Literature review on assessment models of the quality of training devices

Raja Lotfi [a]*, Department: Management Techniques, Higher School of Technology (EST), Hassan II University Casablanca, 20020, Casablanca, Morocco.

Touria Neggady Alami [b], Department: Management Techniques, Higher School of Technology (EST), Hassan II University Casablanca, 20020, Casablanca, Morocco.

Abstract

Assessing the quality of training devices is an ongoing and relevant research topic. It's now more than ever an obvious necessity. However, the literature review on the issue remains disparate and fragmented according to themes and across countries. Our research had a threefold purpose: 1) to propose a review of the literature on assessment models of the quality of training devices; 2) identify assessment models that are most often used; 3) to determine their key issues, strengths and limitations. Our methodological approach was based on two types of data: (1) a set of empirical documents published over the last thirty years (guides and reports); and (2) an analysis of recent studies produced by specialized research centers. Thus, we conducted three successive steps: 1) selection of documents on the basis of criteria such as relevance and consistency; 2) extensive analysis of assessment models, using a validated analysis matrix that consists of four criteria: goals, issues, strengths and limitations; 3) categorization of these models by transversal thematic analysis. At least 21 models have been generated. All have more or less different characteristics. These models have been listed in a thematic classification into four distinct categories. Beyond this classification, we retain two main conclusions. First, no model of them is exhaustive. Each model covers one or at most three dimensions of quality of training device, and thus has limits. Secondly, the quality is complex and plural. Therefore, two approaches are recommended: conceptual (systemic) and methodological (total triangulation). The first allows a combination of several assessment models to cover up dimensions of a training device. The second allows multiple crossing between several actors (assessors), collection instruments and data analysis tools to better describe the quality of training.

Keywords: Assessment models; quality; training device.

---

* ADDRESS FOR CORRESPONDENCE: **Raja Lotfi,** Department: Management Techniques, Higher School of Technology (EST), Hassan II University Casablanca, 20020, Casablanca, Morocco.
E-mail address: rajalotfi@gmail.com / Tel.: 06 14 000 400

1. Introduction

Assessment of the quality of training devices is now more than ever an obvious necessity. No institution (faculty, university, training center, etc.) can be content with "training without evaluation". From the moment when training is expensive in terms of financial, human and time resources, it is important to evaluate in order to evolve (continuous improvement), to account for the use of these resources in order to be sure of making them profitable.

Assessment of the quality of training devices is an ongoing research topic. However, the literature review on the issue is still disparate and fragmented according to themes and across countries. Thus, behind the polysemy of the concepts evaluation, quality and training device, a wide variety of approaches and models, carried out by different actors, is implemented in contexts with different stakes and on the basis of frameworks of reference often not very stable and little shared.

2. Purpose

Our research had a threefold purpose: 1) to propose a retrospective synthesis (review of the literature) on assessment models of the quality of training devices; 2) identify assessment models that are most often used; and 3) to determine their key issues, strengths and limitations.

3. Methodology

Our methodological approach was based on two types of data: written document analysis and bibliographic research.

3.1. Written documentation

In order to document etymologically the concepts of "model" and "quality of training", several documentary sources have been used (empirical, explanatory and normative texts from periodicals and monographs published over the last thirty years). This type of documentation provides scientific information on the theoretical research carried out.

3.2. Bibliographic search

In addition to this written document analysis, a targeted bibliographic search of recent works was added. After noting the diversity and the extent of the scientific contributions relating to our theme, our bibliographic corpus was limited to English and French literature. Thus, it was possible to proceed as follows: 1) selection of writings on the basis of criteria such as relevance and coherence in relation to our theme; 2) review of the most widely used evaluation models, using an analysis matrix consisting of four parameters: objectives, challenges, strengths and limitations ; and (3) classification of these models by vertical thematic analysis.

4. Results

Our approach allowed us to identify 21 different models of training evaluation. Below, we will present them briefly with particular emphasis on the strengths and weaknesses of each of them.

4.1. The Kirkpatrick Assessment Model (1959)

Since the sixties, the predominant model of training's assessment is that of Kirkpatrick (1959). Over the years, it has become a standard for measuring training. This model, with its four levels of evaluation, makes it possible to distinguish the reaction and satisfaction of the learners (level 1), their learnings and acquisition (level 2: knowledge, know-how and know-be), their behaviors (level 3 on transfer: use of

skills acquired in work situations), and the output of training (level 4: organizational benefits and impact of training on performance at work).

However, it has some limitations. Thus, Le Louarn and Pottiez (2010) evoke the hierarchical and causal character between the four levels of the Kirkpatrick model, while Tamkin, Yarnall and Kerrin (2002) denounce its difficult application to all types of training.

Despite the criticisms made against it, the Kirkpatrick model is widely used (Kirkpatrick & Kirkpatrick, 2006; Gilibert, Genty, Gillet & Marchand, 2008) that is why several authors are inspired by it to give rise to variants. Their intentions were not to create entirely new models, but to perfect the existing model in order to improve its efficiency (Rajeev et al., 2009).

4.2. The five levels approach of Hamblin (1974)

Hamblin (1974) was the first to propose a variant of the Kirkpatrick model. Indeed, in addition to Kirkpatrick's four initial levels, he adds a fifth, the financial effects both on the organization and economy.

This level allows a conversion of the Kirkpatrick fourth level (impact of training on performance at work) in financial value. In this way, Hamblin distinguishes between the impacts of training on the performance of the organization on the one hand, and the purely financial results on the other. This is justifiable because their calculation refers to other evaluation indicators.

4.3. The Phillips model (1997)

Jack Phillips also presents a variant close to the Hamblin model. By separating the evaluation of profits and financial costs from the assessment of other organizational results, he confirms the level of return on investment (ROI), that is, the financial and economic results (Goldwasser, 2001).

4.3.1. The CIRO model of Warr, Bird and Rackham (1970)

The CIRO model (Context, Input, Reaction, Outcomes), developed by Warr, Bird and Rackham in 1970, is the first model of our literature's review that focuses on the assessment of the training process, not on its results.

It consists of evaluating four elements: the "context" of the training, the resources invested in the training (input), the participants' reactions to their training (reaction) and the results of training (outcomes).

Even if it's similar to Kirkpatrick's model, CIRO model presents two new contributions (Worthen, Sanders & Fitzpatrick, 2004):

- Expansion of evaluation targets: The CIRO model was the first model to extend the evaluation targets and to analyze not only the results but also the way in which the training was designed and carried out.
- Implementation in several types of training: The second contribution of the CIRO model is its implementation by the authors in the context of assessment of managerial training (assessment of supervisory skills, communication skills and attitudes).

4.3.2. The OEM model of Kaufman, Keller & Watkins (1995)

In this derivative line, the OEM model, which means Organizational Elements Model, is another descendant of the Kirkpatrick model. Recognizing the frequent use and relative simplicity of the Kirkpatrick model, Kaufman, Keller & Watkins try to improve it by suggesting five levels for assessment.

Adapted from Kaufman, Keller and Watkins (1995).

N°	Level	Evaluation Area
1.a	Enabling	Quality input availability like human, financial and physical resources
1.b	Reaction	Methods, means and process acceptability and efficiency
2	Acquisition	Individual and small group mastery and competency
3	Application	Individual and small group utilization within the organisation
4	Organizational Outputs	Payoff to the organization
5	Societal Outcomes	Payoff to the society

Figure 1. The five-levels Evaluation of Kaufman & al. (1995).

4.3.3. The Business Impact Model (Molenda et al., 1996)

This model modifies little that of Kirkpatrick. It retains the four original grades and complements it with two additional strata at the beginning (stratum 0) and at the end of the assessment process (stratum 5):

- Stratum 0 relates to activity accounting. It is similar to what might be called the measurement of operational outputs (number of people trained, number of training courses implemented in a year, for example).

- Stratum 5 concerns the assessment of the impact of training on society (social impact), already recommended by Hamblin (1974) and later by Kaufman & al. (1995). Molenda & al. (1996) suggest evaluating this level using criteria such as adequacy of trained human resources with local or regional needs, improvement of the organizational climate or the evolution of socio-economic indicators. They also propose to carry out this evaluation using tools such as benefit-cost analysis, the organizational climate survey or statistical indicators on socio-economic conditions.

4.3.4. The KPMT "value added" model of Kearns and Miller (1997)

Finding that the existing models were difficult to apply in practice, Paul Kearns and Tony Miller developed in 1997 a new training evaluation model. The KPMT model guides assessment towards the contribution of training to the creation of added value for the organization. This model resembles the Kirkpatrick model to which Kearns and Miller would have added an extra level. This level, called the core level, focuses not on an outcome but on the business objectives and profitability targeted by the organization.

Thus, like several models presented earlier, the KPMT model emphasizes the importance of linking training and organizational needs. Another particularity of the model is to emphasize the importance of assessing the last level of value added. At this level, the objectives set before the start of the training (basic level) are compared with the financial results obtained (level 4). This model seems to place greater emphasis on financial results, leaving aside human resource outcomes and operational results.

4.3.5. The decision based model of Kraiger (2002)

This model was based on the works of Kraiger, Ford and Salas (1993). They were inspired by several taxonomies of learning: the Bloom taxonomy (1956), the taxonomy of Krathwohl & al. (1964) and the

taxonomy of Gagne (1984). It's a learning training assessment model which's centered on the decision-making (Kraiger & Jerden 2007). It is based on three categories (Kraiger, 2002):

- The cognitive category: verbal knowledge, knowledge organization and cognitive strategies ;

- The skills category: compilation (mastering one, then several competencies at the same time) and automation ;

- The affective category: attitude and motivation (emotional disposition, personal effectiveness).

Compared to the Kirkpatrick model, Kraiger's assessment model differs in three points (Table 1):

- First, it identifies three assessment targets instead of the usual four levels.

- Second, the model links the targets and objectives of the evaluation.

- Finally, the model proposes methods or techniques to assess each target.

-

	Target customers	Objects evaluated	Methods of evaluation	Reasons
Target 3: Performance for the Organization	Leaders Human resources	Performance at work, Transfer and Results	Investigation Estimates Profit / Cost Analysis	Feedback to trainer
Target 2: Changes in the learner	Learner Trainer Human resources	Cognitive, Behavioral and Emotional changes	Investigation Interview Test Work Sample	Feedback to learner Decision making
Target 1: Design and Content	Trainer	Design Delivery,Validity and Relevance of training	Investigation, Content Analysis And Expert Judgment	Decision making, Marketing

Adapted from Kraiger & Jerden (2007).

Figure 2. The Kraiger Model: targets, target customers, assessment objects and methods

4.3.6. The Brown and Sturdevant Reed's model (2002)

Brown and Sturdevant Reed model largely differ from the initial Kirkpatrick model. They are interested in the levels of learning and behavioral change in training evaluation. They argue that a distinction must be made between "subjective" and "objective" learning, both in terms of knowledge and behavioral transfer.

They also highlight the importance of considering the work environment and situational aspects in a transfer assessment (Brown & Sturdevant Reed, 2002). This point of view is shared by McCain (2005) and Bournazel (2005). All these authors suggest that, when assessing training, it is necessary to differentiate the evaluation of perceived and effective learning, acquired knowledge and its transferability in real work situations.

4.3.7. The Holton's Evaluation Model (1996)

Holton (1996) proposes an outline of an assessment model based on the results of earlier research and including intermediate variables that may explain learning and transfer. The evaluation focuses on three levels of outcomes already defined by Kirkpatrick: learning, individual performance and organizational outcomes.

However, the author identifies elements likely to influence these results. These are grouped into three main categories: facilitator factors (the influence of individual aptitudes on learning, the influence

of transfer design on performance); environmental factors (the influence of external variables on learning, or transfer conditions on performance); and motivational factors (motivation to learn or to transfer).

In addition, factors that may have a secondary influence, such as personality characteristics or attitudes at work, are taken into account. Thus, the use of this model may lead not only to the assessment of the training outcomes, but also to some explanatory factors that may influence the level of results obtained (Holton et al., 2007). In fact, the interest of this model is that it draws the attention of the professional on a set of variables often neglected. Moreover, its complexity is better suited to researchers than to business practitioners who often resort to the Kirkpatrick model.

4.3.8. Beech and Leather Model (2006): Assessing effects and impacts

The authors of this model presented a synthesis of the models of Kirkpatrick (1959), Warr et al., (1970), Hamblin (1974), Phillips (1997) and Kraiger et al., (1993). In response to criticism of the Kirkpatrick model, Beech and Leather (2006) found that it was not possible to differentiate the immediate (short-term) and subsequent (medium and long-term) effects of a training. Thus, in a recent modeling, they try to answer the various criticisms addressed.

They start from the fact that all these earlier models are not contradictory, but all have limits. Then they propose to revisit and compile the contributions of the various approaches into a single synthetic model, seeking more exhaustiveness.

Levels	Concepts	Tools
Reaction **Learning**	Satisfaction End-of-course assessment	Questionnaires and subjective assessment sheets, Knowledge tests, typical exercises to test the declarative knowledge
Immediate effect level	Declarative And organizational knowledge	Problem solving situations (actions to be carried out, Group exercises, problem solving plan, justification ...) to test organizational knowledge, relevance and appeal of content
Behaviors or **Intermediate effect** **level**	Skills of know-How (expertise)	Self-assessments, explanations of the causes of errors and stress ...
	Affects Attitudes and motivations Trust Self-efficacy	Standardized psychological scales Self-assessments of capacities, Learning motivation questionnaires, Degree of satisfaction, Skills tests, Self-esteem tests ...
« Outcomes » **Ultimate effect level**	Productivity, measures and Critical Incidents	Number of incidents (type and importance), Customer satisfaction indicator (number of Complaints ...)
Financial level	Profitability	Profits Indicators, Indicators of cost variation…

Figure 3. Combined model of Beech and Leather (2006) adapted for training Education

4.3.9. CIPP model of Stufflebeam et al. (1971)

In 1971, Daniel L. Stufflebeam introduced a new assessment model whose acronym CIPP refers to the following concepts:

- ***Context***

Context is the first element that the model invites to evaluate. As in the CIRO model, the evaluation includes an analysis of needs and the formulation of objectives. But it also goes further by including a study of the environmental variables that will condition these needs and objectives.

- ***Input*** **(Resources)**

Inputs are the second element of the CIPP model. They combine the human, material and financial resources invested in training as well as the learning strategies used to achieve the objectives set.

- ***The process***

This model also recommends an evaluation of the training process. The process is the way in which resources are used. From a formative point of view, evaluation of the process must help the trainer to carry out his activity by aiming at an optimal use of resources (Stufflebeam and Shinkfield, 2007).

Summatively, evaluation of the process involves an analysis of participants' reactions to training, as in the CIRO and Kirkpatrick models (Stufflebeam, 2003).

- ***Products***

The products are the last element of the CIPP model. In this respect, the model makes no distinction according to the type of results. It encompasses all the results produced by the training, both desired and undesired results, short-term and long-term outcomes.

On the other hand, the assessor is invited to compare the results obtained with the goals fixed, the resources invested and the environment in which the training takes place.

The Stufflebeam model is a reference that applies to assessments of programs, projects, human resources, products, institutions and systems. It has many similarities with the CIRO model of War et al.

The CIPP is a first model focused on the evaluation of the training process. It is widely used in training device assessment in order to take and implement readjustment decisions.

4.3.10. The IPO model (1990)

The IPO model (Input, Process, and Output) is a third model focused on the evaluation of the training process. Developed by David Bushnell in 1990, the IPO model aims to control the costs of training and presents several similarities with the CIRO and CIPP models.

In developing his model, Bushnell started from the observation that organizations were increasingly looking to offer their employees quality training at a reasonable cost. Similarly, he noted that the costs of developing training tended to increase.

As a result, its IPO model focuses on assessing:

- Resources invested in training (Input) and which can be very varied natures (experience and competence of the trainer, availability of logistics and equipment, allocation of a budget adapted to needs ...);

- Training process (Process) as a formative evaluation. It is used to determine learning goals, to define the most appropriate criteria for designing training and to select the most adequate teaching strategies;

- Ouput and products training: Short-term results corresponding to reactions, learning and behavioral changes of trained individuals; and long-term results (Outcomes) which correspond to an increase in financial results, profitability and competitiveness.

4.3.11. The Roegiers Open Box model (1993)

Drawing on the work of Stuffelbeam (1971), Roegiers and Bourgeois (1993) developed a model that they call "The Open Box". This model invites to analyze the following three contexts:
- The institution (essentially its needs);
- Human resources (learners and trainers);
- The normative framework (legal, social and didactic: pedagogical techniques).

The interaction of these contexts influences the course and outcome of the training.

For Roegiers (1997), this model is specific for the components analysis of any training, project, plan or program, but also for the analysis of the steps involved in developing, implementing and assessing this action.

In addition, Roegiers points out that his model is a model of analysis, a model that attempts to identify the way things happen without necessarily indicating how they should happen. It is therefore a model which is not intended to be prescriptive.

4.3.12. The Bournazel model: The Training Pentagon

The Bournazel model (2005), called the pentagon of training, aims to evaluate a training system. As its name suggests, it is based on five categories of indicators: classical performance indicators (improvement of knowledge and skills), social indicators (equity, equaliy), accompanying indicators (supervision and monitoring), innovation indicators (new information and communication technologies) and economic and financial indicators. These categories of indicators reflect five different dimensions of the training system.

4.3.13. The Quality Circle model (Dejean, 2002)

This model, which is part of a quality approach, is based on annual evaluations (questionnaires completed by the students), as well as four-year evaluations including a multitude of analyzes (needs analysis, training goals analysis, analysis of learning outcomes, training and evaluation methods), in addition to exchanges in which all actors involved in training are present (Dejean, 2002).

This approach has the advantage of showing that the quality of training depends on several factors, in particular co-production and co-assessment between students, trainers and training managers.

4.3.14. The Dejean circular evaluation model (2004)

A variant of the quality circle has been carried out in Spanish universities. This is the circular evaluation (Dejean, 2004). Circular as it is carried out by all the actors concerned: students, teachers, supervisors, rectorate and ministry.

Students evaluate the lessons learned. Teachers self-evaluate, evaluate the different aspects of their teaching (curricula and program, working conditions and available resources, etc.), and perform a meta-evaluation (usefulness of the evaluations made by the students). The rectorate evaluates the trainers teaching and the ministry evaluates the teacher research and production.

The various actors give their opinions as experts on various aspects of training, and the summation of these evaluations provides information on the quality of university training. Therefore, the circular evaluation would better optimize the quality of training (Powell &Yalcin; 2010).

In this context, Dejean (2004) states that "evaluation of training quality can only be approached by taking account of the opinions of a multiplicity of actors (with different interests and issues)".

According to Michalski and Cousins (2001), the assessment of training is limited by a lack of knowledge of the perceptions and expectations of the main actors in the system. These perceptions and expectations would be different for each group of actors.

The sponsors focus on the financial results of the training and pursue a goal of financial profitability. The trainees hope that the training will allow them to increase their professional skills and to progress

in their careers. While training providers consider sponsors and trained as their clients. The important thing for them is that their clients are satisfied with the training offered.

4.3.15. The Noriaki Kano model (1984)

The Kano model is a tool invented by Dr. Noriaki Kano. Its purpose is to evaluate customer satisfaction. It expresses the fact that different quality factors can have a different impact on customer satisfaction.

For Kano, there is no symmetry between satisfaction and customer dissatisfaction. Thus, certain factors may, through their absence, strongly influence dissatisfaction without providing satisfaction when they are present.

Similarly, certain factors can generate high satisfaction without causing dissatisfaction if they are not present.

This is the case, for example, of the unforeseen gift accompanying a service or the unexpected reward.

In short, the Kano model proposes to relate satisfaction of requirements (response to needs) and customer satisfaction. It distinguishes three types of requirements:

- Mandatory requirements, basic expectations ("Basic needs", "Must Have"): These basic requirements are not always expressed. However, they are obvious to the customer and must be imperatively satisfied. These are not a priority component of satisfaction, however if they are absent, they generate dissatisfaction.

- Requirements expressed and proportional expectations ("Performance needs"): The need is expressed and the customer's satisfaction is proportional to the level of performance and quality training. It is a strong lever of customer satisfaction and a priority axis for the quality of training (the ergonomic and logistical conditions of training ...).

- Latent requirements, attractive expectations ("Exciters", Delighters): These requirements correspond to a need not necessarily expressed and sometimes unconscious. It is a lever of attractiveness and innovation. It is a pleasant surprise that can make the difference. It's a source of great satisfaction at little or no cost (free availability of wifi or a photocopier in training class). On the other hand, the absence of this functionality will not cause dissatisfaction or frustration (Riadh and Ladhari; 2005).

In the end, even if its implementation is relatively cumbersome to conduct, it has the advantage of allowing a fairly good understanding of the expectations and needs of customers (Witell and Logfren, 2007). The keys to its success remain its simplicity and especially the fact that it is centered on the users (customers).

Customer Requirements		Dysfunctional Question				
		Like	Expect	Neutral	Live with	Dislike
Functional Question	Like	Q	E	E	E	L
	Expect	R	I	I	I	M
	Neutral	R	I	I	I	M
	Live with	R	I	I	I	M
	Dislike	R	R	R	R	Q

M : Must have ; L : Linear ; E : Exciter ; R : Reverse ; Q : Questionable ; I : Indifferent

Figure 4. Links between the different requirements: the mandatory requirements ("Basic needs", "Must Have"), requirements expressed or linear requirements, ("Performance needs") and latent and attractive requirements ("Exciters", Delighters).

4.3.16. The Servqual model (Parasuraman, Zeithaml & Berry; 1988): a tool for measuring service quality

The service is very different from a product: it is not tangible; it takes multiple forms and is delivered at the same time that it is consumed. Therefore, it's difficult to define what a good service is or to assess its quality.

Three researchers, Parasuraman, Zeithaml and Berry (1988), addressed this difficulty by designing a quality of service evaluation model. They developed an instrument to measure the gap between clients' perceptions and expectations of the quality of service provided.

This model is based on the use of a 22 items scale measuring the gap between expected service and service perceived by consumers / users. This scale is structured in five dimensions or determinants of service quality:

- Physical aspect or tangible elements: Logistics, appearance of human resource and ease of use of equipment;

- Reliability: Ability to provide the service exactly as required;

- Responsiveness or reactivity: Desire to serve, assist and provide prompt service to the customer;

- Insurance: Ability to inspire trust and know how;

- Empathy: Personalized attention, understanding of the customer to provide a human and specific service to customers.

This model allows a follow-up of the customers' trends and expectations. However, its implementation requires adapting the questionnaire to the education sector; to use a representative sample of research, and use the tool periodically and combine it with other internal and external surveys (Parasuraman & al., 2005).

4.3.17. The IMTEE model of Alvarez, Salas and Garofano (2004)

The Integrated Model of Training Evaluation and Effectiveness (IMTEE) draws on the previous models of Kirkpatrick (1959), Kraiger (2002), and Holton (1996).

This model integrates the relationships between the evaluation levels (reactions, self-efficacy, knowledge, skills) and the three categories of transfer factors (individual, educational, organizational) (Green, 2001). The evaluation of the effectiveness of training is defined here as the study of the variables likely to influence the results of training at different levels of the process: before, during and after training (Alvarez et al., 2004).

Alvarez (2004) first draws a distinction between the training evaluation and the evaluation of the training effectiveness. The training evaluation refers to an approach that examines the extent to which the training goals are being achieved. And the evaluation of the training effectiveness aims to study the variables that may influence the training impact. These variables are related to the individual, to the training or to the characteristics of the workplace. This distinction explains the results and makes it possible to formulate recommendations for increasing more the effectiveness of training (El-Akremi & Khalbous, 2004).

Overall, the evaluation levels of the IMTEE model are similar to those of the main known models (Kirkpatrick (1976); Holton (1996); Kraiger (2002). Nevertheless, Alvarez et al., (2004) present a model of six levels of evaluation, including self-efficacy, which is an addition to other models.

The first level of reactions of the trainees refers to the perceived usefulness of the training and its relevance. Reactions are seen as indicators of the content and design of the training. The second level is the post - training self - efficacy. At the third level is cognitive learning in training. Then, the model includes the level of training performance that refers to the behaviors demonstrated at the end of

the training. The fifth level, about transfer, consists of behaviors manifested in the workplace. Finally, the level of outcomes is defined as the impact of behavioral changes on trainees at the end of training.

The study of these IMTEE variables, called efficacy variables, should help to understand why the training results are good or bad.

4.3.18. The motivational parameters model of Noe (1986)

Raymond Noe develops one of the first models dealing with the psychological parameters influencing the effectiveness of training. He essentially identifies eight individual determinants that can motivate the trainee to develop and then transfer his learning: (1) the control locus of the trained person, (2) self-efficacy, (3) his training expectations, (4) feedback received from the previous assessments, (5) his motivation to learn, (6) his motivation to transfer his learning into work, (7) his perception of the working environment and (8) his attitude towards his work and career.

In practice, all variables in this model influence the trainee motivation to learn (Burke and Hutchins, 2007), and only the perceived working conditions have a direct influence on the transfer of skills. Similarly, only the working conditions, the trainee expectations and the feedback received previously influence organizational outcomes (Noe, 2008).

The Noe model has the merit of taking up the results categories of the Kirkpatrick model and listing for the first time several psychological determinants that can influence these results. Nevertheless, this model doesn't take into account the influence of other categories of variables on the training results such as (1) sequencing the training, (2) its content, (3) the pedagogical principles on which it's based, and (4) the support the trainer receives from his supervisors and co-workers (Ryan & Deci, 2002).

Finally, we can observe a relative complexity of the model, which will be found in the majority of the models identified: the variables affect several categories of results and can therefore influence each other.

5. Discussion

After identifying these assessment models, our analysis was enlightened by their comparison and followed by a categorization into four families of models. A comparison of the structure, goals, strengths and weaknesses of the evaluation models reviewed leads us to make several observations and draw several lessons.

First, our 21 identified models can be classified into four distinct categories.

1.   Categories of results-based assessment models

This category consists of 10 evaluation models (Figure 1). All of which are essentially derived from that of Kirkpatrick (1959). They focus on the assessment of the different levels of results produced by training. Therefore, they allow for a summative evaluation (Le Louarn & Pottiez, 2010).

2.   Categories of assessment models focusing on process

Six process-oriented evaluation models were reviewed. Everyone is trying to further develop the idea of formative evaluation that can be found in the CIPP or IPO models. They invite trainees to carry out

an evaluation of each stage of their training in order to develop the culture of assessment and continuous improvement (Muller, Carre & Esnault, 2007).

This also corresponds to the Newby (1992) and Pulley (1994) "responsive evaluation" models, which encourage reflection on the assessment process during training (Figure 1).
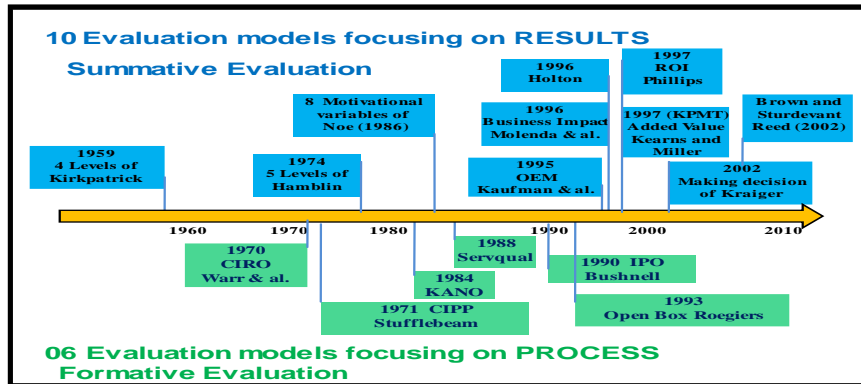

Figure 6. Evolution of the main training evaluation models.

3. Categorie of evaluation models focusing on quality training

It consists of three models mainly aimed at assessing and improving the quality of training (Figure 2). Each model is based on a multitude of evaluations involving different actors and covering several dimensions of the training (multi-dimensional and multi-actor assessment) (Gerard, 2017).

The many actors listed above actually represent four groups with divergent interests in the organization: Investors or funders (creditors, shareholders, and suppliers), producers (administrators, managers, and trainers), customers (beneficiaries, users, consumers of training) and regulators (government, ministry, associations).

4. Categorie of evaluation models focused on training effectiveness

For this category (Figure 2), the training assessment is directly related to organizational effectiveness. Effectiveness is therefore a standard of organizational performance linked to the values of the individuals or groups concerned by the training (Garrett & Brian, 2007; Towler & Dipboye, 2009).
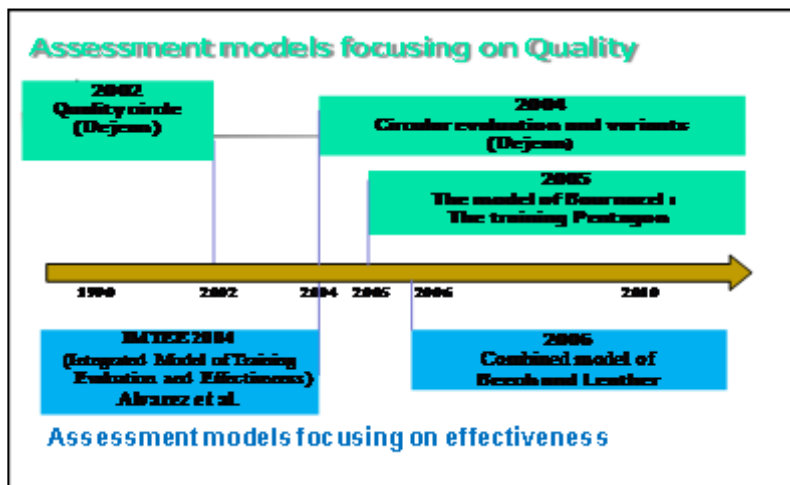

Figure 7. Assessment models focusing on quality and effectiveness.

On the other hand, the training device evaluation task is so complex that it solicits several actors and interferes with several variables not always easy to measure and master.

- *A multitude of evaluation variables*

Our research reveals a multitude of variables that can influence the training results. They can be classified into three categories: (1) internal variables related to the trainee, such as his cognitive abilities, self-efficacy, motivation to learn, expectations, personality traits and attitudes towards training ; (2) external variables related to the work environment, such as supervisor support, colleague support and transfer conditions; and (3) process variables related to the training design, such as the adequacy between the educational goals and the goals of the organization, the teaching methods, the chosen sequencing and the pedagogical content taught (Igalens & Peretti, 2008).

Most models show a problem of definition and delimitation of these variables. A variable in a model may correspond to several variables in another model, which complicates the comparison of evaluations using different models. Almost all variables exert influence at different levels of the training chain. Consequently, they can influence each other.

6. Conclusion

Our review does not claim to be exhaustive. We had the ambition to present a methodological pseudo repertory focusing the main models of assessment of the training quality.

There are a multitude of models that can guide the assessor in the evaluation of training device, but it was difficult for us to judge the quality of a model. In terms of training evaluation, there is no "cookbook" approach" (Nickols, 2000). Indeed, it's so difficult to assess the ability of a model to respond jointly to the needs of researchers, practitioners, trainees, sponsors, shareholders and training professionals, nor to what extent they are used by these communities.

For example, it is difficult to judge the relevance of the various components of the models of IPO (Bushnell, 1990), KPMT (Kearns & Miller, 1997) or IMTEE (Alvarez et al., 2004), since no accessible publication, to our knowledge, presents them in sufficient detail.

All of these models are derived from earlier models, most commonly that of Kirkpatrick and Stufflebeam. Some models try to adapt the evaluation process to a particular context, others extend it to other dimensions that are often overlooked, or use new variables.

Beyond these 21 identified methodological models and the measures they allow, five main conclusions can be drawn.
1) No model is perfect. Each model covers one or at most three dimensions of the quality of a training device, and thus has limits. There is still a lack of tools or methods for simultaneously and rigorously evaluating training outcomes at different levels on the one hand, and for evaluating all of the variables that can influence these outcomes on the other.
2) There are more than thirty determinants that can influence the results of training. They may be related to the psychology of the trainee, his predispositions and aptitudes, the training environment, the skills of the trainers, the training design, or to sponsors and shareholders expectations.
3) These variables present definition and delimitation problems. A variable in one model may correspond to several distinct variables in another, complicating the comparison of assessments based on different models.
4) Numerous studies of these variables produce results that can sometimes be contradictory and disconnected from one another. The quantitative approach used in most studies seems to explain this.
5) Finally, quality is plural. Thus, two approaches are required: one conceptual (systemic approach) and the other methodological (total triangulation). The first allows a combination of several assessment models to cover the maximum dimensions of a training device. The second allows cross-checking of

the results produced by several actors (assessors), using different data collection instruments and various data analysis tools.

References

Alvarez, K., Salas, E. & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human resource development Review, 3*(4), 385-416.

Beech, B. & Leather, P. (2006). Workplace violence in the health care sector: A review of staff training and integration of training evaluation models. *Aggression and violent behavior, 11*(1), 27-43.

Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. *Training & Development Journal, 44*(3), 41-44.

Bournazel, A. (2005). La formation professionnelle: Gestion et Evaluation, le Pentagone de la Formation. Paris: SEFI.

Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Brown, R. E. & Sturdevant Reed, C. (2002). An integral approach to evaluating outcome evaluation training. *American Journal of Evaluation, 23*, 1–17.

Burke, L. A. & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review, 6*(3), 263-296.

Dejean, J. (2002). *L'evaluation de l'enseignement dans les universites françaises. Rapport etabli à la demande du Haut conseil de l'evaluation de l'ecole.* Haut Conseil de l'Evaluation de l'Ecole: Paris.

Dejean, J. (2004). *De Berlin à Bergen, nouveaux enjeux de l'évaluation.* Premier colloque du Comité National d'Evaluation: Dijon.

El-Akremi, A. & Khalbous, R. O. (2004). Mesure de résultats de la formation. approche par les déterminants de l'efficacité. *Congrès 2004 de l'Association de Gestion des ressources humaines, La GRH mesurée*, 1552-1575.

Gagne, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*(4), 377-385.

Garrett, J. & Brian, H. K. (2007). How to measure Management Training and Development Effectiveness. *Journal of European Industrial Training, 14*(9), 1-10.

Gerard, L. (2017). *Le chantier de la qualite en formation: Enjeu de 2017.* Paris: Entreprise & Carrières.

Gilibert, D., Genty, D., Gillet, I. & Marchand, M. (2008). *Evaluer l'efficacite des dispositifs de formation professionnelle.* Paris: Premières journees de l'Association française de psychologie du travail et des organisations.

Goldwasser, D. (2001). Beyond ROI. *Training, 38*(1), 82-90.

Green, E. (2001). *The influence of individual and work environment characteristics on trainee motivation and training effectiveness measures* (Unpublished doctorate thesis). USA: North Carolina State.

Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Holton, E. F., Bates, R. A., Bookter, A. I. & Yamkovenko, V. B. (2007). Convergent and divergent validity of the learning transfer system inventory. *Human Resource Development Quarterly, 18*(3), 385-419.

Hamblin, A. C. (1974). *The evaluation and control of training*. London: McGraw-Hill.

Igalens, J. & Peretti, J-M. (2008). *Audit Social: Meilleures pratiques, methodes, outils.* Paris: Editions d'Organisation.

Kano, N. (1984). Attractive quality and must-be quality. *Hinshitsu, Quality, the Journal of the Japanese Society for Quality Control, 14*(2), 39- 48.

Kaufman, R., Keller, J. & Watkins, R. (1996). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance Improvement, 35*(2), 8-12.

Kearns, P. & Miller, T. (1997). *Measuring the impact of training and development on the bottom line.* London, UK: Pitman Publishing.

Kirkpatrick, D. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors, 13*(11), 3-9.

Kirkpatrick, D. L. (1976). Evaluation of training. In Craig, R. L., (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301 319). New York: McGraw-Hill.

Kirkpatrick, D. & Kirkpatrick, J. (2006). *Evaluating training programs*. San Fransisco, CA: Berret-Koehler Publishers Inc.

Kraiger, K. (2002). Decision-based evaluation. In *Creating, implementing and managing effective training and development. State-of-the-art lessons for practice.* Society for Industrial and Organizational Psychology ed., pp. 331-375). San Francisco, CA: Jossey-Bass.

Kraiger, K., Ford, J. K. & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311 328.

Lotfi, R. & Alami, N. T. (2017). Literature review on assessment models of the quality of training devices. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. 4(1), pp 481-495. Available from: www.prosoc.eu

Kraiger, K. & Jerden, E. (2007). A new look at learner control: meta-analytic results and directions for future research. In *"Where is the Learning in Distance Learning? Towards a Science of Distributed Learning and Training",* ed. SM Fiore, E Salas, pp. 65–90. Washington, DC: APA Books.

Krathwohl, D. R., Bloom, B. S. & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook 11: The affective domain.* New York: David McKay.

Ladhari, R. (2005). La satisfaction du consommateur, ses determinants et ses consequences. *Revue de l'Universite de Moncton, 36*(2), 171-201.

Le Louarn, J. Y., & Pottiez, J. (2010). Validation partielle du modèle d'evaluation des formations de Kirkpatrick. In *communication au 21e congrès de l'AGRH,* Saint-Malo.

Molenda, M., Pershing, J. & Reigeluth, C. (1996). Designing instructional systems. In *Craig, R. (Ed.), The ASTD Training and development handbook: a guide to human resource development* (4th edition). New York: McGraw Hill.

Michalski, G. V. & Cousins, J. B. (2001). Multiple perspectives on training evaluation: Probing stakeholder perceptions in a global network development firm. *American Journal of Evaluation, 22*(1), 37-53.

Muller, J-L., Carre, O. & Esnault, N. (2007). *Motiver aujourd'hui, c'est possible.* Paris: Editions ESF.

Newby, T. (1992). *Training evaluation handbook.* Washington: Pfeiffer & Company.

Noe, R. A. (1986). Trainees' attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review, 11*(4), 736-749.

Noe, R. A. (2008). *Employee training and development*. New York: McGraw Hill.

Nickols, F. (2000). Evaluating training: There is no"cookbook" approach". Retrieved from http://www.nickols.us/evaluating_training.pdf on 1 March 2017.

Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*(4), 41-50.

Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perc. *Journal of Retailing, 64*(1), 12-40.

Parasuraman, A., Zeithaml, V. A. & Malhotra, A. (2005). ES-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research, 7*(3), 213-233.

Phillips, J. (1997). *Une approche rationnelle de l' evaluation des programmes de formation*. Paris: Y compris le calcul de ROI.

Phillips, J. J. (1997). *Return on investment in training and performance improvement programs.* Company, Houston: Gulf Publishing

Skylar Powell, K. & Yalcin, S. (2010). Managerial training effectiveness: A meta-analysis 1952-2002. *Personnel Review, 39*(2), 227-241.

Pulley, M. L. (1994). Navigating the evaluation rapids. *Training & Development, 48*(9), 19-25.

Rajeev, P., Madan, M. S. & Jayarajan, K. (2009). Revisiting Kirkpatrick's model–an evaluation of an academic training course. *Current Science, 96*(2), 272-276.

Roegiers, X. (1997). *Analyser une action d'education ou de formation.* Paris-Bruxelles : De Boeck Universite.

Roegiers, X. & Bourgeois, E. (1993). *Une revision du modèle de Stufflebeam.* Lyon: Communication au colloque de l'ADMEE.

Ryan, R. M. & Deci, E. L. (2002). An overview of self-determination theory. In Deci, E. L. & Ryan, R.M. (Eds.), *Handbook of self-determination research* (pp. 3-33). Rochester, NY: University of Rochester Press.

Stufflebeam, D .L (1971). *Evaluation de l'education et la prise de decision.* Itasca, IL: Peacock Publishers.

Stufflebeam, D. L. (2003). The CIPP model for evaluation. In Stufflebeam, D.L., & Kellaghan, T. (Eds.), *The international handbook of educational evaluation* (Chapter 2). Boston, MA: Kluwer Academic Publishers.

Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, & applications*. San Francisco, CA: Jossey-Bass.

Tamkin, P., Yarnall, J. & Kerrin, M. (2002). Kirkpatrick and beyond: A review of models of training evaluation. *Institute for Employment Studies*. Report number: 392.

Torres, R. T., Preskill, H. S. & Piontek, M. E. (1996). Evaluation strategies for communicating and reporting. Thousand Oaks, CA: Sage Publications, Inc.

Towler, A. J. & Dipboye, R. L. (2009). Effects of Trainer expressiveness, organization and training orientation on training outcomes. *Journal of Applied Psychology, 86*(4), 664-673.

Warr, P., Bird, M. & Rackham, N. (1970). *Evaluation of management training.* London: Gower Press.

Witell, L. & Lofgren, M. (2007). Classification of quality attributes. *Managing Service Quality: An International Journal, 17*(1), 54-73.

Worthen, B. R., Sanders, J. R. & Fitzpatrick, J. L. (2004). *Educational evaluation: Alternative approaches and practical guidelines.* Boston: Allyn & Bacon.