evaluators. New Directions for Program Evaluation, No. 66, Summer 1995.* San Francisco: Jossey-Bass Publishers.

Smith, M. & Glass, G. (1981) *Meta-Analysis of Social Research.* Beverly Hills: Sage.

Smith, N. (1981) Criticism and Meta-evaluation, in Nick L. Smith (ed.) *New Techniques for Evaluation*, pp. 266-274. Beverly Hills: Sage.

Smith, N., Chircop, S., & Mukherjee, P. (1993) Considerations on the Development of Culturally Relevant Standards, *Studies in Educational Evaluation*, 19(1): 3-13.

Sommer, H.& Suter, S. (1993) *Schneeräumung von Urner Passstrassen. Evaluation im Auftrag der Baudirektion des Kantons Uri, mitfinanziert durch NFP 27. Schlussbericht. September 1993.* Altdorf: Ecoplan.

Stanfield, J. & Smith, C. (1984) Management Consulting and Evaluation, *Evaluation and Program Planning* 7: 87-93.

Stufflebeam, D. (1978) Meta Evaluation: An Overview, *Evaluation & the Health Profession* 1: 17-43.

Suter, S. & Walter, F. (1993) *Gestaltungsmöglichkeiten und Auswirkungen verschiedener Submissionsordnungen. Evaluation und Bericht im Auftrag der Bau-, Verkehrs- und*

*Energiedirektion des Kantons Bern, mitfinanziert durch NFP 27 und Eidg. Finanzverwaltung. Dezember 1993. Berne Ecoplan.*

Thierstein, A. (1994) *Design Preis Schweiz. Eine Evaluation. September 1994. St. Gall: SIASR/HSG.*

Widmer, T. (1996a) *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen. Berne: Haupt.*

Widmer, T. (1996b) *Fallstudien zur Meta-Evaluation, 7 Vols., Serie 'Schlussberichte' NFP 27. Berne: NFP 27.*

Widmer, T., Rothmayr, C., & Serdült, U. (1996) *Kurz und gut? Qualität und Effizienz von Kurzevaluationen. Zurich/Chur: Rügger.*

Wuilloud, C. (1994) *Evaluation staatlicher Massnahmen. Effizienz der Massnahmen zur Abwehr von Naturgefahren im Kanton Wallis. Sitten. Unpublished manuscript.*

## **Evaluation Standards**

### **Recommended by the Swiss Evaluation Society (SEVAL)**

by

Thomas Widmer, Charles Landert, and Nicole Bachmann

Translation: Sandy Taut

## **INTRODUCTION**

### **Purposes of the Evaluation Standards**

The following evaluation standards, recommended by the Swiss Evaluation Society (SEVAL Standards), aim to contribute to the professionalization of evaluation in Switzerland. Endorsement of these standards improves credibility, quality, and trust in evaluation. Only the positive collaboration of all stakeholders, i.e., evaluators, clients, and other persons involved in the evaluation, can result in evaluations of high quality. Therefore, the SEVAL standards comprise criteria that should be advocated by all stakeholders.

The SEVAL standards define evaluation objectives. Ideally, an evaluation meets all of these. However, it will not be possible in every single case to take into account each standard to an equal degree. Rather, the SEVAL standards should be adapted with regard to the specific situation. That is, some standards might sometimes be deemed insignificant, while others are attributed an especially high importance. Adjustment to specific conditions should be dealt with in a rational, open way and should be clearly explained. The adaptations should be negotiated and agreed upon by all stakeholders (clients, evaluators, beneficiaries, and others) at the onset of an evaluation.

During the design phase of the SEVAL standards, special care was taken to make them suitable for all kinds of evaluations (excluding personnel evaluations). Consequently, users have to focus the standards according to their individual needs, not only as part of the above-mentioned adaptation process, but also by specifying certain statements contained in the SEVAL standards. This interpretation process should take place in a transparent, comprehensive manner. Thus, even outsiders become empowered to understand every individualized version of the SEVAL standards.

### **Applicability of the SEVAL Standards**

The SEVAL standards can be applied to all evaluations (excluding personnel evaluations) regardless of institutional context, overall approach (e.g., internal or external evaluation), or applied field. In the SEVAL standards, the term "evaluand" is used for programs, projects,

measures, organisations, institutions, policies, products, materials, and other evaluation objects, but excluding persons.

### **Intended Users of the SEVAL Standards**

The SEVAL standards address all persons holding a stake in an evaluation; namely, evaluators, clients, and other persons in a position to support compliance with the SEVAL standards (e.g., those involved in teaching and training evaluation).

### **Development of the SEVAL Standards**

The SEVAL standards are based on *The Program Evaluation Standards* developed by the Joint Committee on Standards for Educational Evaluation (1994, 1999). The SEVAL working group "Evaluation Standards" has closely investigated the topic. This working group, comprised of federal and state representatives as well as field- and research-focused professional evaluators, appointed a committee whose members developed this set of standards with the consent of the members of the working group. The following document contains more detailed procedural suggestions as well as a list of persons involved in the development process.

### **Overview of the SEVAL Standards**

The SEVAL standards are divided into four groups, each addressing one of the following themes: utility, feasibility, propriety, and accuracy. Each of these four attributes will be described in an introductory sentence. The specific standards categorized under each dimension are listed, including a title and a one-sentence description. The standards are then elaborated to further explicate their meaning.

The individual standards, similar to the four major attributes, are not put in a particular order. The SEVAL standards therefore do not stress any standard or group of standards over another. This approach was chosen because the significance of each standard as well as each dimension differs from evaluation to evaluation.

Following the standards, the reader may find a number of supplementary materials:

- An overview (so-called "functional table") where the standards are listed according to their importance for certain steps in the evaluation process. This table allows the user to easily find those standards especially relevant in a specific situation.
- An elaboration of the procedure used to derive the SEVAL standards from *The Program Evaluation Standards* (Joint Committee, 1994, 1999), which includes justifications for the adaptations made

- A list of the members of the SEVAL working group "Evaluation Standards"
- A bibliography

## **THE EVALUATION STANDARDS**

### **Utility**

**The Utility standards emphasize that an evaluation is guided by the information needs of its users.**

#### U1 Stakeholder Identification

Persons involved and affected by the evaluation are identified so that their interests and needs can be addressed.

Persons who should be consulted in the context of an evaluation include the following:

- Those who decide upon the future of the evaluand (often the donor)
- Those who are responsible for the planning and design of the evaluand
- Those who are involved in the implementation of the evaluand
- Those who should or will be directly or indirectly affected by the evaluand (target groups and their social contexts)
- Other groups with an interest in the evaluation findings (e.g., decision makers who plan similar projects, evaluators, and the public).

These persons, groups, and institutions are referred to as "stakeholders."

#### U2 Evaluator Credibility

Those conducting an evaluation are both trustworthy and competent so that the evaluation findings achieve maximum credibility and acceptance.

The credibility of the evaluators strongly influences the feasibility and effectiveness of the evaluation. To be found trustworthy by different stakeholder groups, the following characteristics are crucial: professional competence, integrity, independence, as well as social and communication skills.

### U3 Information Scope and Selection

The scope and selection of the collected information make it possible to answer relevant questions about the evaluand and, at the same time, to take into account the interests and needs of clients and other stakeholders.

When planning an evaluation it must be considered which information is essential for answering the key evaluation questions and which is desirable but insignificant. Available resources should be allocated according to the relevance of the evaluation questions to be answered and the demands by the most important stakeholder groups.

### U4 Transparency of Assessment

The perspectives and rationale used to interpret the findings are described in a way to clarify the bases for value judgments.

The interpretation of information and results constitutes one of the most important and critical steps in the evaluation process. The interpretation is based on theoretical models and values. In order to make the evaluators' judgments convincing, comprehensible, and evaluable, it is necessary to explicitly state these value bases.

### U5 Report Comprehensiveness and Clarity

Evaluation reports describe the evaluand including its context and the purposes, questions, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

To be able to communicate evaluation findings in a convincing manner, the evaluation report (or any other reporting procedure) must be complete and clear. The report should portray precise language (e.g., clear definitions of the most important terms and consistent use of terminology), and it should be comprehensible by the intended audience. A summary of major findings in the form of a table or a graph fosters understanding. Ideally, reporting procedures and layout should be planned so that optimal reception by the target audience results. For some audiences, a detailed final report is inadequate for communicating evaluation findings. Depending on the target group and the particular situation, more attention can be attracted by presentations, workshops, or similar reporting procedures.



## U6 Report Timeliness

Significant interim findings and final reports are brought to the attention of intended users, so that they can be used in a timely fashion.

An evaluation loses most of its effect if its time line does not correspond with the audience's decision-making process. It should be noted that in many cases the report needs to be submitted considerably in advance (e.g., to public service agencies) because of internal processing before decisions can be reached. Furthermore, in many evaluation it is sensible to share interim findings with the client, especially when these results should have an impact on the client's future actions. These feedback loops are to be considered during the planning stage of the evaluation so that appropriate resources can be allocated.

## U7 Evaluation Impact

Evaluations are planned, conducted, and reported in ways that encourage stakeholders to observe the evaluation process and to use evaluation findings.

Whether or not evaluation findings and recommendations are used depends largely upon the expectations of stakeholders regarding their utility. An important prerequisite to promoting positive expectations and to be able to actually meet them is the involvement of stakeholders in the evaluation process. In addition, it has a favorable effect if continuous and clear feedback is given throughout the course of the evaluation.

## Feasibility

**The Feasibility standards ensure that an evaluation is carried out in a realistic, thoughtful, diplomatic, and cost-effective manner.**

## F1 Practical Procedures

Evaluation methods are chosen to collect necessary information while keeping disruption of the evaluand or the evaluation to a minimum.

When planning and implementing an evaluation it is important to use the most appropriate research methods. Equally important is to ensure that the methods and instruments are practical. Neither the evaluand nor the persons affected by the data collection should be unnecessarily burdened. In an evaluation context the most valid research methods often cannot not be applied because they take too much time and effort or they are ethically unacceptable. It is crucial to

discuss with stakeholders the advantages and disadvantages as well as the analytical power of the chosen methods during the planning process.

## F2 Political Viability

The evaluation is planned and conducted by taking into account the different positions of various interest groups, so that their cooperation can be obtained and so that possible attempts by any of these groups to curtail evaluation activities or to bias or misapply the results can be averted.

To avoid negative reactions to the evaluation, it is necessary to identify all groups of persons with an interest in the evaluation. Besides the participants, these may include persons not closely connected with the evaluand (e.g., persons who offer a competitive product). If expectations/needs of the different interest groups are identified and taken into account, negative reactions will be of a more predictable nature or will not manifest themselves at all. Besides the obvious interests of these groups, so-called "hidden agendas" can gain importance. Considerations about the explicit and implicit needs of different actors should also include the client's perspective.

## F3 Cost Effectiveness

Evaluations produce information of sufficient value so that the resources expended can be justified.

An evaluation is cost-effective if its expected benefit is equal or greater than its costs. Costs encompass all necessary resources. They include time spent on supervising the evaluation and financial costs covered by other institutions. Costs therefore comprise the sum of social and monetary resources spent on carrying out the evaluation ("full cost"). The benefit covers the value added by the evaluation (improved effectiveness, budgetary savings, knowledge about a program's level of acceptance, etc.). The cost-benefit ratio of an evaluation should be optimal. For example, out of a number of evaluation designs, the evaluator should choose the one requiring the least effort. Regarding designs of the same cost, the design with the greatest benefit should be selected. If in all cases considered the costs remain greater than the expected benefit, the evaluation should not be conducted.

## **Propriety**

**The Propriety standards ensure that an evaluation is conducted legally, ethically, and with due regard for the welfare of those involved and affected.**

### P1 Formal Agreement

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) is agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or to renegotiate it.

At the beginning of an evaluation, the relations between client and evaluator are usually characterized by mutual respect and trust. This is an optimal basis to establish all rules and obligations of both parties in form of a written agreement (e.g., a contract). Such a formal written agreement should at least regulate budget, time, personnel, report, contents, design, and methodology. In particular, the rights and obligations of those involved should be determined as precisely as possible. If, during the course of time, adaptations become necessary, the conditions of the collaboration can be renegotiated. A formal written contract reduces the likelihood of misunderstandings among the partners and, in case they do nevertheless occur, facilitates their elucidation.

### P2 Protection of Individual Rights

Evaluations are designed and conducted in a way to respect and protect the rights and welfare of human beings.

Human beings possess individual rights that are based on laws and accepted ethical practice, common sense and courtesy. When planning and conducting an evaluation, the rights and welfare of individuals must not be jeopardized. Persons involved in an evaluation should be informed thereof. The predictable consequences of the evaluation have to be discussed in detail. The client should refrain from prompting the evaluator to disregard the above-mentioned principles. If an evaluation leads to well-founded conclusions posing a threat to the welfare of individuals, it needs to be considered carefully to what extent the distribution of these findings is justified.

### P3 Human Interactions

Evaluations are conceptualized in such a way that interactions between the persons involved are characterized by mutual respect.

Evaluators should not impair the dignity and self-respect of persons they come in contact with during the evaluation process. By displaying appropriate behavior, hostility toward the evaluation can be avoided. This is not only a matter of human dignity but also relates to practical considerations. Persons whose dignity and self-respect are threatened not only fall short of their creative potential; they also show behavior limiting the range of evaluation activities. Therefore, it is necessary to study the cultural and social backgrounds of all those involved as well as to understand and consider the significance of specific individuals associated with the evaluation.

#### P4 Complete and Fair Assessment

Evaluations are complete and fair in their examination and recording of strengths and weaknesses of the evaluand, so that strengths can be built upon and problem areas addressed.

A balanced discussion of strengths and weaknesses of the evaluand aims at its complete and fair assessment. Even if the purpose of an evaluation is often to analyze weaknesses, positive aspects should be sought and reported. As a matter of fact, it is often possible to correct weaknesses by building upon existing strengths. In addition, it needs to be considered that correcting shortcomings can lead to weakening the existing strengths. Therefore, it is useful to ask external persons (with a possibly divergent view on positive and negative results) to review the report before its final submission. If, for whatever reason (e.g., because of time or budgetary constraints), it is impossible to collect certain data, these omissions should be explicitly pointed out. Clients should refrain from impeding complete and fair reporting by the evaluator.

#### P5 Disclosure of Findings

The formal parties to an evaluation ensure that the full set of evaluation findings is made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

All those involved or affected by the evaluation should have access to the evaluation report. Since the number of stakeholders is often very large, in many cases the report has to be made publicly accessible. Primary participants (i.e., clients, evaluators, but also other persons) are responsible for adhering to these demands. In addition, the report should be written in such a way as to meet the needs of the audience. For example, adequately communicating about an evaluation might warrant an executive summary of an extensive report and an annex including all methodological details.

## P6 Declaration of Conflicts of Interest

Conflict of interest is dealt with openly and honestly, so that it harms the evaluation processes and results as little as possible.

There are a multitude of situations in which evaluators are confronted with conflicts of interest, partially because they have their own interests that could influence evaluation findings. For example, evaluators are more or less dependent on receiving future requests from clients; they have specific philosophical, theoretical, methodological, and political beliefs; and they are part of a personal and organizational network. Conflicts between certain involved interest groups can surface during an evaluation. This can result in biased evaluation processes, results, and interpretations. Conflicts of interest should generally be avoided. Because of the just-mentioned multitude of possible conflicts of interest, this is often impossible, however. In these cases, the task is to address these conflicts in a way that does not harm the evaluation.

## Accuracy

**The Accuracy standards ensure that an evaluation produces and discloses valid and useful information.**

### A1 Program Documentation

The evaluand is described and documented clearly and accurately, so that it can be clearly identified.

The evaluand, whether it is a measure, a program, or an organization, has to be investigated carefully. It is important to take into account that the evaluand can manifest itself differently depending on time and surrounding environment. The description of the evaluand should clearly state the investigative scope. This will allow the audience to make comparisons with other evaluands. In addition, a precise investigation helps achieve an understanding of causes and effects as well as unintended outcomes. It is especially crucial to note differences between the planned and the actual implementation of the evaluand.

### A2 Context Analysis

The influence the context has on the evaluand needs to be identified.

The evaluation context is defined as the combination of all conditions surrounding the evaluand—for example, the institutional affiliation, social and political climate, characteristics of those involved and affected by the evaluation, structure of political life, neighboring or

competitive public and private activities, or economic conditions. These and other contextual factors are to be examined in enough detail to ensure that the evaluation will be planned, conducted, and reported on accordingly. The contextual knowledge is necessary to design a realistic evaluation and to make it responsive to the existing conditions. Contextual factors often have considerable impact on evaluation outcomes. A well-founded context analysis also allows the evaluator to assess the generalizability of evaluation findings. The context should not be defined too narrowly. On the other hand, the context analysis should not be too detailed to avoid taking much needed resources away from the analysis of the evaluand.

### A3 Described Purposes and Procedures

Purposes, questions, and procedures of an evaluation are documented and described in enough detail, so that they can be identified and assessed.

The purposes of an evaluation, the questions which it is supposed to answer, and the chosen approach need to be documented carefully during the course of the evaluation and should be clearly communicated in reports to the audience. This standard aims at making the evaluation process transparent. When describing the purposes and questions, it is especially important to take into account divergent points of view. The documentation and description of the evaluation process should include a detailed account of management, data collection, data analysis, interpretation, and reporting procedures. It is important to consider changes that may result in incoherence between intended and actual implementation. It needs to be pointed out which discrepancies occurred and why. These, as well as the procedures in general, are to be explicated clearly. Failure to do so can immunize the evaluation against justified criticism. At the same time, it can also encourage unwarranted condemnation.

### A4 Defensible Information Sources

The evaluation uses information sources that are described in enough detail, so that the adequacy of the information can be assessed.

The description of the information sources allows stakeholders to assess the quality of the data collected from these sources. Sources include individuals or groups, documents, audiovisual material, statistical data, etc. The use of different sources makes a comparison of information possible. The credibility of an evaluation can be challenged by a missing or insufficient description of the information sources. Besides a description of the sources, the data gained from them should also be assessed. The trustworthiness of the information should be taken into consideration when interpreting the evaluation results.

#### A5 Valid and Reliable Information

The data collection procedures are chosen or developed and then applied in such a way that validity and reliability of the interpretations are ensured.

To a certain extent, an empirical investigation is always subject to errors. Validity and reliability specify two indicators of sound data collection. These qualities can only be assessed in the specific evaluation context, considering the particular purpose of the data collection. Validity is defined by the extent to which methodologies and instruments really measure what they are meant to measure. A data collection method is more reliable the more consistently it measures (concerning different points of measurement with different instruments and among different people). Validity and reliability are closely linked. When assessing and choosing data collection instruments, both aspects need to be equally reflected.

#### A6 Systematic Review of Data

The data collected, analyzed, and presented in the course of an evaluation are systematically examined for possible errors.

During data collection, recording, analysis, and interpretation, numerous possibilities exist to make errors. These range from insignificant errors during data entry to false interpretations. Therefore, it is crucial to avoid potential sources of errors during the evaluation process. In addition, the collected data need to be examined regarding possible errors by making use of appropriate methods, e.g., plausibility checks, parallel recording, communicative validation, etc. Possible errors and their consequences have to be addressed in the evaluation report. If errors persist, false interpretations and conclusions can result. In addition, inaccurate data can discredit evaluation in general.

#### A7 Analysis of Qualitative and Quantitative Data

Qualitative and quantitative data are analyzed in an appropriate, systematic way, so that the evaluation questions can be effectively answered.

In an evaluation conclusions are drawn based on qualitative and quantitative data analysis. Data analysis should take place in a systematic manner by following rules of methodological soundness. It often proves sensible and useful to include both qualitative and quantitative information in an evaluation. The choice of information and methods of analysis is based on the nature of the evaluation questions and data availability. During the selection process, other factors like knowledge or preferences of persons involved should not play a role. The choices made and their consequences should be critically reconsidered. Strengths and limitations of the methods used should be explicitly stated.

#### A8 Justified Conclusions

The conclusions drawn in the evaluation are explicitly justified, so that stakeholders can comprehend and assess them.

The conclusions arrived at in an evaluation must be explicated and, together with the underlying assumptions and methods used, made transparent to those involved. The scope of the conclusions should also be clarified. Likewise, the report should include a discussion of alternative interpretations with an explanation of why these were dismissed. Only those presumptions shared by important stakeholders should form the basis of the conclusions. Compliance with this standard allows potential users of the evaluation findings to assess whether they are justified. In addition, the credibility of the conclusions is enhanced.

#### A9 Impartial Reporting

Reporting is guarded against distortions by any stakeholder group so that the report fairly reflects the findings.

Many different perspectives characterize the environment in which an evaluation takes place. Stakeholders often hold divergent opinions about the evaluand. An evaluation runs the risk of being dominated or abused by a certain party to the evaluation. An evaluation should refrain from simply accepting one specific point of view. Rather, all relevant perspectives need to be fairly represented. Therefore, it should be guaranteed that the evaluation takes on an independent, objective position. For example, too close a relationship to clients and those responsible for the evaluand should be avoided. The relationship maintained between evaluator and client (as well as other relevant stakeholders) needs to be clarified at the onset of the evaluation process. The role clarification should also include an agreement concerning the right to publish the evaluation report.

#### A10 Metaevaluation

The evaluation itself is subject to evaluation using these and other important sets of standards so that its implementation is guided and stakeholders can assess the strengths and weaknesses of the evaluation upon its completion.

Ineffective evaluations can lead to wrong decisions. At the same time, evaluations can be subject to unwarranted criticism. To avoid these situations, the quality of an evaluation should be examined. A metaevaluation can use the set of standards introduced with this document to assess an evaluation. A metaevaluation can, depending on the specific situation, be designed summatively or formatively. It can be of internal or external nature. As with an evaluation, a metaevaluation can be more or less extensive. Whereas an in-depth, detailed metaevaluation is



only necessary in certain cases, a concise self-evaluation by the evaluation team is a must. The resources needed for the metaevaluation (usually modest) should be set aside in the planning phase of the evaluation.

## **ANNEXES/ADDITIONAL MATERIAL**

### **Functional Table of the SEVAL Standards**

The following table displays each standard according to its importance for certain steps in the evaluation process.

#### Deciding whether to evaluate

U1 Stakeholder Identification

U2 Evaluator Credibility

U7 Evaluation Impact

F2 Political Viability

F3 Cost Effectiveness

P1 Formal Agreement

P6 Declaration of Conflicts of Interest

A1 Program Documentation

A2 Context Analysis

A10 Metaevaluation

#### Defining the evaluation problem

U1 Stakeholder Identification

A1 Program Documentation

A2 Context Analysis

A3 Described Purposes and Procedures

A10 Metaevaluation

Designing the evaluation

U1 Stakeholder Identification

U3 Information Scope and Selection

U4 Transparency of Assessment

F1 Practical Procedures

P1 Formal Agreement

P4 Complete and Fair Assessment

A1 Program Documentation

A3 Described Purposes and Procedures

A4 Defensible Information Sources

A5 Valid and Reliable information

A7 Analysis of Qualitative and Quantitative Data

A8 Justified Conclusions

A9 Impartial Reporting

A10 Metaevaluation

Collecting information

U2 Evaluator Credibility

U3 Information Scope and Selection

U4 Transparency of Assessment

F1 Practical Procedures

F2 Political Viability

P1 Formal Agreement

P2 Protection of Individual Rights

P3 Human Interactions

P4 Complete and Fair Assessment

A1 Program Documentation

A2 Context Analysis

A3 Described Purposes and Procedures

A4 Defensible Information Sources

A5 Valid and Reliable Information

A6 Systematic Review of Data

A10 Metaevaluation

Analyzing information

U4 Transparency of Assessment

F1 Practical Procedures

A1 Program Documentation

A2 Context Analysis

A7 Analysis of Qualitative and Quantitative Data

A8 Justified Conclusions

A10 Metaevaluation

Reporting the evaluation

U1 Stakeholder Identification

U3 Information Scope and Selection

U4 Transparency of Assessment

U5 Report Comprehensiveness and Clarity

U6 Report Timeliness

U7 Evaluation Impact

P2 Protection of Individual Rights

P4 Complete and Fair Assessment

P5 Disclosure of Findings

A1 Program Documentation

A2 Context Analysis

A3 Described Purposes and Procedures

A4 Defensible Information Sources

A8 Justified Conclusions

A9 Impartial Reporting

A10 Metaevaluation

#### Budgeting the evaluation

U4 Information Scope and Selection

F3 Cost Effectiveness

P1 Formal Agreement

A1 Program Documentation

A3 Described Purposes and Procedures

A10 Metaevaluation

Contracting the evaluation

U1 Stakeholder Identification

U2 Evaluator Credibility

U3 Information Scope and Selection

U6 Report Timeliness

F2 Political Viability

P1 Formal Agreement

P2 Protection of Individual Rights

P5 Disclosure of Findings

P6 Declaration of Conflicts of Interest

A1 Program Documentation

A3 Described Purposes and Procedures

A10 Metaevaluation

Managing the evaluation

U1 Stakeholder Identification

U2 Evaluator Credibility

U6 Report Timeliness

F2 Political Viability

F3 Cost Effectiveness

P1 Formal Agreement

P2 Protection of Individual Rights

P3 Human Interactions

P6 Declaration of Conflicts of Interest

A3 Described Purposes and Procedures

A6 Systematic Review of Data

A10 Metaevaluation

#### Staffing the evaluation

U2 Evaluator Credibility

F2 Political Viability

P6 Declaration of Conflicts of Interest

A9 Impartial Reporting

A10 Metaevaluation

#### **Elaborating on the Development of the SEVAL Standards**

The SEVAL standards were derived from *The Program Evaluation Standards* of the Joint Committee on Standards for Educational Evaluation (1994), translated into German by Wolfgang Beywl and Thomas Widmer (Joint Committee, 1999). In a survey, the members of the evaluation standards working group of the Swiss Evaluation Society commented on the set of standards. Based on their comments, a working group committee carried out revisions. The working group again discussed the revised version. The current document is the product of these final discussions.

The following modifications were considered necessary: Generally, the term “program” (or “program evaluation”) was replaced by the term “evaluand” (or “evaluation”). This does not constitute a major modification since the term “program” is used in a broad sense in the American original. Furthermore, to simplify the language, the short definitions of the standards containing “should” were transformed to indicative sentences. In addition to these linguistic adaptations, the following changes were made:

Merging standards A5 and A6 as well as A8 and A9

Rephrasing standards U7, F1, F3, P3, P4, P6, P7, A2, A7, A10, and A12

Deletion of standards P1 and P8.

Following, each modification is explained in detail.

### Merged standards

Standards A5 and A6 address two indicators of high quality work in the social sciences, validity and reliability. Because the assessment of each of them does not make much sense due to their interdependence (see Widmer, 1996, p. 296), the two standards were summarized under one. Since it is usually desirable for evaluations to base their conclusions on both qualitative and quantitative data, standards A8 and A9 were merged as well.

### Rephrased standards

U7 Evaluation Impact: The language was simplified, but the contents were not modified.

F1 Practical Procedures: The language was simplified, but the contents were not modified.

F3 Cost Effectiveness: The standard was deprived of its demand for efficiency because it is inherent throughout the standards and therefore redundant.

P3 Protection of Individual Rights: The original title of this standard ("Rights of Human Subjects") does not adequately reflect its meaning. Its scope is not confined to "human rights."

P4 Human Interactions: The standard was rephrased because the statement "respect human dignity and worth" does not seem convincing due to its asymmetrical nature.

P6 Disclosure of Findings: The statement "along with pertinent limitations" was abandoned, because it is not clear which limitations are meant and because other standards already cover this aspect.

P7 Conflict of Interest: Since impediments caused by conflicts of interest cannot always be averted, the strong wording "does not compromise" was changed to the more moderate "harm as little as possible."

A2 Context Analysis: The language was simplified without modifying the content.

A7 Systematic Information: The statement that all errors should be corrected was removed, because this explicit hint seems unnecessary.

A10 Justified Conclusions: To underline the message of this standard, the call for "comprehensible" conclusions was added because it constitutes an important proposition for assessing the quality of the conclusions.

A12 Metaevaluation: This standard was slightly altered in two ways. Firstly, the technical terms "formative" and "summative" were removed to facilitate understanding. Secondly, the statement "closely examine" was substituted by "assess" to allow for metaevaluations even in small evaluation projects.

### Deleted standards

P1 Service Orientation: Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

Justification for the deletion: One of the principles of the standards is that they should be applicable to the broadest range of context (as they were defined in the introduction). Not all evaluands are connected to service organizations (e.g., evaluations in the private sector). Likewise, it seems not always appropriate to consider the needs of the whole range of target groups. There might be cases where it proves important to concentrate on one specific segment of the spectrum.

P8 Fiscal Responsibility: The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

#### Justification for the deletion:

- The standard mixes different, if connected, demands (careful accountability procedures, ethically responsible allocation of resources, documentation of resource allocation, "otherwise [ . . ] prudent" allocation and expenditure of resources, responsible and adequate expenditure of resources . . . ).
- Standard F3 already covers the major point of this standard.
- After the signing of a contract, the relationship between evaluator and client is similar to any work relationship regulated by the contract. Much more significant would be the events taking place before the signing of the contract (subcontracting by public and private organizations). The client's interests would be represented more effectively as part of these agreements.
- Many contracts in the field of human services include an overall payment based on a list of expected outcomes. As soon as the contract is signed, the right to gain insight into the accountability procedures is void. From this point of view, the standard seems far from reality and naïve.
- The standard does not adhere to the principle of complementarity. Nowhere else in the standards are clients required to supply the funding necessary to conduct an "ethically responsible" evaluation with "responsible and adequate" resources.

*NOTE: The original contains a conversion table which helps translate the standards in the German version of the Program Evaluation Standards into the SEVAL standards. This is not relevant for the American/International audience of this document and is therefore not included here.*



**Members of the SEVAL working group "Evaluation Standards"**

Dr. Paul Ammann, Gesundheits- und Fürsorgedirektion des Kantons Bern, Bern

Dr. Nicole Bachmann, Bundesamt für Gesundheit, Bern\*

Sylvia Furrer Hoffmann, Fürspr., Eidgenössische Finanzkontrolle (EFK), Bern (ab 1.9.1998)

Frohmut W. Gerheuser, Büro für Politikberatung und Sozialforschung, Brugg (AG)

Silvia Kobi, Chancellerie fédérale, Berne (bis 14.5.1998)

Dr. Armin Kühne, Verwaltungskontrolle des Bundesrates (VKB), Bern

Charles Landert, Landert Farago Davatz & Partner, Zürich\*

Dr. Marlène Läubli Loud, Office fédéral de la santé publique, Berne

Dr. Urs Moser, Institut für Pädagogik, Universität Bern, Bern

Prof. Dr. Christian Simon, Schweizerischer Wissenschaftsrat (SWR), Bern

Dr. Heini Sommer, Ecoplan, Altdorf (UR)

Daniela Termignone, Eidgenössische Finanzkontrolle (EFK), Bern (bis 31.8.1998)

Patrick Trees, Geschäftsprüfungskommission (GPK) des Kantons Bern, Bern

Dr. Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich, Zürich (Leitung)\*

\*: Mitglieder des Ausschusses

**References**

Joint Committee on Standards for Educational Evaluation (Ed.) (1994): *The Program Evaluation Standards*, 2nd Edition. Newbury Park: Sage.

Joint Committee on Standards for Educational Evaluation (Ed.) (1999): *Handbuch der Evaluationsstandards. Die Standards des Joint Committee on Standards for Educational Evaluation*. Opladen: Leske+Budrich. (Im Erscheinen)

Widmer, Thomas (1996): *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.

## Annex

The complete list of the original Programme Evaluation Standards is provided below.

*Utility: The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.*

U1 Stakeholder Identification. Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

U2 Evaluator Credibility. The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3 Information Scope and Selection. Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

U4 Values Identification. The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

U5 Report Clarity. Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

U6 Report Timeliness and Dissemination. Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7 Evaluation Impact. Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

*Feasibility: The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.*

F1 Practical Procedures. The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

F2 Political Viability. The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

F3 Cost Effectiveness. The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

*Propriety - The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.*

P1 Service Orientation. Evaluation should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

P2 Formal Agreements. Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

P3 Rights of Human Subjects. Evaluation should be designed and conducted to respect and protect the rights and welfare of human subjects.

P4 Human Interactions. Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

P5 Complete and Fair Assessment. The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

P6 Disclosure of Findings. The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.

P7 Conflict of Interest. Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

P8 Fiscal Responsibility. The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

*Accuracy - The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.*

A1 Program Documentation. The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

A2 Context Analysis. The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified and assessed.

A3 Described Purposes and Procedures. The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4 Defensible Information Sources. The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5 Valid Information. The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

A6 Reliable Information. The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

A7 Systematic Information. The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.

A8 Analysis of Quantitative Information. Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A9 Analysis of Qualitative Information. Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A10 Justified Conclusions. The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

A11 Impartial Reporting. Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

A12 Metaevaluation. The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.