

**QUADERNI DEL DIPARTIMENTO DI SCIENZE
ECONOMICHE E SOCIALI**

**TRAINING PROGRAMS AND PERFORMANCE
MEASUREMENT:
EVIDENCE FROM HEALTHCARE ORGANIZATIONS**

Stefano Baraldi, Antonella Cifalinò

Serie Azzurra: Economia Aziendale – Quaderno N. 52 luglio 2008



UNIVERSITÀ CATTOLICA DEL SACRO CUORE

PIACENZA

I QUADERNI

Possono essere richiesti a:

Dipartimento di Scienze Economiche e Sociali, Università Cattolica, Via Emilia Parmense 84,
29100 Piacenza tel. 0523 599342. Oppure si può ottenere una copia dall'area di download del
Dipartimento al seguente indirizzo: <http://www.unicatt.it/dipartimenti/ScEcoSoc/default.stm>

ALL PAPERS PUBLISHED IN THIS SERIES ARE DOUBLE REVIEWED

TRAINING PROGRAMS AND PERFORMANCE MEASUREMENT: EVIDENCE FROM HEALTHCARE ORGANIZATIONS

by *Stefano Baraldi*^(*) and *Antonella Cifalino*^(*)

1. Introduction

The literature on training has pointed out the relevance of training evaluation (Bartram - Gibson, 1999; Kirkpatrick, 1959a, 1959b, 1960a, 1960b; Kirkpatrick- Kirkpatrick, 2005, 2006; Phillips, 1977, 1997, 2002). This is for many practical reasons (Phillips – Phillips, 2001):

- ▶ as training budgets are growing, it needs to evaluate training effectiveness;
- ▶ executives and managers funding the training programs are requesting evaluation data;
- ▶ training functions need to develop a comprehensive evaluation process around their programs because of increased competition for scarce resources within organizations.

According to this research stream, various criteria of training effectiveness have been identified (namely: reaction, learning, behaviour, and organizational results), with a strong focus on the tools of evaluation, such as in the case of the ROI of training (Phillips, 1977, 1997).

More recently, in response to contemporary business requirements, the literature has argued the need for approaching the topic of training evaluation according to a strategic viewpoint (Robinson and Robinson, 1998; Phillips and Phillips, 2001; Kirkpatrick and Kirkpatrick, 2005). Moving training management from an operational to a strategic approach means that (Robinson and Robinson, 1998):

- ▶ the focus is on what people need to do with respect to organizational goals, rather than on what people need to learn - acquisition of skill and knowledge is a means, rather than an end;
- ▶ the training management is an issue of training process, rather than of training events;
- ▶ the training process is focused on improving performance by changing behaviour, rather than only by teaching;
- ▶ multiple solutions of performance interventions are considered, rather than only structured learning experience;
- ▶ training functions have to operate in partnership with line managers, rather than in parallel;
- ▶ line managers enter the training process proactively, rather than reactively; front-end assessment is mandatory, rather than optional, as work environment barriers to desired performance have to be identified;
- ▶ training effectiveness is measured in terms of contribution to performance change, rather than in terms of the training program quality.

In the meantime, the literature on performance measurement has identified the need for measuring human capital, as essential drivers of organizational competitiveness (Ulrich, 1989, 1997a, 1997b, 1998, 2005). Among the others, knowledge, learning, and competences are mentioned as strategic resources across various sectors. As a consequence, the literature has clearly identified the relevance of performance measurement systems including human capital issues. This is the case of Balanced Scorecard, where innovation and learning perspective of performance is added to financial, customer, and internal processes ones (Kaplan and Norton, 1996, 2004). More specifically, human resource (HR) scorecard models have been developed (Walker, McDonald, 2001).

To sum up, even though the literature on training has recognized the strategic relevance of training, it does not develop how to integrate the training evaluation within the strategic management process. On the contrary, even though the literature on performance measurement and management has in-

^(*) Associate Professor, Catholic University, L.go Gemelli, 1, 20123 - Milan, Italy (stefano.baraldi@unicatt.it)

^(*) Researcher, Catholic University, Via Emilia Parmense, 84, 29100 - Piacenza, Italy (antonella.cifalino@unicatt.it)

cluded human capital issues, it does not focus on how to measure the performance of training as a specific component of HR strategy.

Starting from this theoretical background, the paper is ended to analyse how to apply PM systems to training programs within organizations. This topic is developed according to a strategic approach, i.e. training programs are conceived as levers of training strategy, in connection with both HR strategy and corporate strategy. Therefore, the evaluation of training programs is approached as a strategic performance measurement issue. More specifically, the paper addresses the following research questions:

- ▶ what are the limits of traditional models of training evaluation?
- ▶ is it possible to apply the strategic performance measurement systems to training?
- ▶ if so, how can strategic performance measurement tools be used to evaluate training programs?

Healthcare has been chosen as a relevant sector for providing empirical evidence on this topic, because healthcare organizations (HCOs) are professional organizations where competences (and training as a consequence) assume a strategic relevance. In developed countries, including Italy, health professionals are required to attend programs of continuing medical education. Tian *et al.* recently offered a systematic review of evaluation in formal continuing education, showing that the effectiveness of training programs is not systematically evaluated, particularly in terms of behavioural change of trainees and of patient outcome, and therefore concluded the importance to develop further research on this subject (Tian et al., 2007).

2. Research methodology

Level and unit of analysis

This research is developed on the organizational level of analysis, focused on training programs as unit of analysis. Even if there are many levers for improving human capital, this research is only focused on training programs as tools of HR management. We include different typologies of training programs such as classroom training (live or virtual), and on-the-job training. On the contrary, we exclude both learning performance interventions different from training (such as experiential learning) and non-learning performance interventions (such as work environment redesign, job rotation, motivation interventions, and job aids).

Research approach

The methodology of the paper is largely oriented to action research. The origin of action research can be found in the work of the social scientist Kurt Lewin in the mid-1940s (Lewin, 1946; Lewin, 1947). In the following decades the Lewin's basic idea has developed to become a family of approaches with many different expressions (Coghlan, 2004). Arguing the actual capability of positivist science to create universal knowledge, action research focuses on knowledge in action. In this paradigm, the researcher is not a detached observer, but becomes actively involved with the situation being studied. Thus, the validation of the findings are not coming from the consistency of prediction and control but from the knowledge that "can be applied and validated in action" (Gummeson, 1991). Therefore, an action research process is (French and Bell, 1984) "systematically collecting research data about an existing system, taking action by altering selected variables with the system based on the data and on the hypothesis, and evaluating the impact of actions by collecting more data on the outcomes."

As mentioned earlier, the growing need of a more "phenomenological" way of conducting research in social sciences originated a wide array of action-oriented approaches: constructive (Kasanen *et al.*,

1993), clinical (Schein, 1993), innovation-action (Kaplan, 1998), participatory or collaborative (Eden and Huxham, 1996), interventionist (Jonsson and Lukka, 2006).

Both the limits and the benefits of action research depends on the role researcher could play inside the organization. Actually action researcher cannot be seen as a “friendly outsider” who acts as a facilitator (Coghlan, 2004) but becomes closely involved with problem solving or change processes inside the “real” organization. As Schein noticed (Schein, 1993), “in the traditional situation researchers have to develop a site, gain entry, and establish themselves as ones who will not be too great a pain to have around.” The deal is that the researcher will be allowed to hang around, interview people, maybe even administer questionnaires, but he or she will not do too much harm to the organization. On the opposite, action researchers do not just observe and document existing phenomena but play as change agents, helping to create phenomena that did not exist before (Kaplan, 1998).

This “active” role – not so different from a consultant’s role – has frequently been criticized because of its lack of rigor. Besides, the researcher’s independency could be questioned by the need to satisfying his/her clients’ expectations. On the other hand, action research provides the opportunity to get very rich – and often not accessible in other ways – insights because practitioners are involved over things which actually matter to them (Eden and Huxham, 1996) and turn out more willing to share their ideas and information. Moreover researchers, trying to develop and apply theories in actual organizations, could benefit of what Argyris and Schon called “double loop” or “deuteron” learning (Argyris and Schon, 1974).

For these reasons action research is generally considered a valuable approach in developing and elaborating theory from practice (Coghlan, 1993; Eden and Huxham, 1996; Kaplan, 1998; Westbrook, 1995; Jonsson and Lukka, 2006) and not for a rigorous and detailed theory testing. Although they’re not repeatable, interventions in organizations “provide rich data about what people do and say – and what theories are used and usable – when faced with a genuine need to take action” (Eden and Huxham, 1996). Even more important, action research provides an ideal context for: a) experimenting how different theoretical frameworks can be actually used in the same organization; b) evaluating their feasibility; c) understanding the relationships between them; d) identifying and building a so-called emergent theory.

In recent years, several authors have tried to mitigate the main methodological issue of action research – subjectivity – by arguing that good action research must and can meet both the requirements of rigor and relevance. Eden and Huxham (Eden and Huxham, 1996) set out 12 contentions to suggest some of the standards to which action research might aspire, identifying the characteristics of action research outcomes and processes. Kasanen *et al.* (Kasanen *et al.*, 1993) divided the action research process into the following steps: 1) find a practically relevant problem which also has research potential; 2) obtain a general and comprehensive understanding of the topic; 3) innovate, i.e., construct a solution idea; 4) demonstrate that the solution works; 5) show the theoretical connections and research contribution of the solution concept; 6) examine the scope of the applicability of the solution. Similarly, Coghlan (Coghlan, 1993) stated that action research works through a cyclical four-step process of consciously and deliberately: a) planning; b) taking action; c) evaluating the action; d) leading to further planning.

Action research seemed to us an interesting approach to effectively address the issue of measuring training performance. As mentioned before, we can’t actually rely on a single theoretical framework but we have different contributions coming both from the HR and from the PM literature. Then, we have tried to understand how these different approaches could be made actionable in an organizational setting when the measurement of training performance is critical by its very nature. In HCOs action research is certainly well known, and has received an additional boost from the recent adoption of the principle of “evidence-based medicine” (Bate, 2000). Besides, like any other knowledge-intensive organizations, HCOs look at training initiatives as an essential driver of skills’ develop-

ment, invest a lot of money in them, and are generally struggling in the attempt to evaluate their return.

Research project

We worked for about 12 months with 5 Italian HCOs in order to get an actionable system to measure the performance of their training programs. We selected HCOs that conceive the role of training as a strategic lever; in order to verify this strategic orientation, we developed pilot interviews to training managers. Namely, we worked with:

- ▶ the Carlo Besta Neurological Institute, that is a Research Institute of Public Interest (IRCCS) internationally recognized as a neuroscience leading centre;
- ▶ the Humanitas Clinical Institute, that is a private Research Institute of Public Interest (IRCCS) which provides public services as a full member of the National Health Service;
- ▶ the European Institute of Oncology (EIO), that is a private Research Institute of Public Interest (IRCCS) internationally recognized for its excellence in the prevention, diagnosis, and treatment of cancer;
- ▶ the S. Anna University Hospital, that is a public HCOs located in the city of Ferrara, organized into 8 integrated divisions devoted to the delivery of clinical services as well as to research and education;
- ▶ the Local Health Unit of Lecco, that was created in 1998 by merging three public HCOs working in the province of Lecco, in order to promote efficiency and effectiveness.

The project was structured into five stages. To first stage (project's design) was ended to share the research project design with the HCOs that joined the project. This analysis included the criteria of training program classification, the models of training evaluation, the levels of training program evaluation, the tools of training program evaluation, and the composition of the project teams.

During the second stage (literature review analysis), the researchers presented the results of the literature review during two plenary sessions, one focused on the HR literature and the other focused on the performance measurement literature. This theoretical framework is reported in the following paragraph of this paper.

The third stage was structured into four plenary sessions ended to share experiences of training evaluation developed outside healthcare (IBM, World Bank Institute, US Army, and Isvor Fiat).

The fourth stage (pilot projects) was ended to apply the training evaluation models to a training program organized by one of each Italian HCOs that joined the project. We selected training programs whose learning objectives were considered as strategically relevant, namely:

- ▶ a theoretical training program on the treatment of biological agents organized by the Carlo Besta Neurological Institute;
- ▶ a practical training program on emergency ultrasound organized by the Humanitas Clinical Institute;
- ▶ a on-the-job training program on infectious disease prevention organized by the Local Health Unit of Lecco;
- ▶ two theoretical training programs on pain control, organized by the European Institute of Oncology and by the Sant'Anna University Hospital.

During the last stage (follow-up), the results of the pilot projects were presented in order to develop a plenary discussion and a comprehensive comparative analysis.

Following this methodology, we gained an access to many data and information and the opportunity to have a very "open" and "daily" confrontation with people that are expected to use PM systems. Given the "subjective" nature of action research and recognising the importance of triangulation to clear the validity of data (Eden and Huxham, 1993), we benefit by carrying out a sort of cross-sectional action research project and comparing the findings coming from different organizational

experiences. Our conclusions aims to shed some light on an “emergent” theory about the use of PM systems in the governance of training programs.

3. Theoretical framework

” Levels” of training evaluation

The literature on training has developed many contributions about training evaluation criteria. About 50 years ago, Kirkpatrick published a series of articles in which he outlined four categories (“levels”) of training effectiveness measures - i.e. reactions, learning, behaviour, and results (Kirkpatrick, 1959a, 1959b, 1960a, 1960b). As this framework has had widespread and enduring popularity, both in business and in academy, it has been further analysed and developed (Alliger *et al.*, 1997; Kirkpatrick and Kirkpatrick, 2005, 2006).

The first level measures immediate reactions of trainees towards training programs. These reaction measures are ended to evaluate the trainees’ satisfaction with respect to various elements of the training programs, such as the teaching quality, the relevance of the programs, the organization of the programs. Warr and Bunce (1995) demonstrated that three types of reactions are factorially distinct: training enjoyment, training perceived usefulness, and training perceived difficulty. Alliger *et al.* (1997) proposed an augmented framework dividing training reactions into affective and utility reactions.

Reaction measures are relevant for two reasons: firstly, positive reactions increase trainees’ motivation towards training participation and, as a consequence, learning; secondly, they allow an immediate feed-back for improving teaching and organizational aspects of training programs by comparing immediate reactions against acceptable standards.

The second level measures to what extent learning has occurred, where learning is conceived as knowledge, skill, and/or attitude. While all programs are usually ended to increase the knowledge of the participants, only some programs are ended to improve skills and/or changing attitudes too. Therefore, learning evaluation should be targeted to the specific objectives of the program in order to gather specific measures of learning improvement as a difference between measures before and after training. Alliger *et al.* (1997) proposed an augmented framework dividing learning measures into immediate post-training knowledge, knowledge retention, and behaviour/skill demonstration.

The third level measures to what extent on-the-job behaviour has changed as a result of the program. Even if this level is very difficult to be measured, it is very relevant: if the trainees do not apply what they learned, the program has been a failure even if learning occurred. Measuring behaviour change is necessary, not only to point out if behaviour has changed, but also to determine the reasons why change has not occurred at least as desired.

Finally, the fourth level measures to what extent organizational results have changed because of the training program. Results are measured in term of organizational measures, such as turnover, volume of activity, cost-cutting, or quality indicators. As mentioned earlier, even if this level is very difficult to be measured, it is very relevant: if organizational results did not improve as expected, for the organization the program has been a failure even if behavioural change and learning have occurred.

Even though this model of training evaluation was defined about fifty years ago, it is still relevant both for planning, improving, and evaluating a training program (Bartram and Gibson, 1999; Kirkpatrick and Kirkpatrick, 2005, 2006). Before training, needs assessment allows to point out expected organizational results.

During the training, the Kirkpatrick’s model allows to compare achieved measures against acceptable standards in order to identify immediate feed-back to improve receptiveness and attention (first level), learning effectiveness (second level), action plan about both individual application (third level) and organizational implementation (fourth level).

After the training, the model is very relevant to support the transfer of learning. Moreover, ex-post analysis allows to improve next editions of the same training program.

Phillips further developed the measurement of organizational results deriving from training, and developed the ROI of training (Phillips, 1996a, 1977, 2002). Where Kirkpatrick's fourth level stops at identifying the benefits of the program, Phillips' model:

- ▶ converts the benefits to monetary value and compares the monetary benefits to the fully loaded costs of the program;
- ▶ includes the consideration of intangible, but reportable, benefits of the program.

Recently, Phillips and Phillips (2002, 2004) argued the relevance of the ROI of training across various sectors, including the public sector.

The power of the models suggested by Kirkpatrick and Phillips is their simplicity and ability to help people think about training evaluation criteria. Nevertheless, some limits can be pointed out.

First, these models present three (often implicit) assumptions about their "hierarchical nature" that have been questioned in the literature (Alliger and Janak, 1989):

- ▶ the levels of evaluation are arranged in ascending value of information provided, i.e. a measure of learning provides more information than a measure of reactions, and so forth (Newstrom, 1978);
- ▶ the levels of evaluation are causally linked (Hamblin, 1974);
- ▶ the levels of evaluation are positively intercorrelated (Newstrom, 1978).

Main discussion about these assumptions include (Alliger and Janak, 1989; Alliger et al., 1997):

- ▶ it is not clear that all training in organizations is meant to effect change at all four levels;
- ▶ the hierarchical nature of these models lead to perceive fourth level measures as invariably the best ones, since they are the highest in the hierarchy;
- ▶ there is a critical issue of temporality, because levels 1 and 2 are usually administered directly after training (as well as possibly before, in the case of level 2), while levels 3 and 4 measures occur at some time after training;
- ▶ correlation among levels may be either positive, or negative, or reverse.

Second, these traditional models of training evaluation should be further developed in order to focus the training strategy and to align it to the corporate strategy, within a more comprehensive process of strategic change (Robinson and Robinson, 1998; Kirkpatrick and Kirkpatrick, 2005, 2006).

Process of training evaluation

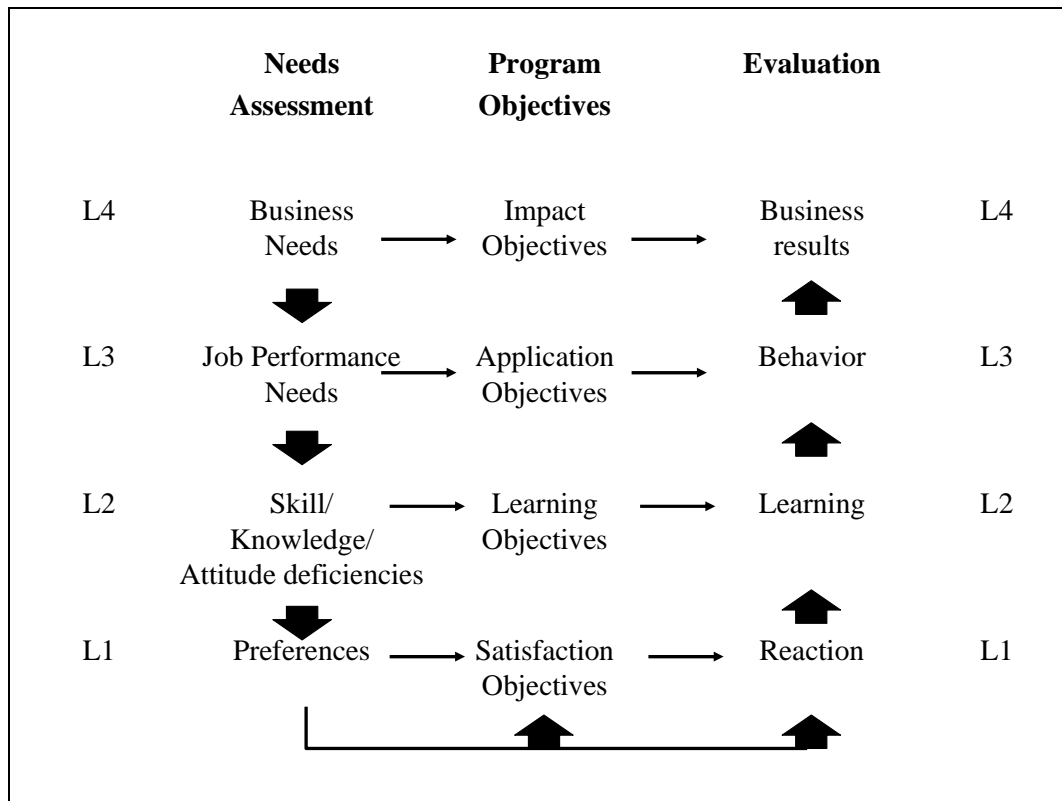
According to the literature, the process of training evaluation should include (Phillips, 1997, 2002; Kirkpatrick and Kirkpatrick, 2005, 2006):

- ▶ the integration of training program evaluation with needs assessment;
- ▶ the selection of training programs to be evaluated at different levels;
- ▶ the design of the proper training unit to be evaluated, when training programs are complex;
- ▶ the time of evaluation at different levels;
- ▶ the methods of data collection;
- ▶ the methods of data analysis, in order to isolate the effects of training from other individual and organizational factors affecting performance;
- ▶ the transfer of learning issue.

Needs assessment. In response to contemporary business requirements, it needs to link training management to the organizational strategy (Robinson and Robinson, 1998). This link is realized through needs assessment, in order to determine if training is the appropriate solution to achieve desired performances (Leibler et al., 2001; Phillips and Phillips, 2001). After the needs are determined and training is identified as the best solution, the training program objectives are developed. Each objective is then linked to a level of evaluation. Therefore, needs assessments levels and training evaluation programs are the same, but the assessment process and the evaluation process work in reverse

(see figure 1). In other words, the planning process should follow these steps: a) starting with the fourth level (organizational results) in order to determine with the line-of-business managers what needs to happen (needs assessment); b) working with the third level (behaviour) in order to identify what behaviours need to be put into practice to achieve the desired results; c) following with the second level (learning) in order to select what knowledge, skills, and attitudes will the targeted employees need to have to behave appropriately; d) concluding with the first level (reactions) in order to attract participants to come to training and be receptive to desired changes.

Figure 1 – Linking needs assessment with training evaluation (adapted from Phillips-Phillips, 2001)



Training program selection. Not all programs should be evaluated at all levels (Kirkpatrick, 1996; Phillips, 1997). According to a cost-benefit analysis, reaction measures should be developed for all training programs. Even though learning measures are more expensive, they should be collected for as many training programs as possible, at least for immediate post-training knowledge. On the contrary, behavioural and organizational impacts should be evaluated for a selection of training programs as they require much more time and resources.

Both qualitative and quantitative criteria may select training programs to be evaluated at levels 3 and 4. The most relevant qualitative criteria include the life cycle of the program, the linkage of the program to operational goals and measures, the contribution of the program to strategic objectives, the cost of the program, the visibility of the program, the size of the target audience, and the top executive interest in the program (Phillips and Phillips, 2001). Quantitative criteria are based on statistical sampling methodology, where the sample number basically depends on the following variables (Phillips, 1996a):

- ▶ the HR staff's expertise on evaluation;
- ▶ the type of the training programs being evaluated;
- ▶ the resources allocated for evaluation;
- ▶ the degree of support from management to training and development;

- ▶ the organization's commitment to measurement and evaluation;
- ▶ the amount of pressure from others to show training effectiveness measures.

Training unit. When a program is composed by many integrated courses, it is critical to design an appropriate unit of analysis. According to Phillips (1996a), when all of the courses in a series must be completed before their common objectives are met, it may be appropriate to wait to evaluate the series as a whole. The decision to evaluate several courses in a series should take into account the training goals, timing of the courses, and cohesiveness of the series. Indeed, it can be difficult to evaluate a series conducted over a long period of time, as cause-and-effect relationships become more confusing and complex.

Time of evaluation. Another critical issue in evaluating training programs at different levels concerns the timing of evaluation (Kirkpatrick, 1996). Reaction measures and immediate post-training knowledge have to be collected immediately after the courses have concluded. On the contrary, it is useful to postpone the collection of knowledge retention and sometimes attitude and skill demonstrations too. Certainly, behavioural change and organizational results have to be gathered after a proper period of time, when the transfer of learning could happen.

Methods of data collection. Moving from level 1 to levels 4 and 5 (ROI), the difficulty and the cost of data collection grow. Questionnaires are the most common method to measure reactions' trainees. Learning data can be collected in many ways according to the learning objectives, i.e. knowledge, skill, and/or attitude. The most common methods used to collect learning measures include pre-post test, post-test, performance demonstrations, peer and self-assessment, skill building exercises (Phillips, 1997).

The measurement of behavioural change and organizational results deriving from training needs to record both prior and subsequent performance (Warr *et al.*, 1999). Levels 3 and 4 data are gathered through a variety of means including: follow-up questionnaires and surveys; observations on the job; follow-up interviews with end users/performers; follow-up focus groups; assignments related to the program; action planning and improvement plans; performance contracting; follow-up sessions; performance monitoring (Phillips, 1997).

Methods of data analysis in order to isolate the effects of training. Though sometimes considered dubious (Warr *et al.*, 1999), Phillips (1996b) pointed out the relevance of isolating the effects of training program from other factors occurring during the same period that could affect organizational results. He also suggested ten methods to accomplish this task (use of control group, trend-line analysis, forecasting, participant estimation, supervisor estimation, management estimation, customer input, expert estimation, subordinate input, differential calculation from other factors).

Transfer of learning. A key issue in evaluating training programs is the transfer of learning (Ulrich, 1997b; Robison and Robinson, 1998; Phillips and Phillips, 2001). Many individual and situational factors of the work climate may either limit or support training effectiveness (Mathieu *et al.*, 1992).

Example of factors inhibiting the transfer of training are (Broad and Newstrom, 1992):

- ▶ lack of involvement by top management in the behaviour change process;
- ▶ high degree of centralization of improvement interventions in the top echelons of the organizations, resulting in little acceptance by lower-level participants;
- ▶ lack of involvement of line managers;
- ▶ unrealistic expectations from training programs (too much is expected too soon);
- ▶ lack of reinforcement on the job to support trainees in applying training to their jobs.

Therefore, a comprehensive evaluation training process has to include the analysis of factors inhibiting and/or promoting the transfer of learning. To this end, it is important to developed a holistic change management process, where training is one of the strategic drivers of change (Ulrich, 1997b; Robinson and Robinson, 1998; Kirkpatrick and Kirkpatrick, 2005). Kirkpatrick and Kirkpatrick (2005) pointed out five foundations that, at least to some degree, allow to improve behavioural change and organizational results deriving from training:

- ▶ a strategic focus, where corporate strategy, organizational cascading strategy, and training strategy are well defined and strictly aligned;
- ▶ the right kind of diffused leadership, where top management team and line managers are involved in the training process in order to contribute to the training strategy definition and to reinforce the transfer of learning on the job;
- ▶ the ability to plan for and manage change effectively, where the training process is integrated with other HR management process such as incentives and career development;
- ▶ an effective performance management system aimed at enhancing the culture of accountability;
- ▶ an effective quality of training processes aimed at increasing success with levels 1 and 2, as appropriate learning is a prerequisite to improve the transfer of learning on the job.

In conclusion, the literature on training has traditionally developed different models for evaluating training performance according to an operational approach. More recently, it is emerged the need for a strategic approach according to which training should be managed as a driver of organizational change. Consequently, it needs to point out structured models of training measurement and management, strictly integrated to the managerial processes, *in primis* to the strategic management process.

Evaluating training with Strategic Performance Measurement (SPM) systems

In the last two decades the traditional approach to PM has radically changed (Neely, 1999). A new “generation” of SPM systems (such as the Balanced Scorecard, and the Performance Prism) got through the need to: a) look at organizational performance from different perspectives (and not only in financial terms); b) point out the capability of an organization to successfully implement its own strategy; c) align individual behaviours; d) stimulate organizational learning.

Of course, human resources has been widely considered an essential driver of organizational performance. Thus the so-called HR performance naturally gained an acknowledged visibility within the frameworks used by SPM systems.

From one hand, many strategic themes related to human and organizational capital – skills, talents, knowledge, culture, leadership, alignment, teamwork, and so on – have been measured at an organizational level in the “learning and growth” perspective of the Balanced Scorecard (BSC; Kaplan, 2004) as well as in the “capabilities” of the Performance Prism (Neely et al., 2002). In this case, the results achieved through training “efforts” have been usually emphasized as a leading driver in the development of organizational competences.

On the other hand, several authors have paid attention to the performance of the HR function, actually “cascading” SPM systems inside the organization and trying to point out its contribution to strategy’s implementation. In 1997 Ulrich (Ulrich, 1997) argued that, although many talk about measuring HR, several issues remained confused (what to measure? how to measure? when to measure?) and the field of HR assessment contained more promise than delivery. Besides, assuming that “the only competitive weapon left is organization” (Ulrich, 1998), he claimed that, to fulfil its business partner role – administrative expert, employee champion, change agent, partner in strategy execution – HR managers should recognise PM as a priority and start to replace concepts with evidence, ideas with results, and perceptions with assessments. Three different models of measuring HR performance could be used (Ulrich, 1989):

- ▶ a *stakeholder* or constituency approach that focuses on the perceived value of the services offered by the HR function to the others organizational units;
- ▶ an *utility* approach that emphasizes the economic utility of the activities carried out by the HR department;
- ▶ a *relationship* approach that aims to point out the ways HR function actually contributes to organizational performance.

Ulrich's recommendations led to a further development of the stakeholder approach. Following the logic underlying the BSC, an HR scorecard has been proposed (Becker et al., 2001) and implemented (Walker and MacDonald, 2001) to effectively capture – and manage – the HR's role in the firm's success. In this kind of framework, the results achieved through training programs are typically “fragmented” between the financial perspective (i.e. cost per trainee hour, percentage of payroll spent in training), the customer perspective (percentage and number of employees involved in training, employees' satisfaction toward training programs) and the internal processes perspective (number of courses taught, number of training days and programs, percentage of new materials in training programs, percentage of training delivered, percentage of training delivered on time).

Finally, Huselid, Becker and Beatty (Huselid et al., 2005) argued that workforce performance and success is a “people” issue whose responsibility is largely of HR professionals. Assuming that firms need a business strategy, a workforce strategy, and a strategy for the HR function, the workforce scorecard has been seen as the framework to highlight the strategic performance of employees rather than the contribution of the HR department to business success. Once again, the assessment of training programs can hardly be addressed by the measures used in the different perspective of the model (workforce success, workforce behaviours, workforce competences, workforce mind-set/culture).

In our knowledge, the literature related to SPM does not provide a definitive answer to the issues that arise in measuring training programs performance. Although the frameworks for a “strategic” way of looking at performance are getting widely known and popular, they fail to help organizations to make their training strategy clear and shared and to deliver it, actually changing individual behaviours (Kirkpatrick and Kirkpatrick, 2006).

4. Evidence from healthcare organizations

This paragraph is ended to report the pilot projects developed by the four teams that were engaged during the research. For each evidence, we present the training program characteristics, the training program evaluation purposes, the methods used for measuring training effectiveness, and the achieved results.

Evaluating a training program on the treatment of biological agents

A first pilot project focused on the training program about the treatment of biological agents, organized by the Carlo Besta Neurological Institute in order to train nurses in response to a regional regulation. The evidence reported in this paper is referred to the first edition of the program, delivered on October 2006 to 20 nurses.

The purpose of the program was to improve knowledge on specific aspects of biological risk management, identified by the regional regulation, and adapted according to the specific characteristics of the Institute. A great emphasis was given to the individual and collective devices for protecting against biological agents, and to the specific treatment of refusals. Therefore, the program was basically ended to improve the application of those devices by nurses, to enhance the proper treatment of refusals, and to minimize the numbers of “biological incidents”. The program was designed as live classroom training, and was developed in various editions. Each edition took 8 hours of interactive plenary lessons, structured into two sessions (four hours per day). The director of the program was the Institute Nurses Director, who involved other 12 trainers selected within the Institute because of their organizational role and/or professional expertise.

In order to evaluate the program, we created a team composed by one external researcher and five professionals working within the Institute – the HR Director, the Nurses Director, two training professionals, one assistant. The specific purposes of the evaluation were:

- ▶ building a rigorous feed-back from the first edition of the program, in order to improve future editions;

- ▶ measuring the learning change occurred in the participants;
- ▶ evaluating the transfer of learning;
- ▶ developing a test on training evaluation, to be applied to other training programs within the Institute.

According to these purposes, we adopted the training evaluation model suggested by Kirkpatrick. Reaction measures were collected by immediate post-training questionnaires distributed both to trainees and to trainers. As the course was the first edition of a series of courses, it was very important to evaluate the quality of the training in order to improve the program during its development. Therefore, two questionnaires of trainees' reactions were designed:

- ▶ a first questionnaire, to be delivered at the end of the first session of the course (first day);
- ▶ a second questionnaire, to be delivered at the end of the second session of the course (second day).

Both questionnaires included questions about teaching enjoyment and perception of individual participation. In addition, the second questionnaire included items concerning an overall satisfaction of the course, and organizational issues.

Each trainer's immediate reaction was collected through a questionnaire ended to catch his/her satisfaction about the program design, and the development of his/her training session.

Great attention was given to the communication process, as the purpose of questionnaires was clearly communicated and anonymity of trainees was guaranteed.

The results of reaction measures reported a general satisfaction of the course by trainees (see table 1) and trainers. Some trainers indicated interesting open comments about the training program design, that resulted very useful to improve future editions.

Table 1 – The evaluation of the course on the treatment of biological agents: trainees' reaction

<i>AREAS</i>	<i>ITEMS</i>	<i>RESULTS</i>
Teaching enjoyment (1-5 scale)	Completeness of the program	3,21
	Clarity of trainers	3,55
	Discussion	2,82
	Quality of teaching aids	2,75
Participation (1-5 scale)	Perception of activation	3,00
	Perception of reflection	3,18
	Perception of knowledge improvement	2,93
Satisfaction (1-5 scale)	Relevance of the program	3,78
	Quality of the program	3,85
	Effectiveness of the program	3,78
Organization (%)	Correctness of duration	100
	Correctness of date	92,90
	Correctness of logistic	92,90

The learning change was measured through a pre-post test ended to evaluate knowledge. The same multiple-choice test was used to measure both pre-learning and post-learning. It was composed by 19 multiple-choice questions, and each question included one right answer. Pre-post test data were analysed according to the methodology of level 2 evaluation suggested on the website of World Bank Institute (WBI). The results of learning change are reported in Table 2. As all respondents answered both the pre-test and the post-test (100% response rate), we could measure the individual learning gains as differences between individual scores on the pre-test and on the post-test. Consequently, we

could also compute the average score of all respondents. The average learning gain was +18,8%, as difference between the mean score on the pre-test for all respondents (51,5%) and the mean score on the post test for all respondents (70,3%), where the mean score is the percentage of correctly answered test items.

Table 2 – The evaluation of the course on the treatment of biological agents: trainees’ knowledge

	<i>PRE-TEST</i>	<i>POST-TEST</i>
Mean Test Score (% correct)	51,5	70,3
Median Test Score (% correct)	50,0	71,1
Standard Deviation	9,9	7,3

Learning data were further analysed in order to evaluate the validity of the learning test. The team analysed:

- ▶ the difficulty of each item composing the test, measured by a difficulty index indicating the percentage of respondents who answered the item correctly;
- ▶ the conformity of each item with the rest of the items of the test, measured by a discrimination index computing the correlation between the score on the item and the score on all other items in the test; if an individual item was in good conformity with the rest of items in the test, its discrimination index should be a high positive number - in other words, if a respondent had performed well in that item, it was expected that she/he would perform well in other items in the test; items reporting low or negative discrimination indexes should be reviewed;
- ▶ the validity of the multiple-choice options of each item of the test, pointed out by the frequency of selection of wrong options with respect to that of the right one.

The analysis of the validity of the learning test resulted very useful to support the improvement of the learning test to be used in future editions.

Finally, learning data allowed to compare individual (anonymous) trainees’ learning change. As individual trainees’ learning change resulted quite differentiated, it emerged the need to put more attention to trainees’ selection and class composition for further editions.

The change behaviour evaluation was measured through two follow-up structured surveys to all trainees and their supervisors. These surveys were conducted one month after the end of the training. Each trainee was invited to indicate his/her perceived degree of application of a list of specific behaviours about the prevention of biological risk before the training (according to the scale “never, sometimes, often, always”). For those behaviours which reported “never” or “sometimes” before the training, each trainee was also invited to write if he/she has or not started to apply these behaviours after the training. Finally, each trainee was invited to indicate the factors inhibiting or supporting expected behavioural change.

Supervisors were formally invited to observe trainees before and after the training, in order to evaluate changes in the application of the same list of behavioural rules about the prevention of biological risks that were asked to trainees. Moreover, supervisors were also invited to indicate factors inhibiting or supporting trainees’ behavioural change.

Great attention was given to the communication process: the purpose of surveys was clearly communicated; anonymity was guaranteed; involvement was emphasized.

The response rates were 85% for trainees and 90% for their supervisor. The results of this analysis showed a great correspondence between trainees’ perception and their supervisors observation in judging those behaviours that were already appropriate before training and those that were not. More interestingly, the analysis offered also a good understanding about the factors affecting the transfer of learning.

Finally, the impact of training upon organizational results was analysed through a follow-up focused group. It was planned two months after the end of the course, and it took 90 minutes. According to what resulted from the behavioural change analysis, the focus group was ended to solicit qualitative judgments about the issue of transfer of learning. Participants were selected as representative sample of the target population (namely trainees, direct supervisors, top management team, clinical risk manager, HR manager). They were invited to the focus group through a letter explaining the purpose of the meeting, detailing the results derived from the other levels of analysis, and indicating the main issues to be discussed. Two researchers attended the focus group, one as an external facilitator and one as an observer. A report of the focus group was written and sent to participants.

The focus group allowed to better analyse some factors inhibiting the transfer of learning emerged in the previous surveys and, more interestingly, could also identify some shared solutions to solve critical issues and to reinforce supportive elements.

Evaluating a training program on emergency ultrasound

A second pilot project focused on a professional training program on emergency ultrasounds, organized by Humanitas Institute, called E-FAST (Extended Focused Assessment by Sonography for Trauma), as a prototype of point-of-care ultrasonography to be used in multiple injured trauma patients.

The course, usually organized by Humanitas, addressed to emergency physicians, and takes 9 hours. The possibility to rapidly learn how to use E-FAST as a point-of-care standard is due to the easiness of the contents, the possibility to develop practical simulation on pathological models, the availability of algorithms for trauma patients, and the short learning curve.

The main purposes of this training program are to develop both knowledge about basic concepts on E-FAST in multiple injured trauma patients, and skills about the use of E-FAST for the detection of both abdominal and thorax trauma. The course may also impact both professional and organizational improvements, if trainees are interested in transferring the learning on E-FAST on their job.

During the last years, 10 editions of the course have been delivered firstly to emergency physicians working in Humanitas, and then to external physicians. Future editions are planned.

We created a team composed by two external researchers and four professionals working in the Institute – the training department director and one of his staff as trainers experts, the emergency department director and one of his collaborators as clinical experts. The specific purposes of the evaluation of this training program were:

- ▶ improving learning evaluation tools in order to give the classroom a specific feed-back on acquired learning at the end of each sessions of the training program, and to design a learning evaluation methodology ended to certificate competences acquired at the end of the training program by single participants;
- ▶ designing a systematic evaluation methodology of the effectiveness of the training program, in order to support its continuous improvement;
- ▶ designing specific tools ended to collect data on the job in order to increase the transfer of learning within the HCOs whose professionals attended the program;
- ▶ developing a research pilot project on training program evaluation, in view of a more systematic application to training programs delivered within the Institute.

According to these purposes, we adopted the training evaluation model suggested by Kirkpatrick. Even though the evaluation design was completed for the four levels, retrospective data are available only for level 1 (reactions) and 2 (learning of knowledge) collected for two editions of the course delivered on March and September 2006.

The design of level 1 included both trainees' and trainers' immediate reactions measured through questionnaires. Trainees' evaluation design included both enjoyment and usefulness criteria, measured with multiple-choice questions, two-way questions, and open-ended questions. Each trainer's

immediate reaction was collected through a questionnaire ended to catch his/her satisfaction about the program design, the organization of the course, and the development of his/her specific training session. Both questionnaires included an introduction explaining the importance of reaction measures, and offering the possibility to receive a consolidated report on reactions. Anonymity of trainees was guaranteed.

Retrospective data on trainees' reactions were collected through a short reaction form - needed for the National Program on Continuous Medical Education promoted by the Italian Ministry of Health - and reported a very high satisfaction (see Table 3).

Table 3 – The evaluation of the course on E-FAST: retrospective data about trainees' reaction (1-5 scale)

<i>ITEMS</i>	<i>EDITION MARCH 2006 (N=24)</i>	<i>EDITION SEPTEMBER 2006 (N=20)</i>
Relevance	4,54	4,63
Quality	4,17	4,63
Effectiveness	4,37	4,68

Taking into account the learning purposes of the program, the design of the second level evaluation included three tools focused both on knowledge and on skill evaluation. Knowledge was evaluated through a pre-post-test, composed by multiple choice questions, so that for each trainee it was possible to compute a learning gain measured as difference between the mean score on the pre-test and the mean score on the post-test, where the mean score is the percentage of correctly answered test items. Retrospective data about knowledge are reported in Table 4, where a 100% response rate is available for each edition.

A job simulation and a video quiz were used as tools of skill evaluation. During the job simulation, each trainee was asked to apply the procedure he/she has learnt (i.e. the set-up of the proper functions of the equipment; the proper execution of the ultrasound scans; the proper interpretation of the E-FAST anatomy). Trainers observed the application of these activities by each trainee, and developed a written evaluation based on a Likert scale.

Table 4 – The evaluation of the course on E-FAST: retrospective data about trainees' knowledge (1-5 scale)

<i>EDITION MARCH 2006 (N=24)</i>	<i>PRE-TEST</i>	<i>POST-TEST</i>
Mean Test Score (% correct)	69,8	95,4
Median Test Score (% correct)	75,0	100,0
Standard Deviation	20,8	5,9
<i>Edition September 2006 (n=20)</i>	<i>Pre-test</i>	<i>Post-test</i>
Mean Test Score (% correct)	71,9	98,5
Median Test Score (% correct)	75,0	100,0
Standard Deviation	16,7	4,9

During the video quiz, each trainee had to fill a multiple choice questionnaire, ended to recognize a proper diagnosis after having observed some ultrasound scans.

The design of change behaviour evaluation was based on monitoring performance data. According to this design, participants are asked to apply E-FAST for 3-6 months after the course, and to register

some clinical data within a format whose validity should be formally certificated by their supervisors. Trainers may periodically review these data in order to evaluate the application on the job of knowledge and skills on E-FAST learnt by single professionals, activating a process of continuous education on the job.

Finally, the design of the last level of training evaluation (organizational results) included a focus group ended to identify qualitative elements that are critical in systematically applying E-FAST within a given organization, focusing the analysis on the individual and organizational factors inhibiting and/or supporting the transfer of learning.

Evaluation criteria were designed to be used both for internal and external applications. Firstly, levels 1 and 2 measures may be immediately used to support a continuous improvement of the training program. At the end of each edition it was planned to organize a de-briefing meeting, in which the scientific director of the training course solicits a genuine discussions among teachers.

Secondly, learning evaluation may be also used collectively, in order to identify the most common mistakes done by participants, and to reinforce some learning issues at the end of each sessions of the training program.

Learning evaluation was also planned to be used as a tool to certificate both the knowledge and the skills reached at the end of the course by single participants. Each trainee is also informed about his/her knowledge gain, the evaluation of his/her simulation, and the evaluation obtained with the video quiz. Such information may also be used to suggest specific assignments to further support individual learning.

Level 3 is used to monitor the knowledge/skill application on the job by each trainee in order to support the continuous education on the job of individual professional competences.

Finally, levels 3 and 4 were planned to be used in order to support the transfer of learning to other hospital emergency departments, if needed. In particular, these measures may support either a more efficient and effective use of the actual organization of an emergency department or an organizational re-engineering, through a change management process aiming at solving factors inhibiting the application of E-Fast and/or strengthening supportive elements.

Evaluating a training program on infectious disease prevention

The third pilot project focused on the evaluation process of an on-the-job training program delivered within the Local Health Unit of Lecco in 2006, ended to define an organizational protocol on infectious disease prevention and to support its application within the organization, in congruence with the regulation established by Lombardy Region about the prevention of infectious disease.

The purposes of this training program were:

- ▶ to improve knowledge about the regional regulation on the prevention of infectious disease;
- ▶ to share a common definition of an organizational protocol on infectious disease prevention, consistent with the regional regulation; the protocol focused on epidemiological survey, as a tool of prevention ended to prompt recognize the sources of the infection and to act for containing the diffusion of the infection;
- ▶ to support a complete and homogeneous diffusion of this protocol within the three social and health districts.

The course was developed as an improvement group, and was recognized by the program of Continuous Medical Education promoted by Lombardy Region. According to this program, an improvement group is a specific methodology of on-the-job training, ended to activate a team engaged in improving a situation that is strategically relevant and highly complex. The program was structured in four meetings with nine participants (both physicians and nurses). An expert facilitator joined the program, with the role of promoting a shared definition of the specific objectives of the training program, a proper selection of the methodology, and a continuous support to participants.

The evaluation of the training program was ended to:

- ▶ build a rigorous feed-back from the application of the improvement group methodology;
- ▶ evaluate the knowledge of participants about the regional regulation on infectious disease prevention, as a pre-requisite to define a proper organizational protocol and to apply it;
- ▶ evaluate the congruence of the organizational protocol with respect to regional regulation;
- ▶ evaluate the transfer of learning about the application of the new protocol within the three social and health districts, in order to support a continuous improvement of the activity of infectious disease prevention;
- ▶ develop a research pilot, in order to test a possible methodology of training evaluation to be more systematically applied within the Local Health Unit.

According to these purposes, we created a team composed by the training department director, a nurse expert on the specific content area, and an external researcher. We adopted the training evaluation model suggested by Kirkpatrick.

Reactions of both participants and facilitator were collected at the end of the program through reaction questionnaires.

Taking into account the characteristics of the improvement group, both trainees' and trainer's reaction evaluations were focused on a critical review of: the perceived achievement of training purposes, the perceived correctness of the program development, the perceived usefulness of the program. The questionnaires included an deep introduction explaining the importance of reaction measures, and offering the trainees the possibility to receive a consolidated report on reactions. Anonymity of trainees was guaranteed.

Both participants and facilitator were generally very satisfied about the training program (see Table 5). This evaluation was very useful to define some methodological improvements, as the duration of the program (that was considered too much short).

Table 5 – The evaluation of the course on infectious disease prevention: trainees' and facilitator's reaction (1-4 scale)

<i>ITEMS</i>	<i>TRAINEES' REACTION (N=9)</i>	<i>FACILITATOR'S REACTION (N=1)</i>
Achievement of training purposes	3,00	2,00
Completeness of contents	2,89	2,00
Engagement of trainees	3,33	3,00
Role of facilitator	3,33	-
Perception of knowledge improvement	3,11	3,00
Perception of behaviour improvement	3,11	3,00
Usefulness of the program	3,11	3,00

Participants' knowledge about the regional regulation on infectious disease prevention was evaluated through a post-test, requiring a minimum percentage of correct answers. All participants passed this evaluation. We decided not to use a pre-test as participants did not know the specific regional regulation on infectious disease prevention at the beginning of the program.

The evaluation of levels 3 and 4 were measured in two ways. First, we could verify the availability, at the end of the program, of the organizational protocol on infectious disease prevention, congruent with the regional regulation, shared by participants, and formally approved by the chief of the prevention organizational unit.

Second, it was planned to analyse the degree of application of the organizational protocol through an epidemiological survey, to be conducted according to proper statistical sampling criteria within the three social and health districts. Participants shared a check-list detailing the main activities required

by the organizational protocol to be checked through the epidemiological survey. This analysis was differed to twelve months after the end of the training programs.

Evaluating two training programs on pain control

Following the guidelines of the Italian Ministry of Health, in recent years the EIO and S. Anna have been engaged in a national project called “Ospedale senza dolore” (hospital without pain), basically aimed to make the clinical staff more active in monitoring, taking care, and reducing patients’ pain. Training, of course, is critical for the project’s success. All the clinical staff (mainly physicians and nurses) has to be involved with a training program whose objective is to improve knowledge about: a) how to recognise and evaluate the pain’s symptoms; b) the treatments (both pharmaceutical and not) for acute and chronicle pain; c) all the aspects – psychological, ethical, juridical, communicational, and so on - of the pain’s experience. Both the EIO and the S. Anna hospital organized a similar training program for their clinical staff. The program, based on 2-3 days of training, is very challenging because of the huge amount of time requested to physicians and nurses who are “detached” from their usual clinical activity. Just to mention some figures, in the case of S. Anna the whole program (involving more than 350 physicians and 1.300 nurses) need about 4.000 training days to be completed.

Given its relevance, the EIO and S. Anna decided to use the BSC to evaluate the results achieved through the “hospital without pain” program. We created a mixed team (EIO, S. Anna) committed to:

- ▶ design a BSC able to provide a comprehensive picture of the results achieved through training initiatives;
- ▶ implement the BSC both at the EIO and S. Anna with the real data of the “hospital without pain program”;
- ▶ evaluate the opportunity to use the BSC in measuring and managing the “training” performance at an organizational level.

After having shared the main findings coming from the existing literature about the use of the BSC in “educational” contexts (i.e. universities; Haddad, 1999) and in the management of projects (Norrie-Walker, 2004; Steward, 2001), the team got through several meetings of brainstorming to gather ideas and suggestions about the “shape” the BSC should have in this particular case. Eventually we were able to drawn a sort of strategy map (see figure 2) and identify the following perspective of the BSC for evaluating training initiatives’ performance:

- ▶ *design* perspective (are we planning the right initiatives?);
- ▶ *delivery* perspective (how are we delivering the initiatives?);
- ▶ *learning* perspective (are the initiatives generating a real learning among participants?);
- ▶ *performance* perspective (does the whole organization actually benefit of training initiatives?).

Table 6 shows in more details the meaning of the key-performance areas identified by the team.

About 30 key-performance indicators (KPI) were then selected in order to measure the results achieved in the different KPAs. Sharing the same structure of the BSC (in terms of perspectives and KPAs), the EIO and S. Anna made different (but, after all, rather similar) decisions about the indicators to be used and their relative weight within the BSC. The KPIs have been then calculated with the actual data related to the already completed initiatives. A software application was used to collect and “feed” the BSC (see figure 3).

Figure 2 – EIO - S. Anna strategy map

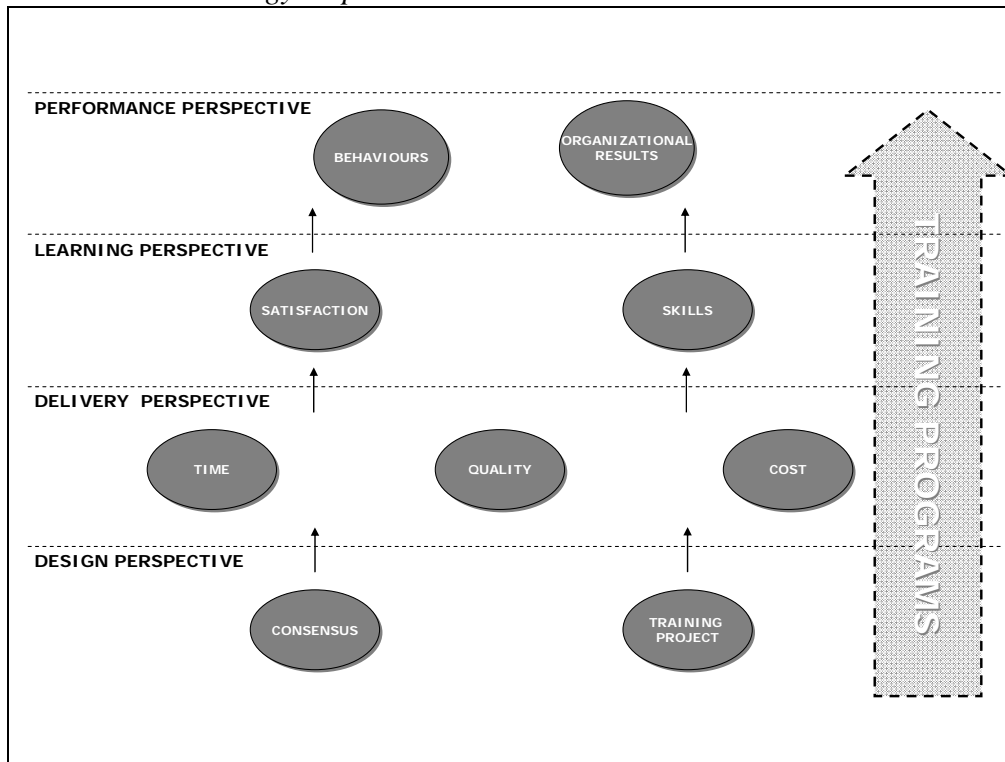
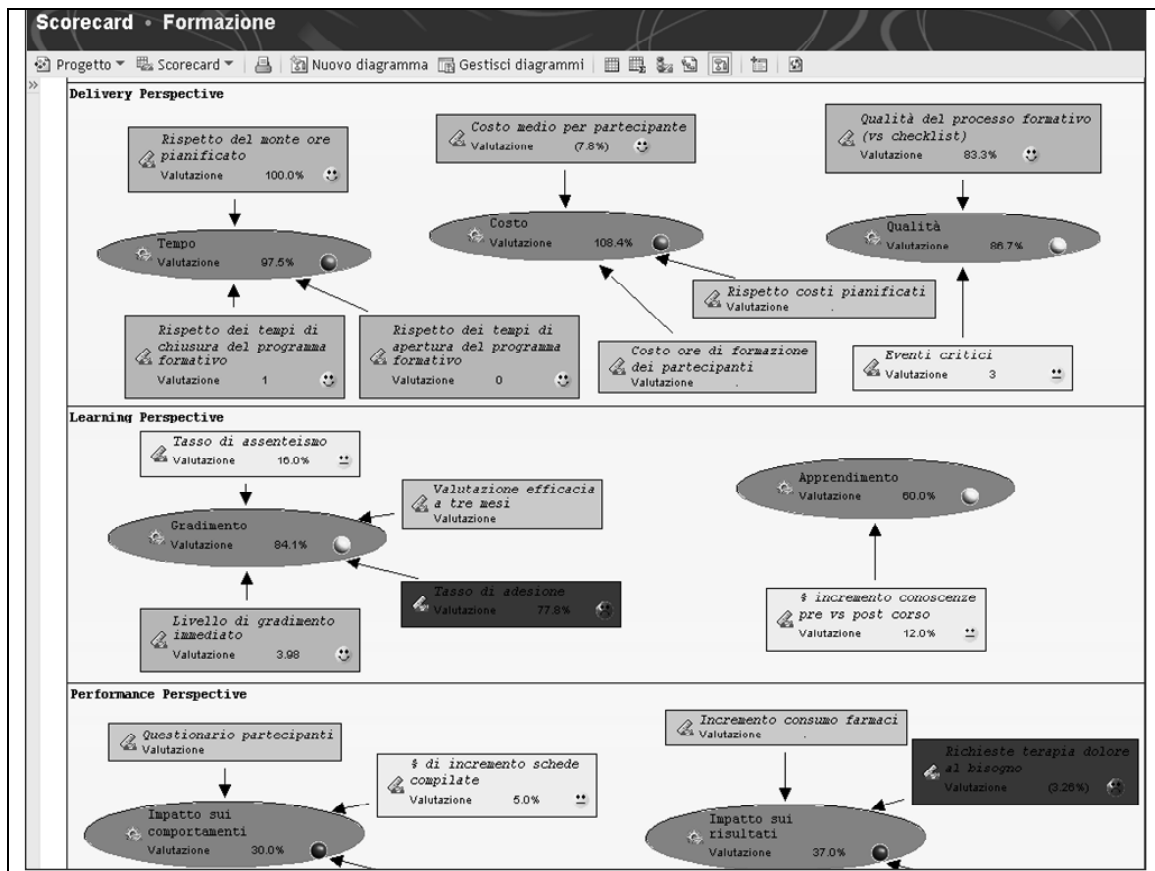


Table 6 – EIO - S. Anna key-performance areas

Design perspective	Consensus	To create an adequate level of consensus towards training initiatives among the main stakeholders.
	Training project	To plan carefully the training initiatives (fit <i>versus</i> organization-wide objective, training need evaluation, training objectives, participants' selection, partners' selection, analytical program, and so on)
Delivery perspective	Time	To deliver training initiatives with respect of the timing previously defined.
	Quality	To deliver training initiatives with respect of the qualitative standards previously defined.
	Cost	To deliver training initiatives with respect of the limitations on costs previously defined.
Learning perspective	Satisfaction	To get a high level of satisfaction by participants in the training initiatives.
	Skills	To improve, through training initiatives, the knowledge and the skills of participants.
Performance Perspective	Behaviours	To influence, through training initiatives, the actual behaviours of participants.
	Organizational results	To improve, through training initiatives, the organizational performance.

Figure 3 – EIO- S. Anna Balanced Scorecard



In general terms, the BSC emphasized:

- ▶ how difficult is the transfer of knowledge in individual behaviours and organizational performance; on average, through their training programs, both the EIO and S. Anna hospital seem to meet their objectives in the design and delivery perspectives but often fail to get the expected improvements in the performance perspective of the BSC;
- ▶ the initiatives that turned out effective in the performance perspective usually based their “success” on a good design; even if they are not a sufficient condition, a broad consensus and a careful planning of training programs seem to be, in the case of the EIO and S. Anna, the basic “ingredients” for their success.

The main lessons learnt applying the BSC methodology to the “hospital without pain” initiatives are the following:

- ▶ “customized” in its structure, the BSC can actually support performance management in the training area; before the implementation of the BSC, this area of performance, despite its relevance, was essentially managed through subjective evaluations and informal judgements;
- ▶ to deliver a substantial benefit the BSC has to be used intensively for all the training programs that: a) have a direct impact on the organizational strategy; b) absorb a relevant amount of resources; both the EIO and S. Anna are now going to implement the BSC in a more extensive way as regard as the training area;
- ▶ designing the BSC to make it a framework able to effectively represent training performance was not time consuming; the members of the team, on the contrary, found the design phase a good opportunity to share ideas and to make their objectives clear;
- ▶ the implementation of the BSC created a sort of common language among the different stakeholder involved with training initiatives;

- ▶ the use of information technology played a critical role in reducing the cost of the BSC and, as a consequence, improving the cost-benefit balance perceived by the team.

5. Discussion and conclusion

The literature on training has traditionally developed many operational tools for evaluating training programs effectiveness for both trainees and their organizations. Our action research project shows the relevance of the evaluation of different typologies of training programs delivered within HCOs (theoretical, practical, and on-the-job training programs). According to participants coming from the five Italian HCOs that developed the four pilot projects, training program evaluation tools are not only applicable, but also very useful, as they allow to:

- ▶ improve the training process quality after the program is finished, when the measures of training program effectiveness (from level 1 to level 4) are used to improve future courses;
- ▶ improve the training process quality during the development of the program, for instance when the measures of level 2 (learning) are used to give specific learning feed-back to the class;
- ▶ increase the integration between training professionals (whose organizational role is to support a rigorous training process) and experts clinicians (whose organizational role is to design the training program according to scientific evidence and to act as trainers experts of the specific content areas), when mixed teams are created in order to plan and evaluate training programs;
- ▶ strengthen the issue of transfer of learning, as measures of levels 3 and 4 allow to point out factors supporting and/or inhibiting the application of what participants learnt on their job within the organization that promoted the program;
- ▶ measure the achievement of specific results expected from training, when the training program evaluation is integrated with the training programs design through needs assessment.

However, if participants coming from different HCOs, even though levels 1 and 2 are immediately measurable during the training program delivery, the evaluation process related to levels 3 and 4 requires a strong commitment by the HCOs where participants work.

Participants also shared the benefits of systematically adopting formal methods of training evaluation assessment. The availability of formal data on training program evaluation allows to act an internal benchmarking aimed to:

- ▶ compare different editions of the same training program, and thus allowing to give specific feed-back to external HCOs that eventually committed the program;
- ▶ compare different training programs, and thus allowing to improve new courses.

Nevertheless, according to a cost-benefit analysis, participants pointed out the need of selecting training programs to be evaluated at different evaluation levels (Phillips, 1997; Kirkpatrick – Kirkpatrick, 2006). In particular, participants agreed that it needs to evaluate from level 1 to level 4 the training programs aimed to support strategic purposes.

Our research project also shows some limits of a traditional approach to training evaluation. The three pilot projects of Carlo Besta Neurological Institute, Humanitas Clinical Institute, and Lecco Local Health Unit suggest that an operational approach to training evaluation does not provide a focused framework for both defining the training program strategy and supporting its realization according to a strategic control paradigm. The lack of a strategic planning of analytical results expected from the training program does not allow to define and communicate a clear focus and alignment about the role of training within a specific project of organizational improvement, according to a participative process that should involve all key people. In particular, the traditional models of training evaluation do not allow to clearly define a strategic focus about the cause-and-effect relationships among the key performance areas of a training program.

On the contrary, participants involved in our research agreed that, once the strategic objectives of the training program are clearly specified, it is possible to align the various organizational stakeholders of the program in order to support and reinforce its implementation. The experiences of the European Institute of Oncology and of S. Anna University Hospital show that the adoption of BSC allows to better focus the training program strategy and consequently support its realization. Indeed, once the organizational purposes of the training program are clearly defined and shared among key people, the formalization of the specific key performance areas of the program allows to measure its success and support its development.

Therefore, our analysis may conclude the relevance of further researching possible links between the training literature and the PM literature, in order to support an “emergent theory” about strategic evaluation models of training programs. The adoption of BSC as a model of strategic control offers a focused framework for designing and implementing the training strategy.

Among the others, possible future research topics regard:

- ▶ the application of BSC not only to single training programs, but also to the overall training strategy (where a single training program is a lever to reach strategic training purposes);
- ▶ the process of training program evaluation, such as the stages of training program evaluation, the actors to be involved – scientific director of the training program, training manager, trainers, trainees’ supervisors, etc;
- ▶ the integration of the training program evaluation process with other managerial processes, *in primis* strategic control management and HR management;
- ▶ the change management process of the training program evaluation development - for instance the diffusion of operational tools of training program evaluation could constitute the first step of an incremental process aimed to develop an accountability-oriented culture about training investments;
- ▶ finally, the evolution of the training manager’s role from a training experts to a performance analyst, in charge for the proper integration of various performance interventions - i.e. structured training programs with other forms of learning and non-learning interventions.

References

- Alliger, G.M. and Janak, E.A., (1989), "Kirkpatrick's Levels of Training Criteria: Thirty Years Later", *Personnel Psychology*, Vol. 42, pp.331-342
- Alliger, G.M., Tannenbaum, S.I., Bennet, W., Trave, H., and Shotland, A., (1997), "A Meta-Analysis of the Relations Among Training Criteria", *Personnel Psychology*, Vol. 50, pp.341-358
- Argyris, C. and Schön, D.A., (1974), *Theories in practice*, Jossey-Bass, San Francisco
- Bartel, A.P., (2000), "Measuring the Employer's Return on Investments in Training: Evidence from the Literature", *Industrial Relations*, Vol. 39 No. 3, pp.502-524
- Bartram, S. and Gibson, B., (1999), *Evaluating Training*, Gower, Hampshire
- Bate, P., (2000), "Synthesizing Research and Practice: Using the Action Research Approach in Health Care Settings", *Social Policy and Administration*, Vol. 34 No. 4, pp. 478-493
- Becker, B.E., Huselid, M.A., Ulrich, D., (2001), *The HR Scorecard*, Harvard Business School Press, Boston
- Broad, M.L., and Newstrom, J.W., (1992), *Transfer of Training*, Addison-Wesley Publishing, New York, NY
- Coghlan, D., (2004), "Action Research in the Academy: Why and Whiter? Reflections on the Changing Nature of Research", *The Irish Journal of Management*, Vol. 25 No. 2, pp. 1-10
- Eden, C. and Huxham, C., (1996), "Action Research for Management Research", *British Journal of Management*, Vol. 7, pp. 75-86
- French, W.L. and Bell, C.H., (1984), *Organization Development: Behavioural Science Interventions for Organization Improvement*, Prentice Hall, New Jersey
- Gummesson, E., (1991), *Qualitative methods in management research*, Sage Publication, London
- Jonsson, S. and Lukka, K., (2006), "There and Back Again. Doing Interventionist Research in Management Accounting", in Chapman, C., Hopwood, A. and Shields, M.D. (Eds), *Handbook of management accounting research*, Elsevier, Oxford, pp. 373-398
- Haddad, K., (1999), "Using the Balanced Scorecard for Improving Finance Education", *Journal of Financial Practice and Education*, Spring-Summer, pp. 92-101
- Hamblin, A.C., (1974), *Evaluation and Control of Training*, McGraw-Hill, New- York, NY
- Huselid, M.A., Becker, B.E., and Beatty, R.W., (2005), *The Workforce Scorecard*, Harvard Business School Press, Boston
- Kaplan, R.S., (1998), "Innovation Action Research: Creating New Management Theory and Practice", *Journal of Management Accounting Research*, Vol. 10, pp. 89-118
- Kaplan, R.S. and Norton, D.P., (1996), *The Balanced Scorecard*, Harvard Business School Press, Boston
- Kaplan, R.S. and Norton, D.P., (2004), *Strategy Maps*, Harvard Business School Press, Boston
- Kasanen, E., Lukka, K., and Siitonen, A., (1993), "The Constructive Approach in Management Accounting Research", *Journal of Management Accounting Research*, Vol. 5, pp. 243- 264
- Kirkpatrick, D.L., (1959a), "Techniques for Evaluating Training Programs", *Journal of ASTD*, Vol. 13 No. 11, pp. 3-9
- Kirkpatrick, D.L., (1959b), "Techniques for Evaluating Training Programs: Part 2 – Learning", *Journal of ASTD*, Vol. 13 No. 12, pp.21-26
- Kirkpatrick, D.L., (1960a), "Techniques for Evaluating Training Programs: Part 3 – Behavior", *Journal of ASTD*, Vol. 14 No. 1, pp.13-18
- Kirkpatrick, D.L., (1960b), "Techniques for Evaluating Training Programs: Part 4 – Results", *Journal of ASTD*, Vol. 14 No. 2, pp.28-32
- Kirkpatrick, D.L. and Kirkpatrick, J.D., (2005), *Transferring Learning to Behavior*, Berrett-Koehler, San Francisco

- Kirkpatrick, D.L. and Kirkpatrick, J.D., (2006), *Evaluating Training Programs: The four Levels*, Berrett-Koehler, San Francisco
- Leibler, S.N., Parkman, A.W., and VanKampen, K., (2001), "How to Conduct a Needs Assessment that Gets Results", in L.L.Ukens (Ed), *What Smart Trainers Know*, Jossey-Bass Pfeiffer, San Francisco, pp.73-81
- Lewin, K., (1946), "Action Research and Minority Problems", *Journal of Social Issues*, Vol. 4, pp.34-46
- Lewin, K., (1947), "Frontiers in Group Dynamics", *Human Relations*, Vol.1, pp.143-153
- Mathieu, J.E., Tannenbaum, S.I., and Salas, E., (1992), "Influences of Individual and Situational Characteristics on Measures of Training Effectiveness", *Academy of Management Journal*, Vol. 33 No. 4, pp.828-847
- Neely, A., "The Performance Measurement Revolution: Why now and what next?", *International Journal of Operation & Production Management*, Vol. 19 No. 2, pp. 205-228
- Neely, A., Adams, C., and Kennerley, M., (2002), *The Performance Prism: The Scorecard for Measuring and Managing Business Success*, Prentice Hall, London
- Newstrom, J.W. (1978), "Catch-22: The Problems of incomplete Evaluation of Training", *Training and Development Journal*, Vol. 32 No.11, pp.22-24
- Norrie, J. and Walker, D.H., (2004), "A Balanced Scorecard Approach to Project Management", *Project Management Journal*, December, pp. 47-56
- Phillips, J.J., (1977), *Return on Investment in Training and Performance Improvement Programs*, Butterworth-Heinemann, Boston
- Phillips, J.J., (1996a), "ROI: the Search for Best Practices", *Training and Development*, February, pp.42-47
- Phillips, J.J., (1996b), "Was it the Training?", *Training and Development*, March, pp.28-32
- Phillips, J.J., (1996c), "How Much is the Training Worth", *Training and Development*, April, pp.20-24
- Phillips, J.J., (1997), *Handbook of Training Evaluation and Measurement Methods*, Butterworth-Heinemann, Boston
- Phillips, J.J. and Phillips, P.P., (2001), "Symposium on the Evaluation of Training", *International Journal of Training and Development*, Vol. 5 No. 4, pp. 240-247
- Phillips, P.P., (2002), *Understanding the Basics of Return on Investment in Training*, Kogan-Page, London
- Phillips, J.J. and Phillips, P.P., (2002), *Measuring ROI in the Public Sector*, ASTD Press, Alexandria, VA
- Phillips, J.J. and Phillips, P.P., (2004), "ROI in the Public Sector: Myths and Realities", *Public Personnel Management*, Vol. 33 No. 2, pp.139-149
- Robinson, D.G. and Robinson, J.C., (1998), *Moving from Training to Performance*, Berret-Koehler Publishers, San Francisco
- Schein, E.H., (1993), "Legitimizing Clinical Research in the Study of Organizational Culture", *Journal of Counseling & Development*, Vol. 71, pp. 703-708
- Stewart, W.E, (2001), "Balanced Scorecard for Projects", *Project Management Journal*, March, pp. 38-52
- Tian, J., Atkinson, N.L., Portnoy, B., and Gold, R.S., (2007), "A systematic review of evaluation in formal continuing medical education", *Journal of Continuing Education in the Health Professions*, Vol. 27 No. 1, pp.16-27
- Ulrich, D., (1989), "Assessing Human Resource Effectiveness: Stakeholder, Utility, and Relationship Approaches", *Human Resource Planning*, Vol. 12 No. 4, pp.301-316
- Ulrich D., (1997), "Measuring Human Resources: An Overview of Practice and a Prescription for Results", *Human Resource Management*, Vol.36 No.3, pp. 303-320

- Ulrich, D., (1997), *Human Resource Champions*, Harvard Business School Press, Boston
- Ulrich, D., (1998), "A new Mandate for Human Resources", *Harvard Business Review*, Vol.76 No.1, pp.124-134
- Ulrich, D., (2005), *The HR Value Proposition*, Harvard Business School Press, Boston
- Walker, G. and MacDonald, J.R., (2001), "Designing and Implementing an HR Scorecard", *Human Resource Management*, Vol. 40 No. 4, pp. 365-377
- Warr, P.B. and Bunce, D., (1995), "Trainee Characteristics and the Outcomes of Open Learning", *Personnel Psychology*, Vol. 48, pp.347-375
- Warr, P.B., Allan, C. and Birdi, K., (1999), "Predicting Three Levels of Training Outcome", *Journal of Occupational and Organizational Psychology*, Vol. 72, pp.351-375
- Westbrook, R., (1995), "Action Research: A New Paradigm for Research in Production and Operations Management", *International Journal of Operations & Production Management*, Vol. 15 No. 12, pp.6-20