



Revue des modèles en évaluation de formation: approches conceptuelles individuelles et sociales

Daniel Gilibert, Isabelle Gillet

► To cite this version:

Daniel Gilibert, Isabelle Gillet. Revue des modèles en évaluation de formation: approches conceptuelles individuelles et sociales. *Pratiques Psychologiques*, 2010, 16 (3), pp.217-238. hal-00588973

HAL Id: hal-00588973

<https://u-bourgogne.hal.science/hal-00588973>

Submitted on 27 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gilibert, D. & Gillet, I. (2010). Revue des modèles en évaluation de formation, approches conceptuelles individuelles et sociales. *Pratiques Psychologiques*, 16, 217-238.

Exemplaire de travail des auteurs

Pratiques Psychologiques rubrique Psychologie du travail

Revue des modèles en évaluation de formation: approches conceptuelles individuelles et sociales*

*Training evaluation models: Individualistic and
social approaches*

D. Gilibert ^{a 1} I. Gillet ^{b 2}

^a EA 4180, laboratoire SPMS, pôle apprentissages-acquisitions-formation-éducation, université de Bourgogne, bureau R 18, campus Montmuzard, BP 26513, 21065 Dijon cedex, France

^b EA4180, laboratoire SPMS, pôle apprentissages-acquisitions -formation-éducation, université de Bourgogne, bureau R 15, campus Montmuzard, BP 26513, 21065 Dijon cedex, France

Reçu le 17 juin 2008 ; accepté le 26 mars 2009

Résumé

Cette revue présente les modèles les plus représentatifs pour effectuer des évaluations de dispositifs de formation. Après avoir évoqué les pratiques, les résistances et les mesures possibles en évaluation de formation, une présentation des principaux modèles est faite. Le modèle à quatre niveaux de Kirkpatrick est notamment présenté et discuté ainsi que le CIRO de Ward, Bird et Rackham visant à poser le problème de l'efficacité de la formation au regard de son adéquation aux besoins et aux contextes de formation. Sont également présentés les modèles de Kraiger sur les niveaux d'apprentissage et la synthèse de Beech et Leather. Au niveau pratique, les finalités, modalités et enjeux de telles évaluations sont alors discutés. Au niveau théorique, la discussion porte sur la nécessité de prendre en compte les interactions entre les acteurs de la formation.

© 2009 Société française de psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

Mots clés : Évaluation de formation ; Erreur fondamentale ; Sentiment d'efficacité personnelle ; Internalisation

* Ce travail s'inscrit dans le cadre d'une convention de recherche de trois ans entre l'université de Bourgogne, le Conseil général de Bourgogne et le service universitaire de formation continue de l'université de Bourgogne: Convention no 06832AAO14S5166 «évaluation de dispositifs de formation et d'insertion» ayant contribué au financement de la thèse du second auteur en qualité de chargée d'étude.

* Auteur correspondant.

Adresses e-mail: Daniel.Gilibert@u-bourgogne.fr (D. Gilibert), isa.gillet4@orange.fr (I. Gillet).

¹ Docteur.

² Chargée d'étude.

Abstract

This review presents the most representative models used to carry out training evaluations. After a recall of the practices, resistances and possible measurements in training evaluation, an overview of the principal models is carried out. The Kirkpatrick's four level-model, in particular, is presented and discussed, as well as the CIRO of Ward, Bird and Rackham aimed at posing the problem of the effectiveness of training, taking into consideration its adequacy to the needs and the contexts of the training. The Kraiger's model on the levels of learning and the synthesis of Beech and Leather are also presented. At a practical level, the finalities, methods and stakes of such evaluations are discussed. At the theoretical level, the discussion relates to the need to take into account the interactions between the stakeholders of the training.

© 2009 Société française de psychologie. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Training evaluation; Fundamental attribution error; Self-efficacy; Internalization

1. Introduction

Cet article a pour objet l'évaluation des formations. Face à l'augmentation croissante de l'offre de formation dans tous les secteurs et la privatisation de nombreuses offres de formation, la qualité de la formation est devenue un enjeu majeur pour les professionnels du secteur et un champ d'étude digne d'intérêt pour les chercheurs. Selon McCain (2005), le but d'une évaluation est de déterminer et d'améliorer la valeur d'une formation. Cette évaluation, qui s'effectue sur la base de standards ou de critères spécifiques à élaborer, apprécie une formation en vue de prendre une décision (maintenir ou non la formation en l'état).

La formation est la plupart du temps un concept qui est assez mal défini, tant elle peut prendre des formes variées adaptées à chaque demande, besoin ou contexte. Le point sur lequel s'accordent les différents manuels en psychologie du travail (Huteau, 2003; Arnold, 2005; Spector, 2006) concerne les étapes successives que devrait idéalement suivre une activité de formation: évaluation des besoins, définition d'objectifs, préparation, activité de formation et évaluation de ses effets. Nous traiterons essentiellement ici de la dernière étape bien que l'activité d'évaluation implique nécessairement un renvoi aux étapes précédentes. Cette contribution a donc pour objectif d'exposer des concepts et des mesures pouvant être utilisés en évaluation de formation par le praticien ou le chercheur. Ces concepts et ces mesures peuvent s'appliquer théoriquement à tout type de formation bien que leur sélection puisse dépendre des objectifs prioritaires de chaque type de formation (Huteau, 2003)³. Il appartient donc à celui qui mène cette évaluation de définir les priorités de son évaluation au regard du cahier des charges précisant les objectifs visés ou de définir des objectifs prioritaires et secondaires qui devront être explorés. Cette présentation sera discutée au regard du risque pratique d'utiliser exclusivement les réactions et réponses des stagiaires (*self-reports*) ou leur acquis pour établir les résultats d'une évaluation de l'action de formation. Concernant la recherche en psychologie, nous discuterons des limites d'une approche exclusivement individuelle centrée sur les personnes formées. Dans ce sens, nous proposons d'explorer le point de vue des différents acteurs sur la qualité de la formation, les déterminants de sa qualité et leurs interactions.

Bournazel (2005), un des quelques auteurs francophones à s'être penché sur le sujet, rapporte que les premières recherches sur l'évaluation remontent au début du xx^e siècle. Selon lui, la science de l'évaluation s'est tout d'abord développée aux États-Unis. La prise de conscience de la nécessité

³ Il est évident qu'une « formation générale » et longue n'a pas les mêmes objectifs qu'une formation professionnelle courte et centrée sur un savoir-faire spécifique immédiatement transférable en situation de travail.

de l'évaluation, selon le même auteur, s'est surtout faite en réponse à l'obligation contractuelle de justifier l'usage des fonds publics, d'estimer les coûts et les résultats des formations. L'évaluation dans ce contexte est «à la fois un instrument de reconnaissance de la spécificité des politiques publiques et un outil de mesure de leurs performances» (Bournazel, 2005, p. 181). L'évaluation est, de plus, employée pour faire face à des crises de compétences (Rosholm et al., 2007) ou dans l'intérêt de besoins commerciaux dans les grandes organisations (Dixon, 1996).

Suite à cette prise de conscience, un grand nombre d'écrits ont été publiés sur ce thème. Phillips et al. (2004) recensent une grande variété d'ouvrages sur le sujet et ce dans plusieurs domaines: apprentissage et développement, sciences sociales, éducation et organisations gouvernementales. De nombreuses sources d'informations sont notamment disponibles: livres, conférences, American Society for Training et Development (ASTD), Association internationale de pédagogie universitaire (AIPU)...

Nous rappellerons dans un premier temps, les constats des études ayant trait à la rareté des pratiques d'évaluation, aux difficultés et aux résistances à l'implantation d'une évaluation et nous présenterons quelques exemples illustrant les différentes dimensions pouvant néanmoins être évaluées. Dans un second temps, nous évoquerons les principaux concepts jugés pertinents dans la littérature proposant des modèles d'évaluation de formation, ainsi que les différentes mesures et agencements suggérés. Dans la discussion, nous tenterons de dégager des axes de développement possibles de ce sujet, en partant d'une approche psychologique et psychosociologique de cet objet.

2. La rareté des pratiques d'évaluation de formation et l'expression des résistances psychosociologiques

S'appuyant sur deux études, une réalisée par la Training Agency (1989) et une par l'Industrial Society (1994; voir Thackwray, 1997), Beech et Leather (2006) concluent que la pratique des évaluations de formations est peu fréquente et que, lorsque des évaluations sont entreprises, elles se limitent le plus souvent à une mesure de satisfaction des stagiaires à l'issue des formations suivies. Ainsi, l'étude de la Training Agency (Bee et Bee, 1994) révèle que 15 % des organisations interrogées ont tenté une évaluation et que seulement 2,5 % ont effectué une analyse de rentabilité à cette occasion. L'enquête de l'Industrial Society indique quant à elle que:

- 20 % des entreprises interrogées n'ont pas entrepris d'évaluation;
- 80 % des répondants l'ont fait mais n'évoquent que l'utilisation de questionnaires de satisfaction, et que;
- 14 % seulement des entreprises réalisant une évaluation ont eu recours à un questionnaire visant à mesurer l'atteinte des objectifs stratégiques des ressources humaines en lien avec les actions de formation mises en place.

Bournazel (2005), examinant les pratiques françaises en matière d'évaluation, signale que la France serait en retard dans ce domaine comparée aux pays anglo-saxons notamment. N'y seraient conduites que des évaluations des politiques publiques et non des évaluations de formation. Dejean (2002), en recensant les pratiques d'évaluation des universités françaises, constate des pratiques relativement éparpillées. Un certain nombre de croyances plus ou moins fondées peuvent expliquer que le recours aux évaluations de formation soit plus rare qu'on ne le pense. Les évaluations seraient, en effet, souvent perçues comme étant inutiles (Sanderson, 1992; Phillips et al., 2004), trop coûteuses en moyens et en temps, trop subjectives car prenant en compte essentiellement des sentiments ou des appréciations fournies par les stagiaires, menaçantes pour les formateurs car

pouvant mettre en lumière des inaptitudes, plus adaptées à certain type de formations, difficiles à implanter, à répliquer et leurs résultats s'avèreraient souvent difficiles à interpréter.

Plusieurs auteurs ont tenté de répertorier les problèmes mentionnés par les professionnels pouvant expliquer ces résistances (Sanderson, 1992 mentionné par Beech et Leather, 2006; Thackwray, 1997; Phillips et al., 2004). Les principaux problèmes évoqués ont trait au manque de temps, au manque de connaissances techniques à propos de l'évaluation et des méthodologies d'évaluation, au manque de soutien de la part des cadres opérationnels percevant l'évaluation comme une perte de temps, au manque de clarté des objectifs définis pour l'évaluation, au fait que les évaluations sont souvent conçues après-coup alors qu'elles devraient être pensées en amont de la formation pour être réellement utiles (Patrick, 1992 rapporté par Beech et Leather, 2006 ; Phillips et al., 2004). Phillips et al. (2004) soulignent de plus qu'il existe trop de variantes parmi les modèles disponibles et que ces modèles seraient parfois trop complexes à mettre en pratique pour un novice⁴. De plus, les acteurs de la formation craindraient qu'une utilisation inadéquate des résultats soit possible. Enfin, les professionnels de la formation parviendraient rarement à intégrer ces évaluations, de manière durable, dans leur système de formation.

Certaines contributions quasi-expérimentales effectuées ponctuellement dans divers domaines de recherches indiquent qu'il est pourtant possible d'effectuer de telles évaluations. Ces études montrent notamment que les effets bénéfiques des formations peuvent être observés et mis à jour à différents niveaux.

3. Quelques études illustratives des principales dimensions pouvant être évaluées

Les études explorant de façon quasi-expérimentale les effets possibles d'une formation donnent quelques indications sur la manière ou les diverses manières dont il est possible de procéder. Elles montrent notamment que les bénéfices d'une formation peuvent souvent être plus variés qu'initialement escomptés avec de simples mesures de satisfaction.

Par exemple, Lin et al. (2004) ont montré qu'une formation initialement orientée vers un but clair et précis (en l'occurrence améliorer l'assertivité dans les relations professionnelles) pouvait avoir d'autres répercussions positives (à savoir une amélioration de l'estime de soi personnelle). La formation dispensée peut donc modifier des attitudes à l'égard de soi-même en les rendant plus positives. Connaître de tels effets permet alors de recadrer les objectifs de la formation dans le sens d'un développement professionnel et personnel.

Des évaluations de formation ont été effectuées en fonction de différentes mesures: subjectives ou au contraire objectives et leurs évolutions durant la formation. C'est notamment le cas d'études visant à évaluer l'intérêt de formations dans le cadre de la prévention de la violence des usagers (Gertz, 1980; Lehmann et al., 1983; Goodykoontz et Herrick, 1990 cités par Beech et Leather, 2006). Dans ces études, des mesures avant et après formation étaient mises en place afin de vérifier une diminution du nombre d'incidents répertoriés (mesures objectives avant et après la formation), ainsi que l'évolution de variables attitudinales (mesures subjectives). Les résultats ont permis de constater, dans les trois études, une diminution du nombre d'incidents alors que

⁴ Concernant ces problèmes méthodologiques, McCain (2005) a établi une liste de biais pouvant avoir un impact sur les résultats observés et auxquels un professionnel de la formation ne pense pas toujours. On peut citer le biais d'échantillon (sélection d'un échantillon non représentatif ou encore de trop petite taille), biais dans les entretiens (s'assurer de la bonne compréhension des questions par exemple) ou encore biais d'acquiescement ou de neutralité en fonction de la présentation des questions.

les changements étaient plus mitigés en ce qui concerne les mesures subjectives à l'égard de son

⁵ L'auteur de la première étude explique ces différences de résultats par le fait que les participants s'y inscrivaient avant tout dans le but de toucher les assurances chômage et non dans le but de retrouver un emploi rémunérateur.

travail (satisfaction ou épuisement professionnel).

Keith et Frese (2005) se sont centrés sur un autre impact possible de la formation: le transfert d'apprentissages. Ils ont ainsi constaté que les formations encourageant à faire des erreurs, en opposition à celles où il convient d'éviter les erreurs, induisaient un transfert d'apprentissage plus important alors qu'elles ne se distinguent pas ou peu en termes d'apprentissages immédiats.

Enfin, d'autres se sont intéressées à évaluer des formations en fonction de leur objectif ultime: le résultat financier pour le stagiaire, comme son salaire après la formation (Regnér, 2002; Rosholm et al., 2007). Des différences d'impact surprenantes entre les formations sont ainsi constatées. Par exemple, dans la première de ces études, un effet négatif de la formation sur le salaire a ainsi été observé – effet pour le moins inattendu – alors que dans la seconde, le salaire avait augmenté de 20 % en moyenne après la formation⁵.

Il ressort de ces études que les auteurs ont recours à des méthodologies quasi-expérimentales issues de la recherche afin d'évaluer les formations (recueil et comparaison des mesures avant et après formation, comparaison des effets avec des groupes témoins, utilisation de techniques statistiques pour éliminer certains risques d'erreur de représentativité statistique). Ces études explicitent une grande variété d'effets ou d'objectifs possibles de la formation qui sont autant d'impacts bénéfiques pouvant justifier l'existence d'une formation: attitudes, apprentissages, comportements et transfert de performance, résultats objectifs et ultimes. L'évaluation des formations n'est pas donc seulement utile pour collecter des preuves de la satisfaction des stagiaires (ce à quoi on se limite souvent dans la pratique) mais, une fois bien conçue, elle devient un outil indispensable pour asseoir durablement les effets bénéfiques d'une prestation de formation, les optimiser, démontrer et renforcer l'utilité réelle de la formation.

Ce sont ces méthodologies et cette variété de mesures qu'on retrouve au sein des modèles d'évaluation de formation qui, ainsi, ne s'arrêtent pas à la seule prise en compte de la satisfaction des stagiaires mais promeuvent une vision plus exhaustive de l'impact des actions de formation. Pour cette revue, nous avons sélectionné les modèles les plus fréquemment cités, présentant une certaine complémentarité et étant suffisamment élaborés en termes de définition des concepts pouvant être évalués.

4. Revue des principaux modèles d'évaluation

En amont de la plupart des tentatives de conceptualisation de ces mesures, on retrouve le modèle de Kirkpatrick (1959) à quatre niveaux d'impacts de la formation. Ce modèle a, par ailleurs, l'avantage de synthétiser le processus complexe de l'évaluation des formations et propose une démarche rationnelle et suffisamment large (mais synthétique) pour correspondre aux besoins des professionnels de la formation (Bates, 2004). Il s'agit du modèle le plus utilisé par les professionnels de la formation mais aussi par les chercheurs travaillant sur l'évaluation des actions de formation (Santos et Stuart, 2003). Nous présenterons par la suite les amendements et compléments essentiels faits à ce modèle qui demeure la référence.

4.1. Le modèle de Kirkpatrick à quatre niveaux d'impacts

Le modèle de Kirkpatrick, dont la première ébauche de référence date de 1959, est un outil incontournable pouvant s'appliquer à l'évaluation de formations quelles qu'elles soient.

Il comprend quatre niveaux correspondants à des niveaux d'impact complémentaires. Les trois premiers niveaux concernent le stagiaire: le premier niveau concerne la satisfaction du stagiaire (ce que le stagiaire a apprécié), le second, les connaissances acquises (ce qu'il sait et a appris), le troisième concerne les changements induits dans la réalisation des tâches (ce qui a changé dans son comportement). Ces trois niveaux sont complétés par un dernier niveau d'impact externe au stagiaire (ce que la formation apporte à l'organisation).

Le premier niveau, appelé «réaction», s'intéresse à la satisfaction des participants suite à la formation et ce sur plusieurs aspects de la formation (tels que les objectifs, le contenu, les méthodes pédagogiques, le style d'enseignement, le matériel mis à disposition). Cette satisfaction est appréhendée sous forme de questionnaires, baptisés à juste titre *happy sheets*. Il s'agit du niveau le plus fréquemment utilisé et des nombreux exemples adaptables à tout type de formation sont facilement accessibles sur Internet⁶. Le deuxième niveau mesure «l'apprentissage» des participants en terme de connaissances, compétences et attitudes acquises lors de la formation. Cette mesure s'effectue le plus souvent par le biais de questionnaires ou d'autres systèmes d'évaluation systématisée (par exemple des examens). Le troisième niveau évalue «les changements comportementaux» dus à la formation et le transfert d'apprentissage. Il s'agit de repérer les connaissances, compétences et attitudes acquises lors de la formation qui donnent effectivement lieu à une réutilisation dans la pratique professionnelle. Cette mesure est, la plupart du temps, réalisée par questionnaires ou entretiens et peut être opérée à plusieurs reprises (au début, en fin et quelques temps après la formation). Le quatrième niveau est le niveau des «résultats» correspondants au calcul de divers indices élaborés en lien avec les objectifs de la formation et définis lors de la conception de cette dernière (Tableau 1).

Ce modèle a l'avantage d'évaluer plusieurs objectifs de la formation permettant ainsi d'appréhender la complexité de l'action de formation (Bates, 2004). Il a de plus ouvert la voie à de nombreux autres modèles. Du point de vue statistique, Arthur et al. (2003) ont montré, dans une méta-analyse basée sur plus de 150 publications sur l'évaluation de formation, que les différents niveaux du modèle de Kirkpatrick sont effectivement affectés par la formation. En prenant ainsi en compte une grande diversité de situations de formation, ils spécifient néanmoins que le niveau le plus facilement affecté est le niveau des apprentissages et le moins celui des réactions en terme de satisfaction, alors que c'est pourtant à celui-ci qu'on pense le plus spontanément.

En dépit de la qualité du modèle principes de Kirkpatrick, plusieurs limites ont été relevées tant au niveau pratique que conceptuel. Au niveau pratique, il traiterait insuffisamment du contexte dans lequel a lieu la formation (prise en compte des besoins de formation, évaluation de la rentabilité, positionnement des acteurs; Bates, 2004). Il serait peu utile pour prendre une décision quant à l'arrêt, la reconduite ou, surtout, l'optimisation d'une formation et il demeure vague quant aux mesures à effectuer précisément (Kraiger et al., 1993 ; Kraiger et Jung, 1997). Enfin, il ne permettrait pas de différencier les effets «à chaud» et «à froid» (Beech et Leather, 2006; Arnold, 2005 ; Bates, 2004 ; Kraiger et Jung, 1997). Ces trois points seront ici présentés ainsi que les amendements proposés.

⁶ On trouvera par exemple sur Internet des questionnaires de satisfaction mis en ligne par l'université de Lausanne. Pour la formation professionnelle, voir le site http://www.cedip.equipement.gouv.fr/rubrique.php3?id_rubrique=83, fiche n° 39 pour les points pouvant faire l'objet d'une évaluation, ainsi que l'ouvrage d'Hadji (1992).

Tableau 1

Modèle de Kirkpatrick (1959, 2007a).

Niveau	Concept	Question traitée	Méthode et indicateurs
Réaction	Satisfaction	Quelle perception les participants ont-ils de la formation ?	<i>Formulaires d'évaluation du programme, questionnaires et entretiens concernant la satisfaction quant à :</i> l'organisation pédagogique les supports l'animation de la formation, etc. <i>et plus rarement des avis sur :</i> la densité des contenus la maîtrise des contenus la transférabilité des savoirs, etc.
Apprentissage	Connaissances	Ont-ils acquis les connaissances et compétences souhaitées lors de la formation ?	Examens de connaissance, Exercices traduisant une connaissance. Observation et entretiens, si possible avant et après la formation Autoévaluation par le stagiaire ou entre pairs. Observations par le formateur. Mesures et entretiens par le gestionnaire de formation ou le commanditaire (nécessité de définir les connaissances visées)
Comportement	Application de l'apprentissage dans les comportements	L'apprentissage des participants a-t-il changé leur comportement ?	Idem en se basant sur des savoir-faire Grille d'observation comportementale À effectuer après quelques mois (Kirkpatrick et Kirkpatrick, 2007b) (nécessité de définir les savoir-faire visés)
Résultats	Application dans l'activité visée ou répercussions pour le commanditaire	Y a-t-il un changement dans les pratiques et ce changement amène-t-il les résultats escomptés ?	Indicateurs de productivité, de rentabilité, de turnover et de coûts humains (nécessité de définir les indicateurs visés)

4.2. Évaluer la formation au regard de son adéquation au contexte: besoins, rentabilité, mesures des apprentissages et objectifs divergents des acteurs

Afin d'appréhender les éléments du contexte de formation, Warr, Bird et Rackham (1970) proposent quant à eux un modèle baptisé le CIRO (évaluation du Contexte, évaluation des ressources ou Input, évaluations des Réactions, évaluation des résultats ou Output)⁷. L'originalité

⁷ Le niveau réaction (R) demeure inchangé comparé au modèle de Kirkpatrick (1959, 2007a). L'évaluation des résultats (output) se subdivise lui en trois sous-niveaux renvoyant plus ou moins aux trois autres niveaux du modèle de Kirkpatrick s'intitulant, respectivement, résultats immédiats, intermédiaires et ultimes en fonction des temps de mesures auxquels il est le plus pertinent d'effectuer ces mesures.

de ce modèle réside dans l'ajout de deux étapes, utiles à mettre en place pour le gestionnaire de formation, afin d'évaluer la qualité de la formation qui sera mise en place ou afin de l'optimiser par la suite. Il s'agit, d'une part, de l'évaluation des ressources disponibles (en termes de temps, de moyens et de connaissances) et, d'autre part, de l'évaluation du contexte qui permet une identification des besoins de formation. Cette étape d'identification des besoins rejoint la procédure d'évaluation proposée par Spilsbury (1995) qui insiste sur l'identification des besoins de formation lors de sa conception. Le mot identification sera remplacé par la suite par le terme générique évaluation (Arnold, 2005). En effet, alors que certains auteurs (Huteau, 2003; Arnold, 2005) considèrent l'évaluation des besoins de formation comme une étape indépendante de l'évaluation des formations, Warr et al. (1970) considèrent que cette identification préalable des besoins et des moyens⁸ est un élément important de la qualité des formations, à l'instar de la pratique des grandes entreprises (Dixon, 1996).

Afin de prendre en compte la logique économique des commanditaires, Hamblin (1974), s'appuyant sur les deux modèles précédents, propose de plus de distinguer deux aspects au sein des outputs (mesure du niveau ultime des «résultats») : l'analyse des retombées de la formation sur l'activité du salarié et l'analyse de rentabilité économique de la formation. D'autres auteurs (Phillips, 1997; McCain, 2005) présentent alors un indice de retour sur investissement, le ROI. Il se calcule en divisant le bénéfice net du programme par le coût du programme (le bénéfice net du programme est obtenu en soustrayant au bénéfice brut du programme son coût). Il s'agit d'un calcul de rentabilité classique emprunté au domaine de l'économie, qui paraît simple à effectuer (notamment lorsque la finalité de l'action de formation est ouvertement financière et à court terme). La difficulté provient néanmoins du fait qu'il faut identifier les gains et coûts induits par la formation et les convertir approximativement en valeurs monétaires à moyen terme (Phillips, 1997)⁹. Cela peut néanmoins s'avérer problématique lorsque les objectifs de la formation correspondent à des éléments plus intangibles et difficilement convertibles en unités financières.

D'autres contributions abordent le problème du choix des mesures à effectuer. Ainsi, Marth (1994), s'intéressant aux niveaux des apprentissages et du changement comportemental, soutient qu'il convient de distinguer les apprentissages «subjectifs» et les apprentissages «objectifs», tant au niveau des connaissances que du transfert comportemental. Ce point de vue est d'ailleurs partagé par McCain (2005) et Brown et Sturdevant Reed (2002). McCain (2005) différencie l'amélioration des connaissances, compétences et habiletés et leur identification ou démonstration effective dans le cadre de la formation. Il apporte, de plus, un nouvel élément à prendre en compte pour la mesure du transfert: l'environnement de travail qui peut faciliter ou entraver

⁸ Bournazel (2005) souligne également le rôle des variables externes à la formation qui peuvent influencer les résultats de cette dernière. Il propose ainsi un outil intitulé le «pentagone de la formation» permettant l'évaluation des organismes de formation. Cet outil repose sur la présence de cinq indicateurs: les indicateurs de résultats classiques, les indicateurs sociaux (prenant en compte les contraintes imposées aux organismes de formation), les indicateurs d'accompagnement (actions d'accompagnement et de suivi indépendantes de la formation mais ayant un impact sur son efficacité), les indicateurs d'innovation et les indicateurs économiques et financiers.

⁹ On trouvera des éléments pour le calcul sur investissement dans quelques dossiers assez complets de l'UQAM: <http://www.observatoiregrh.uqam.ca/index2.php?idcategorie=41>.

Il n'en demeure pas moins que l'estimation du simple coût d'une action de formation en soi est un exercice difficile, voire risqué. En effet, il faut alors prendre en compte et estimer les coûts directs et/ou indirects, les frais de fonctionnement, de structure, d'équipement parfois destiné à plusieurs usages et pouvant enfin donner lieu à d'autres bénéfices. Ainsi, selon le mode de calcul retenu, le coût horaire en formation de l'auteur principal de l'article est multipliable (ou divisible) par 37 selon les éléments retenus ou non retenus; ce qui représente une marge d'erreur possible considérable!

l'utilisation des nouvelles connaissances, compétences et habiletés sur le lieu de travail alors qu'elles sont acquises. Certains auteurs soulignent l'importance de considérer l'environnement et les aspects situationnels lors d'une évaluation (Lupton et al., 1999; Brown et Sturdevant Reed, 2002). Des éléments situationnels circonstanciels ou relativement stables peuvent ainsi perturber ou favoriser la possibilité de démontrer des résultats attendus. Ces éléments suggèrent que, lors d'une évaluation mais aussi lors de l'implémentation d'une action de formation, il convient de différencier les objectifs en terme d'apprentissages perçus et/ou effectifs, les acquis et leur transférabilité dans les situations de travail effectivement rencontrées par les stagiaires.

Enfin, un autre élément important du contexte concerne les finalités des acteurs de la formation. Michalski et Cousins (2001) soulèvent une problématique sous-jacente à toute évaluation de la formation. Ils montrent qu'il existe des points de vue et des finalités poursuivies très différentes de la part des acteurs sociaux de la formation (stagiaires, commanditaires, formateurs prestataires). Les stagiaires espéreraient un développement de leur compétence et de leur carrière ainsi qu'une meilleure adéquation de l'offre de formation à ces besoins. Les cadres ayant la responsabilité du budget de formation voudraient promouvoir une certaine attractivité et crédibilité auprès des salariés des programmes de formation qu'ils financent. Les fournisseurs de la formation, qui l'ont conçue ou mise en pratique, souhaiteraient quant à eux voir la pratique de formation maintenue ou développée et donc mise en valeur auprès des autres acteurs. Les auteurs insistent sur le fait que ces différents points de vue doivent être recueillis, pour une meilleure analyse de la variété des utilités desservies dans les faits par la formation et son évaluation.

En résumé, les principaux compléments apportés au modèle de Kirkpatrick concernent:

- l'identification des besoins et des ressources;
- l'analyse de la rentabilité ;
- la distinction entre apprentissage perçu et effectif, les possibilités de transfert en situation de travail;
- le recueil des points de vues des acteurs du système de formation en ce qui concerne leurs objectifs et leurs attentes.

4.3. Évaluer la formation au regard des caractéristiques des apprentissages et de leur variété

Pour répondre à la critique formulée à l'encontre du modèle de Kirkpatrick, selon laquelle ce modèle n'explicite pas suffisamment ce qui doit être évalué, Kraiger et al. (1993) proposent une typologie des résultats d'apprentissage en s'appuyant sur les théories et les mesures de l'apprentissage. Les résultats d'apprentissage sont, selon ces auteurs, multidimensionnels et les répertorier dans diverses catégories d'apprentissage aurait l'avantage d'aider à choisir les techniques d'évaluation les plus appropriées (cf. p. 323). Les objets de l'apprentissage pourraient ainsi être ainsi classés en trois catégories (indépendantes à la différence du modèle de Kirkpatrick) :

- les résultats «cognitifs» issus de l'évaluation des connaissances acquises ;
- les résultats «conatifs» basés sur l'actualisation de compétences par des comportements concrets ;
- les résultats «affectifs» concernant l'expression d'une valorisation par le stagiaire de divers objets hormis la formation en général.

Ce modèle différencie donc les acquis en fonction de catégories correspondant aux trois composantes des «attitudes» :

- la composante cognitive;
- la composante conative ;
- la composante affective.

Chacune de ces catégories peut alors être précisée et mesurée de différentes façons et ce en fonction du degré d'élaboration des acquis (c'est-à-dire, en fonction du niveau d'expertise déjà atteint par les stagiaires et que l'on pourra évaluer avant, pendant et après la formation à des fins pédagogiques).

Ainsi, les résultats cognitifs vont d'une simple information sur un sujet précis à une connaissance aisée et stratégique sur le même sujet, c'est-à-dire maîtrisée et réutilisable. Les résultats cognitifs comprennent:

- tout d'abord une connaissance déclarative: ces connaissances isolées seraient évaluable dans des tâches de reconnaissance ou de rappel basiques: QCM, QROC à durée limitée si l'on cherche à évaluer l'accessibilité des connaissances ou à durée non limitée si l'on cherche à évaluer l'exactitude de ces connaissances;
- une connaissance organisationnelle: capacité à présenter les informations de façon structurée (capacité à différencier les concepts clefs), reliée (montrer les liens qu'entretiennent ces concepts entre eux) et hiérarchisée (mettre en avant les concepts les plus déterminants) comme le ferait le formateur, lors d'un exposé;
- une stratégie cognitive (capacité à proposer une solution et à la justifier au regard d'autres alternatives dans le cadre d'une étude de cas).

Les résultats conatifs peuvent également être subdivisés en différents niveaux liés à l'acquisition des compétences:

- de l'amas de compétences isolées (maîtrise de procédures simples) ;
- en passant par la compilation de plusieurs procédures (capacité à les hiérarchiser temporellement, sans faire d'erreurs ou sans produire de gestes inutiles, capacité à adapter la procédure à une nouvelle situation) ;
- jusqu'à l'automatisme qui traduit le fait que la performance est fluide, accomplie et individualisée (la tâche pouvant être accomplie rapidement qu'elle soit ou non réalisée simultanément à une autre, ou qu'il existe ou non des informations ambiguës).

Enfin, les résultats affectifs renvoient aux changements motivationnels et aux changements attitudeaux. Les changements motivationnels peuvent renvoyer aux dispositions à l'égard d'un objet d'apprentissage: motivations, évaluation de son efficacité personnelle dans la réalisation de la tâche et motivation envers le but lui-même (évaluables à l'aide d'échelles psychologiques spécifiques de motivation, de sentiment d'efficacité personnelle ou d'entretiens structurés). Les changements attitudeaux correspondent, selon l'objectif de la formation, à l'acquisition des normes de l'équipe, à l'implication organisationnelle, aux attitudes vis-à-vis du travail ou de certaines procédures de travail qui sont nouvelles et aux attitudes vis-à-vis de l'organisation dans son ensemble. Elles peuvent être évaluées par le biais d'échelles d'opinion visant à déterminer le degré d'élaboration de l'attitude, la centralité de l'attitude pour le sujet et l'internalisation de celle-ci.

Pour les auteurs, ces motivations et attitudes peuvent tout aussi bien être des précurseurs que des indicateurs de l'apprentissage (lorsqu'elles évoluent positivement) et faire partie d'objectifs explicites (acquisition d'une citoyenneté organisationnelle, d'une éthique professionnelle, objectif de remobilisation des salariés, objectif de cohésion d'équipe, de développement de l'assertivité, etc.) ou bien n'être qu'un résultat incident (la formation pouvant accroître les attentes de réussites des stagiaires ou leur sentiment d'efficacité personnelle du simple fait des acquis de l'apprentissage) (Tableau 2).

Les auteurs fournissent ainsi de nombreuses idées de mesures, permettant au formateur construisant une formation, d'identifier précisément des objectifs pédagogiques à moyen et court termes, de mettre en place une stratégie pédagogique assortie de mesures correspondantes (pour l'évaluation des acquis) et de tester de nouveaux exercices en cours de formation. Pris isolément, ce modèle se concentre surtout sur les acquis pédagogiques individuels et non sur la place de la formation au sein de l'organisation (Kraiger et Jung, en 1997, feront eux-mêmes remarquer qu'il ne permet pas d'identifier la diversité des «résultats» ou de déterminer le retour sur investissement lié à la formation). Par ailleurs, seules des mesures à froid permettraient de mesurer ce que les stagiaires ont essentiellement retenu de la formation, quelques semaines après. Ce problème est résolu par la synthèse de Beech et Leather (2006).

4.4. Le modèle de Beech et Leather: évaluer les impacts «à chaud» et «à froid»

Beech et Leather (2006), dans une modélisation plus récente, tentent de palier aux différentes critiques adressées aux modèles précédents. Ils proposent alors de compiler les contributions des diverses approches en un seul modèle. Ce dernier modèle est constitué de cinq catégories: réaction, apprentissage ou niveau d'évaluation immédiat, comportement ou niveau d'évaluation intermédiaire, résultat ou niveau d'évaluation ultime et, enfin, niveau financier. Ces catégories sont issues des modèles de Kirkpatrick (1959), Warr et al. (1970), ainsi que de Hamblin (1974). Les niveaux «apprentissage» et «comportement» sont divisés en trois sous-niveaux correspondant au modèle de Kraiger et al. (1993) à savoir «connaissance», «compétence» et «affectif» mais faisant cette fois l'objet d'une mesure en deux temps.

Les auteurs fournissent également des illustrations pour évaluer les différents niveaux et sont extrêmement précis quant au moment où effectuer les diverses mesures dans ce modèle (Tableau 3).

Le principal intérêt du modèle de Beech et Leather (2006) est de spécifier les moments les plus pertinents pour effectuer les mesures consécutives à la formation: les réactions et les acquis au niveau immédiat à la fin de la formation, mais aussi lorsque la personne est retournée en situation de travail ou s'y trouve pour la première fois. Les résultats et retours sur investissement seront évalués à plus long terme (pour cette mesure, comme pour le transfert d'apprentissage d'ailleurs, il faut du temps pour une estimation précise notamment du fait du temps nécessaire à l'apparition des effets et pour mieux prendre en compte d'éventuels effets indirects qui remettraient en cause les premières conclusions). Ce modèle insiste par ailleurs une indépendance possible des différents effets et sur la nécessité d'évaluer des changements et non des états (ces changements représentant les véritables impacts de la formation). Toutes ces mesures doivent donc se faire par comparaison avec un état initial mesuré avant formation ou au tout début de celle-ci. La plupart des auteurs s'accordent également sur la nécessité d'un groupe témoin sans formation. En effet, certaines variables contextuelles peuvent affecter les variations (inflexion des attitudes liées à des réformes en cours, difficulté de transfert des apprentissages du fait du modification

Tableau 2

Classification des résultats multidimensionnels de Kraiger et al. (1993).

Dimensions et concepts	Mesures préconisées et éléments à prendre en compte
<i>Les résultats cognitifs sont composés</i> D'une connaissance verbale: connaissance déclarative (information à propos de quoi), connaissance procédurale (information à propos de comment) et connaissance stratégique (information à propos de qui, quand et pourquoi) D'une métaconnaissance : connaissance organisationnelle (développement d'une structure de significations permettant l'organisation des connaissances) Et d'une stratégie cognitive (développement et application des stratégies cognitives)	Tests de rappel ou de reconnaissance pour vérifier l'étendue des connaissances (questionnaires à réponses obligatoirement courtes, questionnaires à choix multiples ou à réponses vrai/faux) en temps limité pour en vérifier l'accessibilité ou en temps illimité pour en vérifier l'exactitude A utiliser en fin mais aussi en début de formation pour des raisons de feed-back et de variabilité des mesures Rapidité d'analyse des incidents critiques (études de cas) Assemblage d'éléments de connaissance (carte mentale) évalué par comparaison à un standard d'expert ou de formateur Exercices de résolution de problème avec justification des alternatives en terme d'utilité L'exactitude se mesurant par comparaison avec les réponses d'un expert. L'expertise pouvant également s'estimer par l'abandon plus rapide d'une alternative inopérante, une plus grande pertinence dans l'estimation du temps nécessaire ou de la difficulté du problème. De même la performance attendue ou la quantité perçue de ré-apprentissages nécessaires peuvent être des indicateurs valables même s'ils sont autorapportés
<i>Les résultats comportementaux ou en termes de compétence</i> La compilation de savoir-faire élémentaires dans un domaine spécifique pour aboutir à des productions comportementales plus complexes L'automatisme traduit le fait que la performance est fluide, accomplie et individualisée, elle implique une adaptation des modes opérationnels	Tests d'aptitude Observation comportementale ciblée (Grille) basée sur l'analyse des comportements utiles ou inutiles, la rapidité et le nombre d'étapes mises en place Entretiens structurés concernant l'activité (en cas de matériel coûteux) La résistance à de nouvelles méthodes pour réaliser une activité peut traduire une compilation, de même que la transposition spontanée de la procédure dans une tâche non apprise Performance peu ou moins affectée par une mise en situation de tâches multiples Performance moins affectée par mises en situation avec problèmes interférents Mesures indirectes comme: moindre variabilité du niveau de performance sur des tests variés, baisse de rapidité lorsqu'il est demandé de porter son attention sur ce que l'on fait, capacité à remettre en ordre un script d'exécution des tâches et à repérer les étapes clefs

Tableau 2 (Suite)

Dimensions et concepts	Mesures préconisées et éléments à prendre en compte
<i>Pour finir, les résultats affectifs sont déterminés par</i> Les résultats attitudeaux (ciblés par la formation: importance des relations humaines dans le management, éthique professionnelle, acceptation des normes et valeurs centrales, implication organisationnelle, internalisation liée à un modelage comportemental.) Et les résultats motivationnels : les dispositions motivationnelles, l'autoefficacité (efficacité perçue en vue de l'exécution d'une activité particulière) et les buts fixés (importance des buts dans les processus motivationnels)	Questionnaires d'attitudes prenant en compte la direction et la force de l'attitude (accessibilité, centralité et internalisation) Il est important de faire des mesures avant et après formation, notamment en ce qui concerne la force de l'attitude Questionnaires de motivation et d'orientation vers la maîtrise ou la performance Questionnaire de sentiment d'efficacité personnelle, de confiance en soi quant aux savoir-faire Questionnaire ou entretien sur des objectifs avec analyse de leur difficulté, de leur hiérarchisation, de leur indépendance et de leur spécificité

du poste de travail, etc., Arnold, 2005)¹⁰. Les auteurs reconnaissent donc la nécessité de prendre en compte, pour l'interprétation des résultats, les effets contextuels (évaluation des besoins et moyens, différences d'objectifs entre les acteurs, importance de la situation de travail dans la possibilité de transfert des apprentissages.) bien qu'ils ne proposent pas de mesures spécifiques pour ces éléments.

5. Discussion

Passer en revue la littérature concernant l'évaluation des formations permet de prendre conscience de la diversité des impacts possibles de la formation et des raisons pour lesquelles l'action de formation n'amène pas toujours les résultats escomptés par les uns ou les autres. Hormis les déclinaisons conceptuelles de ce que peuvent être les impacts d'une formation, cette littérature fournit des pistes pour mettre en place des procédures d'évaluation et des mesures de ces impacts. Elle soulève alors certaines interrogations tant au niveau de la recherche – quelles peuvent être les relations entre les diverses variables ? – qu'au niveau pratique – comment mettre en place de telles mesures dans le contexte de la formation ?

En termes de recherches on constate, par ailleurs, que les psychologues comme Kraiger, à la différence des gestionnaires, appréhendent le sujet sous un angle essentiellement intra-individuel lié aux problématiques d'apprentissage. Quelle peut alors être alors l'incidence de telles recherches et quelles autres problématiques reste-t-il à traiter ? Au niveau pratique, comment utiliser ces résultats en vue de l'amélioration d'une formation ? Qui doit, ou peut, entreprendre l'évaluation de la formation et comment prendre en considération les variables individuelles et contextuelles ?

¹⁰ Spector (2006) recommande à ce sujet de n'envoyer en formation, dans un premier, que la moitié des salariés pour lesquels cela paraît nécessaire. Cela permet d'évaluer la formation en tant que telle, indépendamment des variables individuelles. Ceci peut être fait par tirage au sort sur un échantillon conséquent ou en contrôlant l'équivalence stricte des deux moitiés quant à d'autres variables (poste, niveau initial, besoin de formation).

Tableau 3

Modèle combiné de Beech et Leather (2006) adapté à des formations de gestion de la violence des usagers.

Niveau de mesure	Concepts	Outils
Réactions immédiates	Satisfaction	Questionnaires et feuillets d'évaluation subjective sur l'intérêt, la pertinence et l'attrait des contenus...
«Apprentissage» Niveau d'impact immédiat	<i>Connaissance déclarative et organisationnelle</i>	Résultats de tests de connaissance, exercices type permettant de tester les connaissances déclaratives Résultats d'exercices de résolution de problèmes (actions à conduire, exercices de groupe, plan de résolution de problème, justifications...) pour tester la connaissance organisationnelle
	<i>Compétences dans la compilation de savoir-faire procéduraux</i>	Tests d'aptitudes Exercices de mise en situation collectifs et individuels, autoévaluations dans une situation antérieure, explications des vecteurs d'erreurs et de stress...
	<i>Affects</i> Attitudes et motivations Confiance Sentiment d'efficacité personnelle	Échelles psychologiques standardisées Autoévaluations de ses capacités quant aux objectifs, optimisme, satisfaction dans la tâche à réaliser
«Comportements» Niveau d'impact intermédiaire Éventuellement analysable par un hiérarchique direct ou indirect (lors d'un entretien annuel)	<i>Connaissances</i>	Tests identiques au niveau immédiat ou appréciation du supérieur
	<i>Compétences</i>	Performance lors d'incidents critiques: résultats ou retours en réunions et séances d'analyse de la pratique
	<i>Affects</i> Attitudes et motivations Confiance Sentiment d'efficacité personnelle	<i>Entretien structuré sur ces thèmes</i> Satisfaction au travail des salariés (groupes-test) Climat d'équipe Indicateurs RH de facilité de recrutement, d'absentéisme, facilité du travail d'équipe...
«Résultats» Niveau d'impact ultime	Productivité, climat et incidents critiques	Mesures de climat d'entreprise Nombre d'incidents, type et importance, Indicateur de satisfaction client (groupe tests et nombre de réclamation...)
Niveau financier	Rentabilité de l'action	Indicateurs des bénéfices ou indicateurs de variation des coûts (liés, par exemple en milieu hospitalier, aux réclamations et remboursements, aux indemnités ou à la prescription excessive de consommables de santé, etc.)

5.1. Au niveau pratique, le recueil et l'interprétation des résultats d'évaluation : évaluer quoi, comment et par qui ?

Les résultats d'une évaluation classique en terme de satisfaction nous permettent de préciser si les stagiaires sont satisfaits ou non à l'issue de la formation et sur quels points ils

En tant que tels, ces résultats n'indiquent pourtant ni si la formation est de qualité ou non, ni s'il faut l'améliorer, ni quelles modifications effectuer, et ce d'autant si l'évaluation est sommaire, imprécise ou ponctuelle. De la même façon pour ce qui est des résultats d'apprentissages (résultats d'examens), l'évaluation en fin de formation de ceux-ci spécifie rarement ce qu'il convient de modifier dans la formation ni dans quelle direction chercher (utiliser une pédagogie plus attractive, mettre en place de nouvelles techniques pédagogiques, améliorer la motivation, jouer sur le temps et le rythme d'apprentissage, favoriser une rétention à plus long terme...). Un modèle d'évaluation plus large, par le biais des mesures qu'il propose, permet de situer voir d'isoler les problèmes à résoudre. Ainsi le modèle de [Kraiger et al. \(1993\)](#) permet d'être assez exhaustif dans l'observation des résultats d'apprentissage et d'indiquer exactement où en est le processus d'apprentissage pour le resituer dans le cadre d'un modèle d'apprentissage empirique ou théorique. Une des principales qualités des modèles d'évaluation de formation, au travers du large spectre des mesures collectées, est qu'ils permettent donc de mettre en perspective les différents éléments pouvant intervenir et d'isoler précisément des points d'amélioration. Ce faisant, ils permettent de ne pas se fourvoyer sur des pistes vaines et usantes ou, à l'autre extrême, sans avoir à promouvoir une satisfaction client visant à masquer d'éventuelles insuffisances (ce qui tendrait à détériorer le sens du travail pour le formateur). La variété des dimensions évaluées, ainsi que la relative indépendance des résultats pouvant être observés, laisse donc entrevoir tout le gain que représente une mesure des impacts de la formation comparativement à une simple mesure de satisfaction ou des acquis en fin de dispositifs.

Une fois le(s) point(s) problématique(s) identifié(s), la capacité d'innovation du pédagogue ou de l'ingénieur pédagogique est alors à solliciter. Une solution au niveau méthodologique consiste à comparer les nouvelles modalités de formation aux anciennes (comparaison transversale de méthodes pédagogiques différentes et mesures longitudinales de la motivation des participants et de leurs acquis avant et après formation) en utilisant des indicateurs systématiquement identiques et des modèles d'évaluation constants. À ce propos, [Pershing et Pershing \(2001\)](#) ont montré que dans la pratique très peu de formateurs utilisaient spontanément des indicateurs précis et constants afin d'évaluer les réactions des participants. La méthodologie d'évaluation de formation relève donc d'une compétence spécifique et d'une rigueur d'observation qui, en contre partie, permet une identification plus précise et donc plus rapide et économique des points à améliorer et surtout une évaluation des contributions effectives et non supposées de la formation et de ses modifications successives.

Un des points les plus épineux de l'évaluation de formation concerne les enjeux des participants à la formation. À ce propos, [Dejean \(2002\)](#) énonce que «l'évaluation de la qualité dans l'enseignement ne peut être approchée que par la prise en compte d'avis d'une multiplicité d'acteurs (aux intérêts et aux enjeux différents [...])»¹¹. S'assurer de leur participation active peut permettre d'éviter une mise en place et un recueil bâclés ou encore à une «fin de non-recevoir»

¹¹ Les pratiques d'évaluation des universités sont généralement assez disparates. Une des pratiques les plus originales s'inscrit dans le contexte d'une démarche qualité (réalisation, évaluation et amélioration), à l'instar de Bordeaux-II. Elle se base sur des évaluations annuelles (questionnaires remplis par les étudiants) ainsi que des évaluations quadriennales comprenant une boucle de la qualité (analyse des besoins, finalités de l'enseignement, objectifs et programme, niveau des étudiants, modalités d'enseignement et des épreuves) et des échanges-débats où sont présents un président de séance, les enseignants concernés par l'évaluation, des enseignants extérieurs à l'enseignement évalué et des étudiants ([Dejean, 2002](#)). Cette conception a l'avantage de montrer que la qualité d'une formation dépend de plusieurs facteurs notamment d'une coproduction entre enseignants et étudiants. Une autre conception très particulière a été mise en pratique dans des universités espagnoles. Il s'agit de l'évaluation circulaire ([Dejean, 2002](#)). Circulaire car elle est réalisée par l'ensemble des acteurs concernés: étudiants, enseignants, rectorat et ministère. Les étudiants évaluent les enseignements. Les enseignants

des résultats de cette évaluation par les acteurs eux-mêmes. La nature et la gestion de ces enjeux sera nécessairement différente selon la personne qui réalise l'évaluation (formateur prestataire, commanditaire ou personne indépendante). À première vue, le formateur semble être le mieux placé pour mener à son terme de façon utile une telle évaluation. Il est au cœur de la formation et de l'action de formation. Il en a également apprécié toutes les étapes et a vu son déroulement sur le terrain quand bien même devrait-il faire le deuil de sa subjectivité et faire preuve de recul face à cette évaluation. Pour les formateurs, la pratique de l'évaluation de leur propre action de formation, ne peut être qu'un gain pour adapter leurs méthodes pédagogiques au public (par exemple en effectuant des mesures, en fin mais aussi en tout début de formation, sur les savoir-faire du public, ses connaissances et ses attitudes). Les employeurs des stagiaires ou hiérarchiques directs (et plus globalement les «commanditaires» de la formation) tout comme les formateurs connaissent les objectifs de la formation. Leur participation est donc essentielle pour définir les critères d'une évaluation de l'adéquation au besoin. Il demeure pourtant peu probable qu'ils obtiennent des retours quant à la formation qui soient parfaitement objectifs de personnes avec lesquelles ils entretiennent par ailleurs des rapports de subordination (rapports difficiles à occulter lors d'un «entretien annuel d'évaluation» avec les stagiaires par exemple). Au regard des retombées attendues sur le terrain, leur participation à l'évaluation du dispositif de formation s'avère pourtant indispensable pour définir ou redéfinir le besoin de formation, les mesures les plus adéquates pour rendre compte des objectifs ultimes, ainsi que pour mettre en place des moyens nouveaux pour les atteindre (Warr et al., 1970; Dixon, 1996; Spilsbury, 1995; Huteau, 2003; Arnold, 2005). L'intervention de personnes extérieures (consultant, psychologue du travail, ingénieur de formation spécialisé) est un gage d'objectivité pour l'évaluation, quoique ces personnes ne connaissent pas a priori les objectifs d'une formation spécifique pouvant s'inscrire dans un contexte et une histoire institutionnels. Elles auront certes plus de difficultés à interpréter les résultats d'évaluation d'une formation si elles ne sont pas informées par les autres acteurs des modalités pratiques et des contraintes liées à l'implantation de la formation. Hormis leurs expériences antérieures dans le secteur, leur rôle peut se justifier en terme de compétences pour la mise en place technique du dispositif d'évaluation (mise en place pouvant s'avérer fastidieuse pour le débutant) mais aussi en ce qui concerne l'animation d'un comité de pilotage impliquant les autres acteurs. Leur position extérieure, garante de neutralité, peut leur permettre d'amener les différents acteurs à s'impliquer et se coordonner dans l'amélioration de l'action de formation. Ils peuvent ainsi jouer un rôle de médiateur et en leur faisant expliciter et partager des perceptions souvent différentes des objectifs ou des utilités que dessert ou pourrait desservir la formation (Michalski et Cousins, 2001).

5.2. Quelques pistes de réflexion pour le psychologue ou la recherche en psychologie

Un premier enjeu pour la recherche actuelle concerne la complémentarité des différentes mesures d'évaluation. En effet, contrairement au postulat initial de Kirkpatrick (1959), les différents niveaux de son modèle d'évaluation de formation n'entretiennent pas nécessairement une relation hiérarchique (Bates, 2004). Par exemple, on pourrait supposer que la satisfaction

s'autoévaluent, évaluent les différents aspects des enseignements (climat, conditions de travail et moyens disponibles, fonctionnements des différents services, interaction avec les étudiants) et effectuent une métaévaluation (évaluation de l'utilité des évaluations faites par les étudiants). Le rectorat évalue l'enseignement des enseignants et le ministère évalue la recherche des enseignants (Dejean, 2002). Les différents acteurs donnent ainsi leurs avis en tant qu'experts sur différents aspects de la formation. L'évaluation circulaire permettrait donc de mieux appréhender et optimiser la qualité d'une formation.

(évaluation du premier niveau) soit une condition minimale mais non suffisante pour qu'il y ait apprentissage (évaluation du second niveau). Néanmoins, [Alliger et Janak \(1994\)](#) concluent dans leur méta-analyse que les liens entre «réactions» et «apprentissages» ou «transfert comportemental» sont faibles (respectivement, $r=0,07$ et $r=0,05$). Les niveaux «apprentissage» et «transfert comportemental» seraient partiellement liés ($r=0,13$) et pourrait partiellement prédire les «résultats» ($r=0,40$ et $r=0,19$). Les liens entre les différents niveaux sont donc relativement faibles et ces niveaux peuvent donc amener des résultats contradictoires entre eux, quant à l'efficacité d'une formation. Cette relative indépendance doit amener à réfléchir sur ce qui caractérise chaque niveau. Un autre problème relatif à la mesure de ces niveaux concerne l'utilisation de mesures dites subjectives (les réactions mais aussi l'autoperception par le stagiaire de ses acquis et de la transférabilité de ses acquis).

Selon [Kirkpatrick \(2007a\)](#), évaluer la satisfaction des stagiaires et considérer leur point de vue demeure essentiel, car cela permet de connaître l'état d'esprit des participants lorsqu'ils retournent sur leur lieu de travail. D'une façon plus générale, prendre en compte les différents niveaux d'impact et la perception qu'en a chaque acteur paraît justifié, selon qu'on se pose la question de :

- l'image commerciale de la formation;
- de sa qualité pédagogique effective;
- des changements effectivement induits dans les pratiques, par la formation, ou bien;
- de son utilité sociale pour le commanditaire (ou pour le stagiaire ou le prestataire).

Il est vraisemblable que les différents niveaux correspondent initialement à des points de vue d'acteurs différents. Mais il est également vraisemblable que chaque acteur donnerait une définition différente et choisirait des indicateurs opérationnels différents pour chaque niveau. Ces différences mériteraient d'être explicitées dans le cadre de la mise en place de procédure d'évaluation ([Michalski et Cousins, 2001](#)) et même d'être explorées, auprès des différents acteurs, en utilisant des procédures d'exploration des représentations sociales de la formation et des techniques de remises en causes ([Moliner et Tafani, 1997](#)).

Une autre question qui découle de l'indépendance des différents effets de la formation concerne le rôle des variables psychologiques individuelles. Ainsi, il n'est pas rare de voir des problématiques liées à l'articulation entre les différents niveaux que l'on tente d'expliquer par l'intervention de variables individuelles. Par exemple, les problèmes d'apprentissages sont parfois appréhendés par l'intervention chez les stagiaires de problème de motivation à l'apprentissage ou de croyance en leur efficacité personnelle ([Quinones, 1995](#) ; [Colquitt et al., 2000](#); [Tracey et al., 1995](#)). La facilité qu'il y a à obtenir des mesures rapportées par les stagiaires, regroupés en session de formation, n'est sans doute pas étrangère à ce recours à des mesures individuelles. Il peut également être tentant de voir dans les caractéristiques individuelles du public la cause des problèmes de formation. Plusieurs études dans le domaine de l'insertion ([Castrà, 1994, 2003](#)) ont par exemple mis en évidence une représentation commune des intervenants en insertion qui se traduit par une tendance à surestimer le rôle des causes dispositionnelles pour interpréter les problèmes rencontrés par les demandeurs d'emploi, au détriment d'explications référant au contexte social ou contexte de formation dans lequel les individus sont insérés. Un modèle quasiment personnalologique sous-tend ainsi les pratiques d'interventions et la tendance assez systématique des acteurs de la formation à placer la modification des croyances et des comportements individuels (autonomie, responsabilité, etc.) comme un préalable à toute autre acquisition. En ce sens, lorsque la recherche se focalise sur des mesures individuelles, elle contribue à son insu à cette vision dispositionnalisante.

De la même façon, il est bien possible qu'en basant l'évaluation des formations sur l'évaluation de réponses individuelles des stagiaires (leurs réactions affectives, leurs attitudes à l'égard de la formation ou leur situation de travail, leurs apprentissages, leurs transferts comportementaux) certaines pratiques ou démarches de recherche finissent par réduire la compréhension et l'interprétation des effets de la formation à l'évaluation des dispositions des stagiaires en formation (leurs motivations initiales, leurs aptitudes à mettre en œuvre les apprentissages, à faire face à des situations professionnelles plus ou moins impliquantes)¹². S'intéressant à la part des variables individuelles et contextuelles, Hertenstein (2001) rappelle la distinction qu'il convient de faire entre, d'une part, l'évaluation en soi de l'action de formation et, d'autre part, la recherche des mécanismes individuels pouvant expliquer pourquoi certains stagiaires ont atteints ou non les objectifs visés. L'utilisation de variables individuelles renvoyant à des dispositions, si elle peut être utile à la démarche de recherche, est vraisemblablement à proscrire dans le cadre d'une démarche d'évaluation de la formation. Par ailleurs, la question qui se pose ici concernant les caractéristiques individuelles des stagiaires peut également se poser pour ce qui concerne les caractéristiques individuelles des formateurs qui pourraient être évaluées à l'occasion d'une évaluation de la formation¹³. En se focalisant ainsi sur des variables individuelles, les recherches ou les pratiques en évaluation de formation risquent de légitimer les freins à l'évaluation de formation évoqués en introduction de cet article. Elles risquent également d'occulter le rôle des variables contextuelles et les effets plus tangibles de la formation, en ne considérant que les impacts perçus (*self-reports*) par une catégorie d'acteurs. À l'inverse, s'intéresser à l'articulation des enjeux des différents acteurs est plus susceptible d'aider ces acteurs à expliciter leurs enjeux de façon constructive et à mesurer de façon objective les impacts effectifs, pouvant ainsi contribuer à une amélioration de la formation.

Enfin, hormis la signification des dimensions d'impact et les différences de représentations des divers acteurs, un autre champ de recherche utile peut également concerner l'impact des représentations causales de la qualité de la formation sur l'action des partenaires en situation de formation. Ainsi, en soulignant chez les enseignants et les formateurs en insertion, le risque d'une interprétation dispositionnalisante, Monteil (1990), Gosling (1992) et Castra (1994) posent la question de la contribution de cette perception au cercle vicieux de l'échec scolaire et de l'échec en insertion. De plus, cette représentation semble également structurer la pratique de formation en plaçant la modification des croyances et des comportements individuels (autonomie, responsabilité, etc.) comme un préalable aux apprentissages. Par exemple, il n'est pas rare dans le secteur de la formation de chercher à sélectionner le public en fonction de ces qualités difficilement objectivables. Il n'est pas rare non plus que la demande plus ou moins implicite faite

¹² De la même façon que pour l'apprentissage, le problème de la possibilité de transfert des apprentissages sur le lieu de travail est parfois traité par une mesure de la motivation au transfert du stagiaire (Chiaburu et Day, 2006). D'autres variables individuelles sont également explorées pour expliquer des résultats d'évaluation parfois décevants, comme l'engagement au travail et l'implication organisationnelle (lesquelles auraient un impact sur les réactions des participants, les connaissances déclaratives et le transfert, toujours par l'intermédiaire de deux variables médiatrices l'efficacité personnelle et la motivation, Tracey et al., 2001). Enfin, certaines variables individuelles sont également évoquées pour rendre compte du transfert d'apprentissage et de l'efficacité d'une formation comme l'autorégulation et l'autocontrôle (Kehr et al., 1999), les attitudes de confiance envers la formation (Ryman et Biersner, 1975), les variables de personnalité et les cinq facteurs de personnalité majeurs (le Big-Five ; Colquitt et al., 2000).

¹³ Ainsi, les pratiques ou les recherches sur l'évaluation des styles et des aptitudes pédagogiques des formateurs posent les mêmes problèmes en termes d'utilisation et de décision: convient-il de mettre en place une procédure de sélection des stagiaires et des formateurs pour optimiser les effets des dispositifs, d'améliorer leur formation ou d'optimiser leur synergie ?

aux intervenants concerne la modification de ces attitudes ou encore que l'on affiche ouvertement la formation comme visant à une « remobilisation » ou un « développement personnel ». Dans le quotidien d'une formation, il s'agit alors bien souvent de persuader le stagiaire, orienté vers la formation de façon un peu rapide, de ses motivations pour ce qui lui est transmis, de son efficacité personnelle par rapport à l'objet d'apprentissage ou de sa capacité à transférer de façon autonome ses apprentissages (à défaut de résultats comportementaux tangibles ou objectivables, on tend à convaincre son public de ses acquisitions, voire de sa responsabilité dans les résultats ultérieurs). Cette modification des croyances est un résultat bien connu concernant le sentiment de contrôle général des stagiaires qui tend à évoluer positivement suite à la formation (qu'il s'agisse d'élèves, de salariés ou de demandeurs d'emploi; Beauvois et Le Poutier, 1986; Dubois, 1988; Dubois et Trognon, 1989; Gangloff et Sourisse, 1995; Bressoux et Pansu, 1998; Bressoux et Pansu, 2001 ; Pansu et al., 2003). Le sentiment de contrôle des stagiaires est parfois même utilisé comme un indicateur de l'effet d'une formation au même titre que les mesures de satisfaction (Hertenstein, 2001 ; Gist et al., 1989 ; Sharp et al., 1997 ; Inouye et al., 2001). Pourtant rien ne permet de penser que ces croyances générales soient utiles à développer; c'est-à-dire prédictives des apprentissages ou de transfert d'apprentissage. Ces attitudes peuvent donc très bien n'être que les conséquences d'un processus de formation (ou de socialisation normative) achevé et ne contribuent pas nécessairement aux autres objectifs de la formation (un point que Kraiger et al. questionnent quant à l'effet internalisant du modelage comportemental).

La recherche pourrait alors s'attacher à voir, d'une part, dans quelle mesure les sentiments des stagiaires évoluent suite à leur formation et, d'autre part, rechercher ce qui favorise une telle évolution (méthodes pédagogiques, acquis effectifs, démonstration des acquis ou attentes positives développées chez les stagiaires). Elles pourraient également vérifier dans quelle mesure les formateurs cherchent à transmettre de tels sentiments et s'ils amènent des changements subjectifs (réactions positives à la formation, sentiment de compétence) et/ou objectifs (transfert d'apprentissages et possibilités effectives de transfert)¹⁴.

Si l'on étend ces préoccupations à l'ensemble des acteurs (stagiaires et commanditaires), cela permettrait d'entrevoir quelles peuvent être les représentations de chacun en ce qui concerne la qualité de la formation, les comportements induits par ces représentations (choix d'orientation, choix d'investissement, etc.). Cela pourrait également éclairer la façon dont les acteurs se coordonnent dans l'action de formation de façon synchrone ou non.

En définitive, les divers modèles d'évaluation proposent un certain nombre de critères dans la lignée de celui de Kirkpatrick. Les différents critères ou niveaux évalués sont justifiés eu égard aux finalités pratiques de telles évaluations: valoriser la formation par une démarche qualité, optimiser les apprentissages et leur possibilité de transfert, en mesurer les retombées pour le commanditaire en fonction de ses objectifs, estimer la rentabilité financière de l'action formative pour le commanditaire et le prestataire. Néanmoins, au niveau de la recherche en psychologie, les évaluations de formation semblent être appréhendées de façon relativement individuelle (par le biais des réponses des stagiaires devant y répondre) en omettant de prendre en compte les points de vue des autres parties prenantes (les commanditaires ou les prestataires, avec lesquels les stagiaires sont en interaction et avec lesquels des divergences peuvent exister). Cette centration excessive

¹⁴ Quelques résultats (Gilibert et al., 2008) laissent supposer que les mesures psychologiques (estime de soi, contrôle personnel et sentiment d'efficacité personnelle) sont liées aux ressentis des étudiants (satisfaction et apprentissages perçus) sans pour autant être prédictifs des apprentissages évalués par un examen. De la même façon, chez des adultes en formation professionnelle ces sentiments sont liés à la satisfaction à l'égard de la formation sans pour autant être prédictifs de leur insertion professionnelle.

semble également favoriser une centration du propos sur les caractéristiques dispositionnelles propres aux stagiaires, lesquelles pourraient affecter les effets de la formation, au détriment de l'étude plus complexe de leur interaction avec les autres acteurs.

Références

- Alliger, J.M., Janak, E.A., 1994. Kirkpatrick's levels of training criteria: In: Craig Eric Schneier, Craig J. Russel, Richard W. Beatty and Lyod S. Baird (Eds.). *The Training and Development Sourcebook*, Second edition. Human Resource Development Press, Amherst, Mass. pp. 219–27.
- Arnold, J., 2005. Training. In: Arnold, J. (Ed.), *Work Psychology*. Prentice Hall, pp. 356–383.
- Arthur, W., Bennett, W., Edens, P.S., Bell, S.T., 2003. Effectiveness of training in organisations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology* 88, 234–245.
- Bates, R., 2004. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning* 27, 341–347.
- Beauvois, J.-L., Le Poutier, F., 1986. Norme d'internalité et pouvoir social en psychologie quotidienne. *Psychologie française* 31, 100–108.
- Bee, F., Bee, R., 1994. *Training needs analysis and evaluation*. Institute of Personnel and Development, London.
- 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, 27–43.
- Bournazel, A., 2005. *La formation professionnelle: gestion et évaluation, le pentagone de la formation*. SEFI, Paris.
- Bressoux, P., Pansu, P., 1998. Norme d'internalité et activités évaluatives en milieu scolaire. *Revue française de pédagogie* 122, 19–29.
- Bressoux, P., Pansu, P., 2001. Effet de contexte, valeur d'internalité et jugement scolaire. *L'Orientation scolaire et professionnelle* 30, 353–371.
- Brown, R.E., Sturdevant Reed, C., 2002. An integral approach to evaluating outcome evaluation training. *American Journal of Evaluation* 23, 1–17.
- Castra, D., 1994. Explications causales et modes d'exercice professionnel en travail social. *Cahiers internationaux de psychologie sociale* 22, 33–50.
- Castra, D., 2003. *L'insertion professionnelle des publics précaires*. Presses universitaires de France, Paris.
- Chiaburu, D.S., Day, R., 2006. Social cognitive factors influencing training motivation and training effectiveness. In: Saul Carliner (Ed.), *ASTD Research-to-Practice, Conference Proceedings*. American Society for Training & Development, USA, pp. 27–36.
- Colquitt, J.A., LePine, J.A., Noe, R.A., 2000. Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology* 85, 678–707.
- Dejean, J., (2002). *L'évaluation de l'enseignement dans les universités françaises. Rapport établi à la demande du Haut conseil de l'évaluation de l'école*, Haut conseil de l'évaluation de l'école, Paris, pp. 1–108. Pour information également disponible à l'adresse: <http://lesrapports.ladocumentationfrancaise.fr/BRP/024000165/0000.pdf>.
- Dixon, N., 1996. New routes to evaluation. *Training & Development* 50, 82–85.
- Dubois, N., 1988. The norm of internality: Social valorization of behavior and reinforcements in young people. *Journal of Social Psychology* 128, 431–439.
- Dubois, N., Trognon, A., 1989. L'apport de la notion de norme d'internalité à l'approche des pratiques de formation. In: Beauvois, J.-L., Joule, R.-V., Monteil, J.-M. (Eds.), *Perspectives cognitives et conduites sociales*, 2. Delval, Cousset, pp. 213–224.
- Gangloff, B., Sourisse, M., 1995. Influence de la durée du chômage et des stages chômeurs longue durée sur l'évolution du conformisme à la norme d'internalité. *Psychologie et psychométrie* 16, 5–19.
- Gertz, B., 1980. Training for prevention of assaultive behaviour in a psychiatric setting. *Hospital and Community Psychiatry* 31, 628–630.
- Gilibert, D., Genty, D., Gillet, I. & Marchand, M. (2008). *Évaluer l'efficacité des dispositifs de formation professionnelle. Premières journées de l'Association française de psychologie du travail et des organisations*, Paris. Disponible sur <http://asso.univ-lyon2.fr/afpto/IMG/pdf/GILIBERT.pdf> [consulté le 25 avril 2009].
- Gist, M.E., Schwoerer, C., Rosen, B., 1989. Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology* 74, 884–891.
- Goodykoontz, L., Herrick, C.A., 1990. Evaluation of an in-service education program regarding aggressive behaviour on a psychiatric unit. *Journal of Continuing Education in Nursing* 21, 129–133.
- Gosling, P., 1992. *Qui est responsable de l'échec scolaire?* Presses universitaires de France, Paris.

- Hamblin, A.C., 1974. Evaluation and control of training. McGraw-Hill, Maidenhead.
- Hadji, C., 1992. L'évaluation des actions éducatives. PUF, Paris.
- Hertenstein, E.J., 2001. Goal orientation and practice condition as predictors of training results. *Human Resource Development Quarterly* 12, 403–419.
- Huteau, M., 2003. Comment évaluer les effets des méthodes de formation? In: Lévy-Leboyer, C., Huteau, M., Louche, C., Rolland, J.P. (Eds.), *La psychologie du travail*. Éditions d'organisation, pp. 289–308.
- Inouye, J., Flannelly, L., Flannelly, K., 2001. The effectiveness of self-management training for individuals with HIV/AIDS. *Journal of the Association of Nurses in Aids Care* 12, 73–84.
- Keith, N., Frese, M., 2005. Self-regulation in error management training: emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology* 90, 677–691.
- Kehr, H.M., Bles, P., von Rosenstiel, L., 1999. Self-regulation, self-control, and management training transfer. *International Journal of Educational Research* 31, 487–498.
- Kirkpatrick, D.L., 1959. Techniques for evaluating training programs. *Journal of American Society for Training and Development* 13, 3–9.
- Kirkpatrick, D.L., 2007a. The four level of evaluation. Info Line 0701, 1–16.
- Kirkpatrick, D.L., Kirkpatrick, J.D., 2007b. Implementing the four levels. Berrett-Koehler Publishers, San Francisco.
- 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.
- Kraiger, K., Jung, K.M., 1997. Linking training objectives to evaluation criteria. In: Quinones, M.A., Ehrenstein, A. (Eds.), *Training for a rapidly changing workplace: Applications of psychological research*. American Psychological Association, Washington, DC, pp. 151–175.
- Lehmann, L.S., Padilla, M., Clark, S., Loucks, S., 1983. Training personnel in the prevention and management of violent behavior. *Hospital and Community Psychiatry* 34, 40–43.
- Lin, Y.R., Shiah, I.S., Chang, Y.C., Lai, T.J., Wang, K.Y., Chou, K.R., 2004. Evaluation of an assertiveness training program on nursing and medical students' assertiveness, self-esteem, and interpersonal communication satisfaction. *Nurse Education Today* 24, 656–665.
- Lupton, R.A., Weiss, J.E., Peterson, R.T., 1999. Sales Training Evaluation Model (STEM): A conceptual framework. *Industrial Marketing Management* 28, 73–86.
- Marth, J., 1994. Evaluating Organizational Training Programs. Annual Meeting of the Midwestern Educational Research Association, Chicago.
- McCain, D.V., 2005. Evaluation Basics. ASTD Press, Alexandria.
- 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, 37–53.
- Moliner, P., Tafani, E., 1997. Attitudes and social representations: a theoretical and experimental approach. *European Journal of Social Psychology* 27, 687–702.
- Monteil, J.M., 1990. Éduquer et former: perspectives psycho-sociales. PUG, Grenoble.
- Pansu, P., Bressoux, P., Louche, C., 2003. Theory of the social norm of internality applied to education and organizations. In: Dubois, N. (Ed.), *A sociocognitive approach to social norms*. Routledge, London, pp. 196–230.
- Patrick, J., 1992. Training: Research and practice. Academic Press, London.
- Pershing, J.A., Pershing, J.L., 2001. Ineffective reaction evaluation. *Human Resource Development Quarterly* 12, 73–90.
- Phillips, J.J., 1997. Return on investment in training and performance improvement programs. Gulf Publishing Company, Houston.
- Phillips, J.J., Phillips, P.P., Hodges, T.K., 2004. Make Training Evaluation Work. ASTD Press, Alexandria.
- Quinones, M.A., 1995. Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology* 80, 226–238.
- Regné, H., 2002. A nonexperimental evaluation of training programs for the unemployed in Sweden. *Labour Economics* 9, 187–206.
- Rosholm, M., Nielsen, H.K., Dabalen, A., 2007. Evaluation of training in African enterprises. *Journal of Development Economics* 84, 310–329.
- Ryman, D.H., Biersner, R.J., 1975. Attitudes predictive of diving training success. *Personnel Psychology* 28, 181–188.
- Sanderson, G., 1992. Objectives and evaluation. In: Truelove, S. (Ed.), *Handbook of training and development*. Blackwell Business, Oxford, pp. 114–142.
- Santos, A., Stuart, M., 2003. Employee perceptions and their influence on training effectiveness. *Human Resource Management Journal* 13, 27–45.

- Sharp, C., Hurford, D.P., Allison, J., Sparks, R., Cameron, B.P., 1997. Facilitation of internal locus of control in adolescent alcoholics through a brief biofeedback-assisted autogenic relaxation training procedure. *Journal of Substance Abuse Treatment* 14, 55–60.
- Spector, P., 2006. *Industrial and Organizational Psychology: Research and Practice*, 4th edition. John Wiley & Sons, Hoboken, NJ.
- Spilsbury, M., 1995. *Measuring the Effectiveness of Training*. IES Co-operative Research Programme, Poole.
- Thackway, B., 1997. *Effective evaluation of training and development in higher education*. Kogan Page, London.
- Training Agency, 1989. *Training in Britain*. HMSO, London.
- Tracey, J.B., Tannenbaum, S.I., Kavanagh, M.J., 1995. Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology* 80, 239–252.
- Tracey, J.B., Hinkin, T.R., Tannenbaum, S., Mathieu, J.E., 2001. The influence of individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly* 12, 5–23.
- Warr, P., Bird, M., Rackham, N., 1970. *Evaluation of management training*. Gower, Aldershot.