

L'évaluation du retour sur les attentes à l'aide du cycle de Construction et de gestion qualité des évaluations de l'impact des formations (cycle CGQEIF)

Evaluation of the return on expectations using the Construction and quality management cycle for training impact evaluations (CGQEIF cycle)

Avaliação dos retornos sobre as expectativas com base no ciclo de Construção e gestão da qualidade das avaliações de impacto das formações (ciclo CGQEIF)

Mauro Dos Santos Paulo¹

ID ORCID : 0009-0002-0303-7864

Haute école pédagogique du canton de Vaud, Lausanne, Suisse

Jean-Luc Gilles

ID ORCID : 0000-0002-3549-4683

Haute école pédagogique du canton de Vaud, Lausanne, Suisse

Jean-Michel Rigo

ID ORCID : 0000-0002-0031-526X

Université de Hasselt, Belgique

Yann Barbaras

Haute école pédagogique du canton de Vaud, Lausanne, Suisse

Anissa Bentayeb

Haute école pédagogique du canton de Vaud, Lausanne, Suisse

1 Les auteurs tiennent à remercier les responsables et les équipes de formation des deux terrains de mise à l'essai du cycle CGQEIF: Espace compétences et la FORDIF. Ils remercient également pour leurs contributions à cette R-D toutes les personnes qui ont accepté d'être interrogées par voie de questionnaires, d'interviews ou de focus groups.



MOTS CLÉS: évaluation de l'impact des formations, évaluation du retour sur les attentes, modèle de Kirkpatrick-Phillips, recherche et développement

Les évaluations de l'effet des formations sont peu pratiquées, notamment parce que les évaluateurs manquent d'instruments méthodologiques facilitant leur mise en œuvre. Les auteurs présentent une recherche-développement (R-D) qui a abouti à une méthode d'évaluation des retours sur attentes (ROE) en formation qu'ils ont intitulée cycle de Construction et de gestion qualité des évaluations de l'impact des formations (cycle CGQEIF). Ce cycle a été mis à l'essai dans deux contextes de formation continue : une première fois, dans le secteur de la santé avec une formation d'assistants techniques spécialisés en salle d'opération (ATSSO) et ensuite, dans le cadre d'un module de formation de directeurs d'institutions de formation (FORDIF) en éducation. Les résultats des deux mises à l'essai nous ont permis d'améliorer la méthode et d'en renforcer la pertinence. Les leçons que nous en tirons seront exploitées lors d'une troisième mise à l'essai dans le contexte de l'enseignement supérieur.

KEY WORDS: evaluation of return on expectations, evaluation of training impact, Kirkpatrick-Phillips model, research and development

Evaluations of the impact of training are not widely practiced, partly because evaluators lack the methodological tools to facilitate their implementation. The authors present a research and development (R&D) project that has resulted in a method for evaluating return on expectations (ROE) in training, which they have entitled the cycle de Construction et de gestion qualité des évaluations de l'impact des formations (cycle CGQEIF). This cycle was tested in two continuing training contexts: firstly, in the health sector, with a training course for specialized technical assistants in operating rooms (ATSSO), and secondly, as part of a module for training directors of training institutions (FORDIF) in education. The results of both trials enabled us to improve the method and reinforce its relevance. The lessons we have learned will be applied to a third trial in the context of higher education.

PALAVRAS-CHAVE: avaliação de efeito, avaliação dos retornos sobre as expectativas, modelo de Kirkpatrick-Phillips, investigação e desenvolvimento

As avaliações do efeito das formações são pouco praticadas, principalmente porque os avaliadores carecem de instrumentos metodológicos que facilitem sua implementação. Os autores apresentam uma investigação-desenvolvimento (I&D) que resultou numa metodologia de avaliação dos retornos sobre as expectativas (ROE) em formação, a qual denominaram ciclo de Construção e de Gestão da Qualidade das Avaliações do Efeito das Formações (ciclo CGQAEF). Este ciclo foi testado em dois contextos de formação contínua: inicialmente no setor da saúde, com uma formação de assistentes técnicos especializados em salas de operação (ATESO), e, posteriormente, no âmbito de um módulo de formação de diretores de instituições de formação (FORDIF), na área da educação. Os resultados dos dois testes permitiram melhorar a metodologia e reforçar sua pertinência. As lições extraídas dessas experiências serão aproveitadas numa terceira fase de testes, desta vez no contexto do ensino superior.

Introduction

L'évaluation de l'impact des formations correspond au niveau 4 du modèle de Kirkpatrick (Kirkpatrick, 1959, 1996 ; Kirkpatrick & Kirkpatrick, 2006). Malgré son ancienneté et bien qu'il ait fait l'objet de critiques, principalement sur son caractère hiérarchique et causal (Alliger & Janak, 1989), ce modèle est très utilisé dans le monde de l'éducation tout au long de la vie, comme le soulignent Alsalamah et Callinan (2022, p. 36): « *One of the best-known and most widely used models for evaluating training and development programs is Kirkpatrick's (1959) four-level evaluation model (the four levels being a reaction, learning, behavioral and results)* »².

Aux trois premiers niveaux d'évaluation de Kirkpatrick, l'évaluateur s'intéresse à la satisfaction des personnes formées (niveau 1), à leurs apprentissages (niveau 2) et au transfert de leurs acquis dans l'environnement de travail (niveau 3). Concernant l'évaluation de l'effet (niveau 4),

2. L'un des modèles les plus connus et les plus largement utilisés pour évaluer les programmes de formation et de développement est le modèle d'évaluation en quatre niveaux de Kirkpatrick (1959), ces quatre niveaux étant les réactions, les apprentissages, les comportements et les résultats (traduction libre).

il évalue les retombées de la formation par rapport aux attentes des commanditaires en consultant les données disponibles ou en interrogeant les acteurs des départements concernés.

Notre présentation du cycle CGQEIf, qui vise à faciliter le travail des évaluateurs soucieux de mettre en œuvre des évaluations de l'impact basées sur une méthodologie solide, débute par un cadrage théorique où nous situons le niveau d'évaluation de l'impact dans le modèle de Kirkpatrick, les enjeux en matière de retour sur les attentes des parties prenantes, et où nous rappelons les origines de la méthode, ses étapes ainsi que ses liens avec une approche qualité en évaluation. Nous abordons ensuite les objectifs et le processus de cette recherche-développement (R-D) avant de présenter l'adaptation des étapes du cycle qualité au contexte particulier du niveau 4 d'évaluation de l'impact des formations. Nous enchaînons sur la mise en œuvre de ces étapes du cycle qualité dans deux contextes de formation assez différents. Enfin, nous analysons les résultats de ces deux mises à l'essai, présentons les limites de notre R-D et terminons par quelques conclusions et perspectives en matière de pistes d'améliorations du cycle qualité en vue de prochaines mises à l'essai.

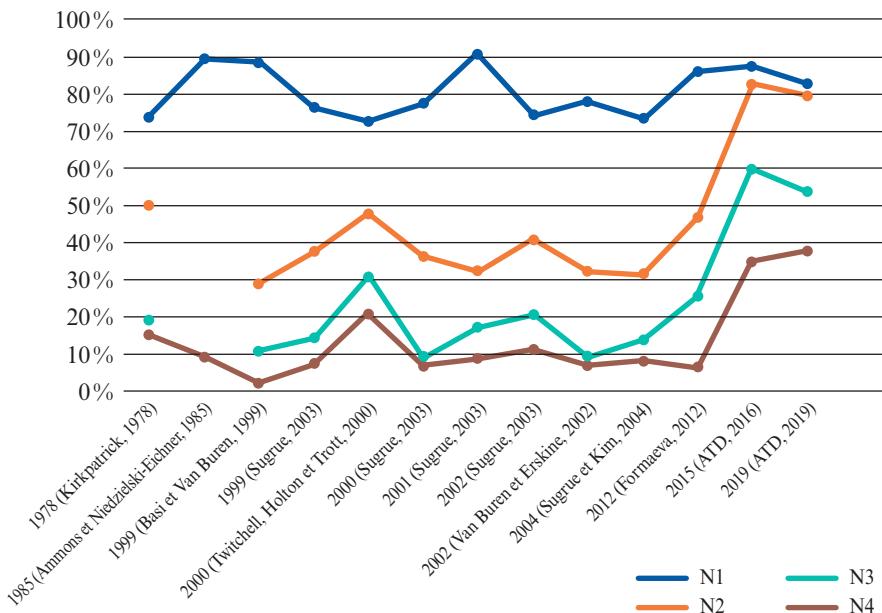
Cadre théorique

Pottiez (2021, p. 15) présente les résultats d'une série d'études qui s'intéressent à l'utilisation des quatre niveaux du modèle de Kirkpatrick depuis une quarantaine d'années. Ces études exploitent les données des rapports annuels de l'*American Society for Training and Development* (Ammons & Niedzielski-Eichner, 1985; ATD, 2016, 2019; Basi & Van Buren 1999; Formaeva, 2012; Kirkpatrick, 1978; Sugrue & Kim, 2004; Twitchell et al., 2000; Van Buren & Erskine, 2002). La figure 1 offre une vue synthétique du tableau présenté par Pottiez : elle montre que plus les niveaux augmentent, moins ils sont évalués. En effet, les pourcentages de la courbe du niveau 4 montrent une fréquence de mise en œuvre qui se situe en deçà de 20% jusque dans les années 2010, avec une hausse de la mise en œuvre de dispositifs d'évaluation du niveau 4 - évaluation de l'effet, depuis le début des années 2010. Cependant, comparativement aux trois premiers niveaux (N1 - évaluation de la satisfaction ; N2 - apprentissages et N3 - transfert), le niveau 4 - évaluation de l'effet reste le moins utilisé.

Figure 1

Évolution des fréquences de mise en œuvre des quatre niveaux de Kirkpatrick aux USA depuis la fin des années 1980

Évolution des fréquences de mise en œuvre des quatre niveaux du modèle de Kirkpatrick dans les entreprises américaines depuis 1978 (d'après le tableau de Pottiez, 2021, p. 15)



Source. Dos Santo Paulo, 2023

Selon Pottiez (2021, p. 12-13), la tendance est similaire en Europe. Ces données vont dans le sens des propos de Dunberry et Péchard (2007, p. 14-15) qui signalent, en s'appuyant sur les données de nombreuses enquêtes aux États-Unis et au Canada, que le niveau 4 est peu évalué. Ces auteurs expliquent que cela serait principalement dû au fait que les responsables de formations se sentent démunis sur un plan méthodologique quand il s'agit d'évaluer l'effet des formations au sein de leurs organisations, en particulier lorsqu'ils souhaitent des dispositifs solides d'évaluation d'impact offrant des garanties de fiabilité des résultats obtenus.

Les statistiques de mise en œuvre de l'évaluation de l'impact de la formation, présentées plus haut, montrent que ce niveau est moins évalué comparativement aux trois autres. Parmi les causes de cette sous-représentation

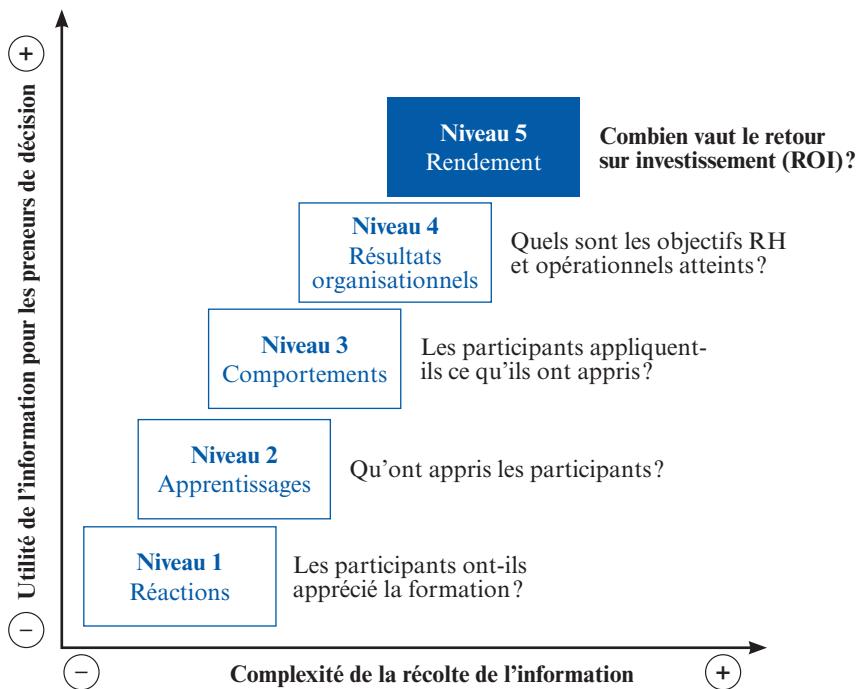
figure le fait que les praticiens manquent de marches à suivre pour réaliser de telles évaluations. Ils ne voient pas bien quelles opérations et quelles démarches doivent être entreprises pour évaluer ce niveau. Autrement dit, le niveau 4 de Kirkpatrick manque d'instrumentation (Dunberry & Péchard, 2007). Dès lors, les évaluateurs ont des appréhensions face à la mise en œuvre d'une évaluation de l'impact du fait de sa complexité et de sa lourdeur (Dunberry & Péchard, 2007 ; Monnot, 2014 ; Pottiez, 2021).

Ces constats nous ont amenés à mettre au point la méthode du cycle de Construction et de gestion qualité des évaluations de l'impact des formations (cycle CGQEIF) que nous présentons dans cet article et qui vise à aider les évaluateurs dans la mise en œuvre du niveau 4 dans les milieux de la pratique. Cette méthode a ensuite été mise à l'essai dans deux contextes différents : une première fois, dans le cadre de l'évaluation de l'impact de la formation d'assistants techniques spécialisés en salle d'opération (ATSSO) et ensuite, dans le contexte de l'évaluation de l'impact d'un module de la formation de directeurs d'institutions de formation (FORDIF). Nous décrirons ces deux contextes de formation plus loin dans la section qui traite de la mise en œuvre du cycle CGQEIF.

Comme nous l'avons expliqué plus haut, le modèle en quatre niveaux d'évaluation (satisfaction-apprentissage-transfert-impact) proposé par Kirkpatrick (Kirkpatrick, 1959, 1996 ; Kirkpatrick & Kirkpatrick, 2006) est bien connu des praticiens du monde de la formation tout au long de la vie (Alsalalah & Callinan, 2022 ; Pottiez, 2019, 2021). La littérature montre un intérêt grandissant pour le niveau 4 d'évaluation de l'impact qui devrait permettre de vérifier si la formation répond aux attentes et si l'investissement est justifié. Certains auteurs, tels que Phillips (1997) et Phillips et Stone (2002) ont développé des indicateurs financiers pour alimenter un cinquième niveau retour sur investissement ou *return on investment* (ROI) qui s'ajoute aux quatre premiers du modèle de Kirkpatrick pour former le modèle Kirkpatrick-Phillips à cinq niveaux.

L'ajout d'un cinquième niveau se justifie par le fait que les entreprises et les organisations investissent souvent des sommes considérables dans la formation du personnel sans en connaître les potentielles retombées (Gilles & Chochard, 2022) et, comme l'indique Gérard (2003), elles ne peuvent plus se contenter de former pour former. Ainsi, l'évaluation devrait permettre de démontrer le retour sur investissement de la formation en mettant en évidence les bénéfices concrets et mesurables qu'elle apporte à l'organisation.

Figure 2
Le modèle de Kirkpatrick-Phillips



Source. Chochard, 2012

Concernant le niveau 4, selon Meignant (2014), l'impact d'une formation peut non seulement se mesurer en référence aux indicateurs directs de performance de l'organisation, par exemple, la réduction des coûts ou l'augmentation de la production, mais aussi à l'aide d'indicateurs indirects comme la diminution des incidents, la réduction des plaintes ou encore l'augmentation de l'adaptabilité des personnes formées.

Dès lors qu'on s'intéresse plus aux retours sur les attentes des parties prenantes de la formation qu'aux retours financiers, le niveau 4 intéressera plus particulièrement les institutions sans but lucratif où «la mesure du ROI se focalise souvent plus sur l'efficacité de l'accomplissement d'une mission spécifique que sur sa valeur monétaire» (Wang et al. 2002, p. 208, cités dans Dunberry et Péchard, 2007). En lien avec le niveau 4, il est alors question de *return on expectations* (ROE) ou retour sur les attentes

(Goldwasser, 2001). Selon Pottiez, «le ROE est une véritable démarche collaborative d'évaluation de l'impact, qui commence en amont de la formation par l'identification des attentes des commanditaires et leur traduction en critères de succès, qui deviendront ensuite en aval les critères d'évaluation» (2019, p. 5).

Ainsi, une formation n'exerce pas un effet uniquement sur l'apprenant, mais également sur toute une série d'acteurs *lato sensu*. En effet, les commanditaires de la formation ne sont pas les seuls à avoir des attentes. Les apprenants, lors d'une formation, cherchent à acquérir certains résultats tels que des compétences, une amélioration de leurs performances au travail, etc. (Pottiez, 2021). Les gestionnaires, quant à eux, ont des attentes managériales telles que le développement des compétences de l'équipe ou la réduction de coûts. Mais d'autres effets intangibles peuvent influencer des groupes encore plus larges. Selon Pottiez (2021), ces effets «peuvent concerner tant le personnel et l'organisation du travail (par ex : renforcement de la cohésion des équipes, plus forte implication ...) de l'entreprise dans son ensemble (par ex : meilleure image)». Par exemple, les équipiers des personnes formées auront donc très probablement des attentes explicites ou implicites concernant les formations suivies par leurs collègues dans la mesure où elles affectent aussi leur propre travail. D'autres parties prenantes plus éloignées peuvent également avoir des attentes explicites ou implicites, comme les (futurs) élèves des enseignants qu'on forme ou les patients des soignants en formation.

Depuis deux décennies, des chercheurs en ingénierie des évaluations s'attellent à produire des outils pour évaluer les différents niveaux d'évaluation des formations dans le cadre d'une approche qualité. Basé sur la roue de Deming (voir figure 3, plus loin), le cycle de construction de gestion qualité des tests standardisés ou cycle CGQTS (Gilles, 2002 ; Gilles & Leclercq, 1995) constitue un des premiers essais de mise en œuvre d'une spirale de qualité dans le domaine de la construction d'évaluations des apprentissages. Le cycle CGQTS comporte huit étapes : 1) Analyse des objectifs de l'enseignement ; 2) Conception du dispositif d'évaluation ; 3) Construction des modalités de questionnement ; 4) Entraînement des évalués ; 5) Passation de l'évaluation ; 6) Correction et discussion ; 7) *Feed-back* aux évalués et 8) Macrorégulation du cycle. Cette spirale qualité a d'abord été proposée pour le niveau 2 de Kirkpatrick, l'évaluation des apprentissages. Ensuite, elle a été adaptée à la construction et la gestion qualité des évaluations du niveau 3 de Kirkpatrick, l'évaluation

du transfert des acquis de la formation dans l'environnement de travail (Gilles & Chochard, 2022). Par la suite, la méthode en cycle a encore été adaptée pour le niveau 1 de Kirkpatrick (Kaeser Boutillier et al. 2022). Finalement, Gilles & Rigo (2022) ont transposé le cycle qualité au niveau 4 du modèle de Kirkpatrick, l'évaluation de l'impact des formations. Ces derniers l'ont intitulé cycle de Construction et de gestion qualité d'évaluations de l'impact des formations (cycle CGQEIF).

Dans une première version, le cycle d'évaluation de l'impact d'une formation comprenait huit étapes, auxquelles a été ajoutée plus récemment une étape 0 Enjeux (Gilles & Rigo, 2022). Voici en résumé la description des neuf étapes du cycle CGQEIF (La figure 4, plus loin, présente un schéma du cycle CGQEIF).

Étape 0. Enjeux

Cette étape s'est avérée nécessaire étant donné l'importance de clarifier les finalités de l'évaluation de l'impact. S'agit-il de rendre des comptes aux autorités qui ont décidé de mettre en place l'action de formation ? Doit-on s'intéresser à l'impact de la formation sur les destinataires finaux, par exemple les patients, dans le cadre d'une formation dans le monde de la santé ? La formation dont on évalue l'impact faisait-elle partie d'un plan affectant toute l'organisation dont sont issues les personnes formées ? Par exemple, lorsqu'une formation concerne la gestion de projets numériques dans les établissements scolaires, peut-on en évaluer l'impact sans tenir compte des enjeux associés à une stratégie globale d'introduction du numérique dans l'espace régional où se situent ces établissements ? Il s'agit donc à cette étape préliminaire 0 de s'interroger sur les finalités de la formation, d'en clarifier les enjeux, et ce, afin de pouvoir délimiter le périmètre du dispositif d'évaluation d'impact et d'identifier les parties prenantes en vue de l'étape suivante d'analyse.

Étape 1. Analyse

Cette étape est la plus chronophage, car elle a comme objectif de clarifier les attentes liées à l'organisme de formation et de les prioriser afin de dégager celles qui pourront être évaluées. En première approche, il s'agit d'identifier les parties prenantes directes ou indirectes concernées par la formation ainsi que les facteurs influençant l'impact que peut avoir la formation. À cette étape, des groupes de travail seront mobilisés. Ils seront composés de personnes représentatives des parties prenantes identifiées et

animés par l'équipe des évaluateurs. L'idée est de dégager les principales attentes des parties prenantes au dispositif de formation. Ce travail, qui prend souvent la forme de *brainstormings*, débouche sur un produit clé à cette étape : une matrice acteurs/attentes, une sorte de tableau de spécifications où les dimensions d'impact ont été clarifiées en fonction des attentes regroupées en catégories.

Étape 2. Design

Cette étape consiste à élaborer un plan du dispositif de récolte de données, incluant son agencement dans le temps. Autrement dit, il s'agit de choisir la méthodologie de récolte de données en fonction des caractéristiques du milieu et des possibilités offertes par le contexte de formation et les profils des parties prenantes.

Étape 3. Opérationnalisation

Ici, il s'agit d'élaborer les instruments qui permettront la récolte des données. On cherche à construire et à opérationnaliser les outils qui seront utilisés à l'étape 5 - Récolte des données en fonction du plan établi à l'étape précédente. Les sources d'information peuvent être multiples lorsqu'on évalue l'impact d'une formation. Les aspects qualitatifs et quantitatifs sont complémentaires. Par exemple, il est possible de questionner différents profils de parties prenantes à l'aide de questionnaires qui peuvent inclure des items sous forme d'échelles de Likert et des questions ouvertes. Des interviews ou des *focus groups* peuvent également compléter ce type de récolte. Il est également envisageable de créer des *query* (instructions écrites dans un langage informatique de requête) en vue de tirer des informations utiles de bases de données préexistantes.

Étape 4. Information

Cette étape consiste à informer les parties prenantes à propos du dispositif d'évaluation d'impact qui se met en place. Elle est particulièrement importante pour provoquer l'adhésion des personnes concernées par la récolte des données et leur explique les enjeux et l'utilité du dispositif d'évaluation. Il s'agit également de préciser les mesures d'anonymisation des données et les garanties de confidentialité. Des informations techniques et logistiques sont également communiquées, par exemple à propos du temps de passation, des possibles relances, des rétroactions qui seront fournies, etc. L'idée générale ici est d'être le plus clair et le plus transparent

possible avec les parties prenantes à propos du dispositif d'évaluation d'impact et de ses enjeux afin que les étapes suivantes se passent dans les meilleures conditions.

Étape 5. Collecte des données

Cette étape consiste à récolter les données selon le plan établi à l'étape 2 - Design et avec les instruments élaborés à l'étape 3 - Opérationnalisation. Elle inclut tous les aspects logistiques à prévoir, qui vont de la réservation de locaux, avec toute l'intendance qui l'accompagne, jusqu'à la mise en place de moment de prises d'information à distance, par exemple, en organisant des interviews ou des *focus groups* en visioconférence.

Étape 6. Traitements

Le but est de traiter et d'analyser les données récoltées en fonction des dimensions d'impact clarifiées à l'étape 1 - Analyse. Des synthèses des résultats aux différentes dimensions d'impact sont produites en vue d'alimenter les supports des rétroactions de l'étape suivante.

Étape 7. Rétroactions

Les résultats de l'étude sont présentés aux commanditaires ainsi qu'aux différentes acteurs concernés par l'évaluation de l'impact de la formation identifiés à l'étape 1 - Analyse. L'objectif est donc de communiquer le plus clairement possible les informations issues de cette évaluation de l'impact de la formation pour permettre aux responsables, aux commanditaires et à toutes les parties prenantes d'effectuer un bilan des points forts et des points faibles et de dégager des pistes d'amélioration, s'il y a lieu.

Étape 8. Macrorégulation

Cette étape de contrôle et de régulation de l'évaluation de l'impact a pour objectif de cerner les aspects à améliorer dans le dispositif d'évaluation mis en place. Il s'agit de passer en revue les actions entreprises à chaque étape et les produits qui en ont découlé en vue de tirer les leçons pour une prochaine mise en œuvre de l'évaluation d'impact. Cette analyse fournit également des informations qui peuvent être injectées dans le dispositif de formation. Par exemple, si cette macrorégulation montre qu'une catégorie d'acteurs concernés par la formation a été négligée à l'étape 1 - Analyse et que la formation ne prend pas leurs attentes en compte, cela devrait également faire l'objet d'une régulation sur le plan du dispositif de formation.

Gilles (2002) signale que le concept de qualité renvoie à un exercice difficile, voire impossible. En effet, comme l'indique la définition d'ISO 8402 «la qualité n'existe pas une fois pour toutes (...) La qualité s'inscrit dans un processus progressif infini qui fait que sa limite (l'excellence) n'est jamais atteinte». Dans une démarche perpétuelle d'amélioration de la qualité du processus d'évaluation, le cycle d'évaluation de l'impact fait appel aux directives et aux normes de qualité formalisées dans les *Standards for educational and psychological testing* (AERA-APA-NCME, 1999). Ces normes ont été réparties en critères de qualité (Gilles, 2002 ; Gilles & Charlier, 2020) qui sont, selon nous, indissociables du processus en huit étapes du cycle CGQEIF décrit plus haut. Ces critères de qualité constituent en quelque sorte des points de repère qui nous permettent de déterminer la pertinence des actions entreprises aux différentes étapes. Chaque étape du processus doit répondre à un ou à plusieurs de ces critères de qualité, que nous résumons ci-dessous.

La validité

Ce critère concerne la couverture des attentes à évaluer. Il s'agit de vérifier si le dispositif d'évaluation de l'impact évalue réellement toutes les attentes qui ont été identifiées comme prioritaires.

La fidélité

Il s'agit ici de s'assurer que le dispositif d'évaluation de l'impact de la formation est fiable et sans biais. Les informations qu'il renvoie doivent être objectives.

La sensibilité

Le dispositif d'évaluation de l'impact de la formation doit fournir des informations suffisamment nuancées pour permettre de déceler des variations fines sur les dimensions d'impact analysées.

La diagnosticité

Ce critère de qualité porte sur la capacité du dispositif d'évaluation d'impact à faire un diagnostic précis des points forts et des points à améliorer en lien avec les attentes évaluées dans le dispositif.

L'équité

Ce critère de qualité concerne l'impartialité et la justesse de l'outil d'évaluation. Il faut, entre autres, se prémunir contre les conflits d'intérêts et veiller à ce que l'outil tienne compte des attentes de toutes les parties prenantes.

La praticabilité

Ce critère de qualité est lié à la facilité de mettre en œuvre le dispositif d'évaluation de l'impact d'une formation. Il doit être réalisable dans les délais impartis et efficient sur le plan de la mobilisation des ressources humaines et matérielles disponibles.

La communicabilité

Ce critère de qualité fait référence à la transparence, à la clarté et à l'intelligibilité des informations communiquées aux parties prenantes du dispositif d'évaluation de l'impact. Les informations non confidentielles doivent être communiquées et comprises par les parties prenantes.

L'authenticité

Il faut s'assurer que les items et les instruments qui seront utilisés pour évaluer l'impact de la formation aient du sens pour les parties prenantes et soient en lien avec leur vécu.

L'objectif de notre R-D en ingénierie des évaluations en formation

Les constats issus de la littérature et de nos observations sur le terrain nous ont poussés à entreprendre cette recherche-développement (R-D) d'une méthode en cycle de Construction et de gestion qualité des évaluations de l'impact des formations (cycle CGQEIF). Nous nous sommes inspirés d'autres cycles qualité utilisés en ingénierie des évaluations en formation et ayant fait leurs preuves aux trois premiers niveaux d'évaluation du modèle de Kirkpatrick (Boutillier et al., 2020; Chochard et al., 2017; Chochard & Gilles, 2016; Gilles, 2002; Gilles et al., 2017; Gilles & Leclercq, 1995). Considérablement à ces cycles, des critères de qualité tels que la validité, la fidélité, la sensibilité, la communicabilité permettent d'envisager des évaluations d'impact d'autant plus fiables qu'elles auront été l'objet de contrôles à différentes étapes du cycle CGQEIF.

Dans ce contexte, le but de notre recherche est de proposer un cycle qualité en vue d'outiller les évaluateurs souhaitant mesurer de la façon la plus fiable possible l'impact de leurs formations. L'idée est d'inscrire cette démarche dans une perspective d'amélioration continue, non seulement des formations évaluées, mais aussi du dispositif d'évaluation de l'impact. Le cycle que nous proposons doit être adaptable à divers contextes de formation et facile à mettre en œuvre. Sa praticabilité dépend de la simplicité des étapes proposées, que les acteurs des milieux de la pratique doivent pouvoir facilement appréhender. Enfin, la méthode doit permettre des contrôles qualité liés aux critères de qualité.

Méthodologie

Nous privilégions la méthodologie de la recherche-développement (R-D) en éducation (Gascon & Germain, 2017; Loiselle, 2001 ; Loiselle & Harvey, 2007 ; Van der Maren, 2014). Ce modèle cyclique de R-D se rapproche au plus près de notre outil d'évaluation de l'impact d'une formation : une itérativité entre la théorie et le terrain, les phases de développement et les phases de test, qui nous a paru essentielle dans notre recherche.

Outre le développement du produit, la R-D peut aussi poursuivre une finalité d'analyse de l'objet développé, de son contexte d'application, des perceptions des acteurs dans leur expérience d'utilisation et de son utilité. Dans notre cas, le développement d'un cycle CGQEIF implique l'utilisation des critères de qualité en ingénierie des évaluations. Ces critères aideront à juger de la pertinence des actions entreprises et de la fiabilité des résultats aux différentes étapes du processus cyclique.

Harvey et Loiselle (2009) proposent six phases pour la mise en œuvre d'une R-D en éducation-formation que nous transposons à l'étape 5 - Mise en œuvre du processus R-D.

La mise en œuvre du processus R-D

Dans la partie qui suit, nous allons détailler la mise en œuvre du processus R-D proposé par Harvey et Loiselle que nous avons utilisé dans le cadre du développement du cycle CGQEIF.

Phase 1 - Origine - Clarification de l'idée initiale et des objectifs de cette R-D visant à développer un cycle CGQEIF

En complément des informations apportées dans l'introduction, signalons que l'idée de créer un cycle CGQEIF est née au sein du Groupe international de recherches et expertises en ingénierie des évaluations en formations (GIRIEF)³ qui soutient la R-D en ingénierie des évaluations en formation. Les objectifs de cette R-D sont triples :

- 1) développer une méthode de construction et de gestion de la qualité des évaluations de l'impact des formations, le cycle CGQEIF, qui soit fiable et facilement appréhendable par les professionnels de la formation continue ;

3. <http://www.girief.net>

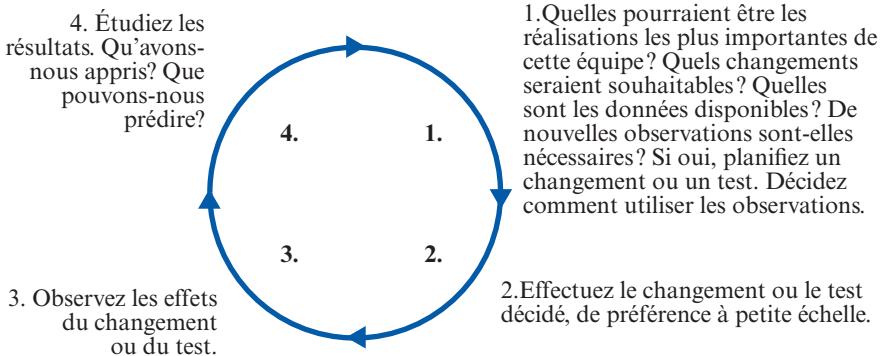
- 2) tirer les leçons de deux mises à l'essai du cycle CGQEIF dans deux contextes différents en formation continue en Suisse romande : la première, dans le secteur de la santé, dans le cadre d'une formation d'assistants techniques spécialisés en salle d'opération (ATSSO) et la seconde, dans le domaine de l'enseignement, dans le contexte de la formation de directeurs d'institutions de formation (FORDIF);
- 3) finalement, et de façon indirecte, participer à l'amélioration continue des dispositifs de formation et contribuer à leur efficience sociétale.

Phase 2 - Référentiel - Définition du cadre théorique de référence et des développements antérieurs dans le domaine des cycles qualité en évaluation.

Notre démarche en cycle qualité est inspirée du cycle *Plan-Do-Check-Act*, ou roue de Deming (Deming, 1986). Selon nous, il est essentiel de définir un processus étape par étape pour garantir des évaluations robustes, quel que soit le niveau de Kirkpatrick envisagé. Dans ce domaine, nous avons été influencés par les travaux d'Edward Deming, considéré par beaucoup comme le fondateur du mouvement de la gestion de la qualité (Juran, 2003 ; Peters & Waterman, 1982 ; Suarez, 1992) Une première idée clé pour une démarche qualité est de pratiquer l'amélioration continue et de penser les processus comme des systèmes (Deming, 1986). C'est ce que nous avons transposé dans le domaine de la construction des évaluations de l'impact en gardant à l'esprit sa célèbre citation : « Si vous ne pouvez pas décrire ce que vous faites comme un processus, vous ne savez pas ce que vous faites ». Une autre idée clé que nous avons reprise et adaptée de Deming est celle du cycle *Plan, Do, Check, Act* (PDCA), qu'il a toujours appelé le cycle de Shewhart et qui est basé, du moins partiellement, sur la méthode scientifique. Gilles & Chochard (2022) signalent à ce sujet :

L'origine du cycle PDCA est bien résumée par Best & Neuhauser (2006) : l'évaluation constante de la politique et des procédures de gestion conduit à une amélioration continue. Ce cycle a également été appelé cycle de Deming, cycle Plan-Do-Check-Act (PDCA) ou cycle Plan-Do-Study-Act (PDSA). Bien que Deming ait popularisé le cycle auprès du grand public - un cycle qu'il a appelé le cycle de Shewhart - la plupart des gens l'ont néanmoins désigné le cycle de Deming. Le cycle de Shewhart comporte les quatre étapes suivantes : 1) Planifier : identifier ce qui peut être amélioré et le changement nécessaire ; 2) Faire : mettre en œuvre le changement de conception ; 3) Étudier : mesurer et analyser le processus ou le résultat ; 4) Agir : si les résultats ne sont pas ceux escomptés. (Best & Neuhauser, 2006, p. 142-143). (Gilles & Chochard (2022), notre traduction, p. 3538-3539).

Figure 3
Le cycle de Shewhart représenté par Deming, traduit en français (1986, p. 88)



Étape 5. Répétez l'étape 1, avec les connaissances accumulées.

Étape 6. Répétez l'étape 2, et ainsi de suite.

Pour Deming (1986, p. 88), le cycle PDCA (Shewhart, 1931, 1958) est considéré comme une procédure utile à suivre pour améliorer la qualité. La figure 3 présente les quatre étapes et les questions liées à ce cycle d'amélioration de la qualité, indépendamment d'un contexte de production particulier.

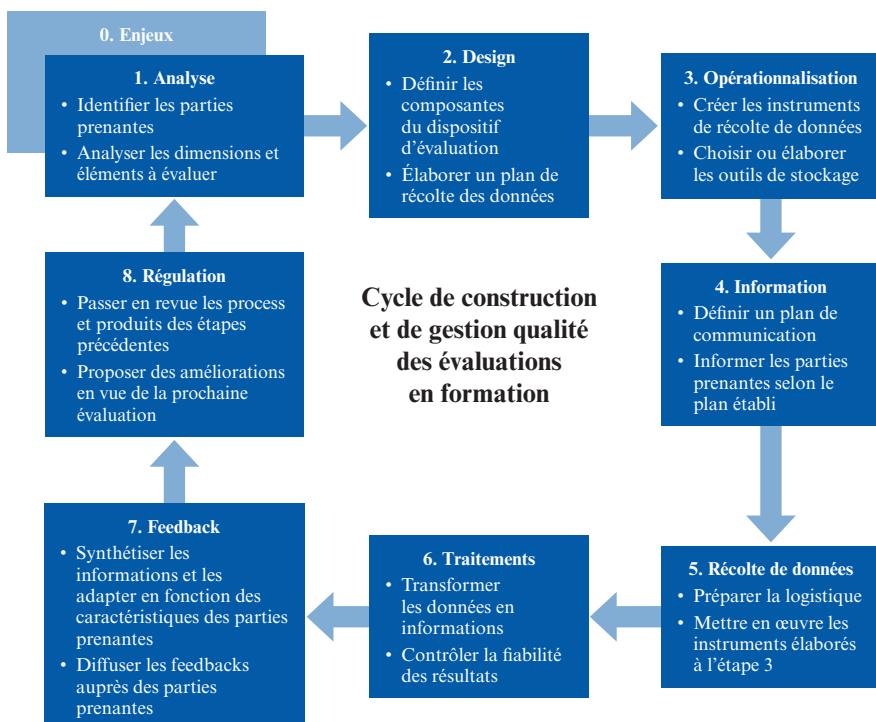
En ingénierie des évaluations en formation et dans le contexte des cycles qualité aux quatre niveaux du modèle de Kirkpatrick, les quatre étapes initiales de Deming ont été portées à huit, puis à neuf avec l'introduction d'une étape 0 - Enjeux pour mieux répondre aux besoins et aux exigences contemporaines en matière de gestion qualité dans le domaine de l'évaluation en formation (Boutillier et al., 2020; Gilles, 2002; Gilles & Leclercq, 1995; Rigo & Gilles, 2022). L'étape préliminaire visant à identifier les enjeux permet de renforcer les étapes suivantes 1 - Analyse et 2 - Design, qui se réfèrent à l'étape planification du cycle de Deming.

Comme signalé plus haut, notre R-D, en lien avec le niveau 4 - Évaluation de l'impact du modèle de Kirkpatrick (1959, 2006), repose sur une méthodologie en cycle qualité déjà éprouvée pour les trois autres niveaux de Kirkpatrick. Progressivement, au fil de ces recherches

cumulatives, un cycle qualité générique émerge pour la construction et la gestion qualité des évaluations en formation. Il peut aujourd'hui se schématiser tel que présenté sur la figure 4.

Cette démarche qui vise à améliorer la qualité des évaluations en formation est inséparable d'une réflexion sur les critères de qualité en ingénierie des évaluations. Nos réflexions dans ce domaine s'inspirent des recommandations qui figurent dans les *Standards for educational and psychological testing* (AERA, APA, NCME, 2014). Par ailleurs, elles s'en éloignent, dans une certaine mesure, dès lors que nous souhaitons conserver un certain recul critique dicté par des cadres contextuels et axiologiques qui peuvent varier en fonction des contextes et des niveaux d'évaluation (Gilles & Charlier, 2020). Sachant que les enjeux d'une évaluation

Figure 4
Cycle de construction et de gestion qualité des évaluations en formation



de formation ne sont pas les mêmes pour une évaluation de la satisfaction et pour une évaluation de l'impact, les critères qualité ne vont pas nous guider de la même manière. Si les mêmes normes sont appliquées uniformément, cela pourrait conduire à des évaluations qui ne sont pas pertinentes ou adaptées aux objectifs spécifiques de chaque niveau. Pour rappel, ces critères de qualité en ingénierie des évaluations en formation ont été rappelés dans la partie intitulée Cadre théorique.

Phase 3 - Méthodologie - Construction des outils et méthodes du cycle CGQEIF

La méthode d'évaluation de l'impact des formations que nous proposons comporte les neuf étapes précisées plus haut : 0) clarification des enjeux ; 1) analyse des impacts recherchés ; 2) design du dispositif de récolte de données ; 3) élaboration des instruments ; 4) information des parties prenantes ; 5) collecte des données ; 6) traitement des résultats ; 7) rétroaction aux parties prenantes et 8) macrorégulation. Pour chacune de ces étapes, nous avons établi un premier référentiel en tenant compte des éléments suivants : a) le but poursuivi par étape et les liens avec les critères de qualité ; b) les opérations à effectuer et les produits finaux de l'étape ; c) les ressources à mobiliser et d) les contrôles qualité. Cette démarche a été entreprise à chaque étape du cycle.

Phase 4 - Opérationnalisation - Mises à l'essai du cycle CGQEIF dans deux contextes différents

Cette phase fut extrêmement chronophage et s'est étalée sur trois ans. Elle a bénéficié des apports d'étudiants du Master en sciences et pratiques de l'éducation (MASPE) qui y ont consacré un stage ou un mémoire (Dos Santos Paulo, 2023).

a) Première mise à l'essai du cycle CGQEIF dans le contexte de la formation d'assistants techniques spécialisés en salle d'opération (ATSSO)

Espace Compétences est un centre de formation à but non lucratif dans le domaine de la santé qui propose diverses formations dans le secteur hospitalier (formation de cadres, préparation au brevet fédéral, logistique, stérilisation, ...). Ce centre « offre un éventail diversifié de formations immédiatement transférables et applicables dans la pratique » et « travaille avec des méthodologies pédagogiques innovantes pour permettre une meilleure intégration des enseignements et favoriser le changement dans la pratique des participants » (Espace Compétences, 2025).

Il y a une vingtaine d'années, les assistants techniques spécialisés en salle d'opération (ATSSO) étaient appelés aides de salle, un métier qui, à l'époque, ne nécessitait aucune qualification spécifique. Pour donner suite au combat de divers acteurs du monde de la santé afin que ce métier soit reconnu et pris en main avec des formations reconnues, l'appellation ATSSO a été créée. Aujourd'hui, il faut au minimum six mois d'expérience en tant qu'aide de salle avant de pouvoir s'inscrire en formation ATSSO.

L'ATSSO est le garant du bon fonctionnement des dispositifs et des appareils médicaux de haute technologie utilisés au bloc opératoire. Il fait office de personne charnière faisant le lien entre le bloc opératoire stérile et l'extérieur. C'est aussi l'ATSSO qui accueille le patient à son arrivée au bloc et qui, une fois l'intervention terminée, le réinstalle dans son lit. Cette formation n'avait pas subi de révisions depuis sa création. Ainsi, la directrice désirait évaluer la formation et y apporter des améliorations en fonction, notamment, des résultats de l'évaluation de l'impact de cette formation.

La mise en place de l'évaluation de l'impact de la formation ATSSO a duré un an. Cela englobait la recherche de terrain, les démarches administratives avec les parties prenantes de la formation, l'évaluation de la formation en elle-même ainsi que la restitution des résultats de l'évaluation de l'impact de cette formation.

b) Seconde mise à l'essai du cycle CGQEIF dans le contexte de la formation de directeurs d'institutions de formation (FORDIF)

Pour cette deuxième mise à l'essai du cycle, nous évaluons un module de formation proposé par la FORDIF intitulé *Piloter un établissement de formation à l'ère du numérique*. Il fait partie intégrante d'une formation certifiante décernant un DAS (*Diploma of Advance Studies*) en gestion et direction d'institutions de formation.

Les institutions de formation ont comme objectif de former aux mieux les citoyens de demain. En Suisse, le Plan d'Études Romand (PER) est un curriculum éducatif utilisé dans les cantons romands. Il s'agit d'un plan-cadre qui définit les objectifs, les contenus d'apprentissage et les compétences attendues pour l'enseignement dans les écoles primaires (degrés 1 à 8) et secondaires (degrés 9 à 11). Ainsi, l'éducation au numérique s'inscrit «dans des perspectives socioculturelles, technologiques et d'application, quels que soient les supports. Elle vise un équilibre approprié entre les processus et les conséquences de la transition numérique en général, comme au sein de l'École» (PER, 2023).

L'objectif général du module dont nous avons évalué l'impact à l'aide du cycle CGQEIF est de former les dirigeants d'institutions à élaborer une stratégie numérique en collaboration avec leur équipe, en prenant en considération et en coordonnant les aspects pédagogiques, organisationnels, éducatifs, juridiques et technologiques.

Phase 5 - Résultats

a) Première mise à l'essai du cycle CGQEIF dans le cadre de l'évaluation de l'impact de la formation d'assistant technique spécialisé en salle d'opération (ATSSO)

0. Enjeux

- Les enjeux de l'évaluation de l'impact de la formation ATSSO ont été discutés avec la directrice du centre de formation en santé Espace Compétences, situé en Suisse romande. Pourquoi cherche-t-on à évaluer la formation ? Dans ce cas, la formation n'avait jamais été régulée en 20 ans, soit depuis sa création.
- Une liste des personnes à mobiliser dans le processus R-D a été établie, de même qu'un premier relevé des parties prenantes ayant des attentes fortes concernant la formation ATSSO. Deux groupes distincts ayant des attentes différentes ont été identifiés lors de ces entrevues. Le premier, que nous avons appelé groupe gravitationnel, est constitué des personnes qui travaillent en interaction directe avec l'assistant technique spécialisé en salle opératoire (par exemple, le chirurgien, l'instrumentiste et l'anesthésiste). Le deuxième, que nous avons appelé groupe gestionnaire, était constitué des responsables directs dans ces hôpitaux (responsable de service, responsable d'unité) afin d'avoir une vision macro sur les attentes.

1. Analyse

- L'équipe des chercheurs-évaluateurs a préparé une ébauche d'analyse des attentes de la formation qu'elle a soumise à la relecture critique de responsables du centre de formation en santé et qui a été discutée au cours de plusieurs allers et retours.
- Les attentes de la formation ont été identifiées en fonction des différentes parties prenantes.
- Les domaines d'attentes des parties prenantes ont été priorisés dans une matrice acteurs/attentes qui constitue une forme de table de spécifications adaptée aux évaluations d'impact de formation.

2. Design

- Dans le contexte de cette mise à l'essai, nous avons choisi de limiter les attentes à tester chez les acteurs aux priorités définies de rang 1.
- Nous avons identifié trois types d'acteurs importants à interroger : i) le groupe gravitationnel, soit les professionnels travaillant dans ou autour de la salle d'opération ; ii) le groupe décisionnel, soit les responsables hospitaliers (niveau bloc opératoire, département, hôpital) et iii) le groupe patient.

3. Opérationnalisation

- À l'aide de la table de spécifications et selon les priorités, des questionnaires ont été élaborés et adaptés aux différents groupes d'acteurs identifiés. Le groupe gravitationnel est composé de personnes qui collaborent directement avec l'ATSSO et le groupe des gestionnaires des infrastructures dans lesquelles l'ATSSO travaille. Les attentes et le degré de sensibilité des données récoltées n'étant pas les mêmes, nous avons dû créer deux questionnaires différents.
- Pour des raisons de praticabilité, le nombre d'items par questionnaire est limité à cinq ou six questions fermées, avec possibilité d'ajouter un commentaire libre. La population que nous avons sollicitée est surchargée au quotidien. Donc, afin de favoriser un maximum de participation, nous avons dû nous adapter à leur emploi du temps et faire en sorte qu'ils puissent répondre au questionnaire à leur meilleure convenance.
- Les questionnaires ont été coconstruits avec les responsables du centre formation Espace Compétences en étant particulièrement attentif au critère de qualité authenticité. Il était en effet crucial que les questions fassent sens pour les participants et soient ancrées dans leur réalité quotidienne.
- Des *focus groups* ont été mis en place pour approfondir l'un ou l'autre des thèmes, en fonction des résultats aux questionnaires. Nous avons ainsi pu compléter les réponses aux questionnaires par des informations de nature qualitative.

4. Information des parties prenantes

- Une première information des parties prenantes a eu lieu par courriel.

- Les parties prenantes ont en outre été informées par nos relais qui ont distribué les questionnaires.

5. Collecte des données

- Les données des questionnaires ont été collectées en juin 2022.

6. Traitement des résultats

- Différents types de graphiques ont été réalisés pour faire apparaître les points saillants des résultats : gains bruts et relatifs (pour les questions qui permettent une comparaison de la situation avant et après la formation ATSSO), distribution de la satisfaction, etc.

7. Rétroaction aux parties prenantes

- Nous avons organisé une séance de rétroaction où les premiers résultats ont été discutés avec les commanditaires de la formation ATSSO et une partie de l'équipe de formation.
- Un rapport consolidé des résultats leur a été fourni.

8. Macrorégulation

- L'analyse rétrospective a eu lieu lors de deux séances avec les partenaires d'Espace compétences et les parties prenantes de la formation ATSSO. Elle a abouti à un relevé des points forts ainsi que des aspects à améliorer sur le plan de la méthodologie en cycle et guidée par les critères de qualité en ingénierie des évaluations en formation. Ces points ont servi à une macrorégulation pour la deuxième mise à l'essai du cycle CGQEIF (voir point suivant).

b) Seconde mise à l'essai du cycle CGQEIF dans le cadre de l'évaluation de l'impact de la formation de directeurs d'institutions de formation (FORDIF)

À ce jour, les étapes 0 à 8 du cycle ont été réalisées. L'étape 8 de macro-régulation a bénéficié des analyses menées par Dos Santos Paulo (2023) dont les résultats ont contribué à alimenter la réflexion sur les autres étapes à l'issue de la seconde mise à l'essai du cycle.

0. Enjeux

- Nous avons discuté des enjeux de l'évaluation de l'impact de la formation FORDIF avec les responsables du module de formation qui portait sur la gestion et sur la conduite de projets numériques par les responsables d'établissements scolaires.
- Nous avons également abordé et clarifié les profils des personnes à

mobiliser dans le processus R-D et des parties prenantes concernées par la formation.

- Cette seconde mise à l'essai a montré que nous pourrions établir trois types de parties prenantes concernées à des niveaux différents : les bénéficiaires finaux au niveau micro ; les responsables de formation au niveau meso et les commanditaires à un niveau macro.

1. Analyse

- L'analyse des attentes des parties prenantes concernées par la formation a bénéficié de l'expérience acquise lors de la première mise à l'essai.
- La formalisation d'une matrice acteurs/attentes a pu être établie avec moins d'allers-retours et a été validée par les responsables du module de formation.
- À partir de la validation de la matrice acteurs/attentes, les étapes suivantes du cycle ont pu être menées avec une relative autonomie de l'équipe d'évaluation, ce qui a accéléré le processus tout en impliquant une partie des bénéficiaires qui ont joué le rôle de personnes ressources. Dos Santos Paulo (2023, p. 65) signale que l'intégration des parties prenantes dès l'étape d'analyse contribue, lors de cette mise à l'essai, au respect de quatre critères de qualité sur huit : 1) la validité via les rendez-vous avec les parties prenantes pour établir des matrices acteurs/attentes ; 2) la fidélité à l'aide d'un dispositif qui laisse peu de place aux biais attribuables à la subjectivité lors de la récolte et des traitements des données ; 3) l'équité, car nous prenons en compte toutes les attentes des différentes parties, et 4) l'authenticité par l'utilisation d'un dispositif qui fait sens pour les différentes parties prenantes.

2. Design

- Deux types de questionnaires en ligne composés d'échelles de Likert (1 à 7) portant sur le sentiment d'accomplissement des attentes ont été proposés : l'un pour les participants à la formation et l'autre ciblant leurs collaborateurs. Les participants ont évalué leur propre sentiment d'accomplissement des attentes avant et après la formation. Pour leur part, les collègues directs ont estimé le sentiment d'accomplissement des attentes des participants. Pour rappel, les participants à la formation ne sont pas les seuls que la

formation influence, l'équipe qui gravite autour d'eux est également influencée.

- Nous avons également envisagé des interviews individuelles et des *focus groups* dans le plan du dispositif d'évaluation d'impact avec les participants à la formation afin d'approfondir l'analyse des données quantitatives récoltées grâce aux questionnaires.

3. Opérationnalisation

- Structuration d'une banque d'items en fonction de la matrice acteurs/attentes.
- Création d'un ordinogramme de flux qui permet de clarifier les parcours possibles dans les questionnaires, en fonction des réponses fournies.
- Validation des instruments d'évaluation par les bénéficiaires qui ont joué le rôle de personnes-ressources (ce qui améliore la validité, la fidélité et l'authenticité des instruments, en particulier). Les échanges ont eu lieu par courriel et par visioconférence. Nous avons passé en revue chaque item des deux questionnaires. Les interviews ont duré approximativement une heure. L'idée était de s'assurer de la validité et de l'authenticité des items en vue d'une récolte de données la plus fiable et objective possible. Dès qu'il y a eu un doute ou une incompréhension, nous avons discuté du problème et avons systématiquement pris en compte et vérifié les suggestions avec les personnes-ressources.

4. Information des parties prenantes

- Nous avons communiqué avec les parties prenantes par courriel, en utilisant la base de données partagée par les responsables de la formation. Le courriel précisait les objectifs de l'évaluation et le temps de passation et sollicitait l'autorisation de poursuivre la démarche de questionnement en ligne. Nous avons choisi ce moyen de communication pour des questions de praticabilité, car notre population réside dans les différents cantons romands
- Le courriel d'information contenait également une demande de propagation du questionnaire auprès des collègues.
- Après deux relances, le taux de participation a été de 61% : parmi les 32 participants, 20 nous ont répondu.

5. Collecte des données

- Les données des questionnaires ont été collectées en mai-juin 2023 à l'aide de l'outil *Google Form*.
- La stratégie de propagation du questionnaire auprès des collègues n'a pas fonctionné. D'après certains commentaires dans les questionnaires, la période de récolte de données ne coïncidait pas bien avec la période de l'année scolaire. En effet, la période de mai-juin est un moment crucial dans la clôture du programme éducatif.
- La stratégie des *focus groups* n'a pas fonctionné : aucun participant ne s'est inscrit. Cette fois-ci, la période proposée coïncidait avec les vacances scolaires.

6. Traitement des résultats

- L'outil Google Sheets a servi à traiter les données en mode collaboratif à distance et à produire des graphiques de synthèse diffusables en ligne.
- Pour certains items, des comparaisons pré-post ont été possibles et ont permis le calcul de gains bruts et de gains relatifs en lien avec le niveau d'atteinte des attentes des parties prenantes avant et après la formation.
- La possibilité de laisser des commentaires a permis des traitements qualitatifs en complément des analyses quantitatives.

7. Rétroaction aux parties prenantes

- Le rétroaction s'est déroulée via une visioconférence à distance. Le groupe était constitué des évaluateurs et des responsables de la formation.
- Cette rétroaction a pris la forme d'une réunion de débreffage avec les responsables de la formation. En première partie, nous sommes repassés sur chaque étape du processus, sur les critères et sur des produits liés aux étapes. Puis, nous avons présenté les résultats et en avons discuté. En dernier lieu, nous présentons une discussion sur les pistes d'amélioration du dispositif de formation en fonction du retour sur les attentes qui ont fait l'objet de l'évaluation de l'impact.

Discussion

Prise de recul sur les actions entreprises, évaluation de la méthode et pistes d'amélioration

Dans le contexte du Séminaire international de recherche en lien avec le projet CORETEV (SIR-CORETEV) qui s'est tenu à Lausanne en mai 2022 (Gilles et Rigo, 2022; Rigo et al., 2022) une ébauche de référentiel pour le cycle CGQEIF a été produite et soumise à la critique des participants du séminaire. Le tableau 1 présente la version amendée par les pairs, qui constitue un document de référence utilisé lors des mises à l'essai dans les deux contextes évoqués plus haut.

Certaines de ces prescriptions inscrites dans le référentiel n'ont pas été mises en œuvre lors des mises à l'essai, par exemple, l'exploitation de bases de données qui n'était pas pertinente dans les deux contextes.

Dos Santos Paulo (2023) s'est livré, après l'évaluation de la formation, à une analyse de l'application du cycle CGQEIF avec la méthode KISS (Hyatt, 2020 ; Warner, 2020) en partant des résultats de la seconde mise à l'essai dans le contexte FORDIF. KISS est l'acronyme de *Keep, Improve, Start, Stop*. Pour chaque étape du cycle CGQEIF mis en œuvre lors de l'évaluation de l'impact du module de formation FORDIF, Dos Santos Paulo s'est posé les quatre questions suivantes : 1) Qu'est-ce qui a bien fonctionné et qu'il faudrait garder (*Keep*) ? 2) Qu'est-ce qui pourrait être amélioré (*Improve*) ? 3) Qu'est-ce qu'on aurait pu faire de plus (*Start*) ? 4) Qu'est-ce qu'il faudrait arrêter (*Stop*) ? Ces questions ont guidé le processus d'amélioration du cycle proposé dans la section suivante.

Nous présentons dans le tableau 2 une synthèse de ses analyses basées sur ces quatre questions.

Conclusion

Le modèle du cycle CGQEIF que nous proposons permet d'évaluer l'impact d'une formation en mesurant les retours sur attentes (ROE). Ce niveau 4 du modèle de Kirkpatrick s'intéresse aux effets globaux d'une formation et à son impact. Alors que le niveau 1 porte sur l'évaluation de la satisfaction des participants en prenant en compte leurs réactions après formation, les niveaux 2 et 3 s'intéressent quant à eux aux retombées pour la personne formée en ce qui concerne les apprentissages réalisés (niveau 2)

et le transfert des acquis dans l'environnement professionnel (niveau 3). La littérature dans le domaine montre la faible fréquence de mise en œuvre du niveau 4, qui est de loin le moins utilisé dans l'évaluation des formations. Et ce n'est pas parce qu'il est moins utile, mais parce qu'il est perçu comme plus difficile à mettre en œuvre, en particulier lorsqu'on souhaite s'assurer de la qualité des données récoltées. Il existe donc un besoin quant à la conception et à l'instrumentation d'une méthode d'élaboration d'évaluations de l'impact des formations qui soit robuste et aisément mise en œuvre, et ce, dans des contextes de formation très divers.

C'est pour répondre à ce besoin que nous avons conçu et mis à l'essai dans deux contextes différents une méthode en cycle qualité que nous avons intitulée cycle de Construction et de gestion qualité des évaluations de l'impact en formation, ou cycle CGQEIF, basée, comme le cycle PDCA dont il est inspiré, sur un paradigme d'amélioration continue des processus, où chaque mise à l'essai apporte des éléments qui bonifient la mise à l'essai suivante. Dans la mise au point théorique du cycle CGQEIF, nous avions imaginé récolter dans des bases de données existantes beaucoup d'informations concernant l'accomplissement des attentes liées à la formation. Nos mises à l'essai nous ont appris que cela dépend du type de ROE que les commanditaires d'une formation mettent en lumière lorsque nous les interrogeons sur les enjeux des formations évaluées. Notre hypothèse de consultation de bases de données reste probablement valide pour apporter des informations sur l'accomplissement des attentes, mais elle est encore à confirmer étant donné que nous n'avons pas eu l'occasion de consulter de telles bases de données dans les contextes de nos deux mises à l'essai. Dans l'état actuel de notre R-D, ce sont les questionnaires (incluant les comparaisons pré/post) et les *focus groups* qui se sont avérés des outils fiables et praticables.

Pour la mise à l'essai du cycle CGQEIF, nous avons délibérément choisi des contextes de formation assez différents qui devaient nous permettre d'améliorer et de commencer à généraliser ce modèle. Dans cet ordre d'idée, nous entamons une nouvelle mise à l'essai dans un contexte encore différent des deux qui sont présentés ici : l'étude de l'impact d'une réforme des études de bachelier en médecine dans une université en Belgique néerlandophone, l'Université de Hasselt (UHasselt). Cette réforme complète la pédagogie centrée sur l'étudiant par un enseignement centré sur le patient. Cette forme de pédagogie universitaire se centre également sur les compétences liées aux différents rôles qu'un médecin est appelé à jouer. Nous avons démarré cette nouvelle mise à l'essai du cycle CGQEIF en nous concentrant, comme

nos deux premières mises à l'essai nous l'ont appris, sur la définition des enjeux avec les parties prenantes de cette formation universitaire. Dans cette démarche, nous nous sommes appuyés sur le modèle de compétences utilisé à l'UHasselt, soit le modèle CanMEDS4, développé au Canada.

Nous espérons que, dans ce contexte, le cycle CGQEIF pourra encore être amélioré, notamment lors de l'étape de macrorégulation avec la méthode KISS. Il faudra cependant entreprendre d'autres mises à l'essai dans d'autres contextes de formation tout au long de la vie pour stabiliser les étapes du cycle CGQEIF.

Sur le plan des perspectives futures liées au cycle CGQEIF, nous envisageons une instrumentation à l'aide des technologies numériques. En effet, l'usage accru du numérique dans l'évaluation des formations concerne aussi l'évaluation du retour sur les attentes. Nous pourrions améliorer la praticabilité de la méthode en offrant aux évaluateurs une plateforme en ligne qui les aiderait dans la mise en œuvre de chaque étape du cycle. Un autre avantage d'une instrumentation numérique résiderait dans la facilitation du travail collaboratif à distance lors de la construction et de la gestion des dispositifs d'évaluation de l'impact des formations.

Enfin, ne perdons pas de vue que les informations apportées par l'évaluation des retours sur attentes alimentent la régulation et l'amélioration continue de la formation elle-même. Dès lors, une autre perspective future réside dans l'adoption d'une approche longitudinale pour les évaluations de l'impact des formations. Cela comprendrait un suivi de la mise en œuvre de plusieurs cycles CGQEIF qui se succéderaient à la suite de formations répétées dans le temps, comme dans le cas du cursus d'études en médecine de l'UHasselt. Nous pourrions ainsi mesurer à l'aide des cycles CGQEIF successifs les effets de régulations continues dans le dispositif de formation sur l'accomplissement des attentes des parties prenantes.

Révision linguistique: Marie-Claire Legaré

Mise en page: Emmanuel Gagnon

Résumé en portugais: Eusébio André Machado

Réception: 29 septembre 2023

Version finale: 22 septembre 2024

Acceptation: 20 février 2025

4. (<https://www.royalcollege.ca/ca/fr/canmeds/canmeds-framework.html>).

Tableau 1
Ébauche de référentiel lié au cycle CGQEIf élaboré lors du SIR CORTEVE

	Buts poursuivis	Opération et produits finaux	Ressources	Contrôles qualité
0. Enjeux	Clarifier les intentions liées à l'organisation de la formation afin d'identifier les dimensions d'impact et les types de données qui pourraient informer sur l'atteinte (ou non) des impacts recherchés.	Questionner les commanditaires pour savoir pourquoi ils organisent la formation et le problème qu'ils cherchent à solutionner avec la formation.	Les commanditaires de la formation dont on évalue l'impact.	Légitimité et pertinence des acteurs interrogés.
1. Analyse	Identifier les acteurs sur lesquels la formation a un effet et leurs attentes. Identifier les facteurs (autres que la formation) qui pourraient influencer l'impact.	Formaliser les types de données qui permettent d'objectiver le problème de l'impact et de voir évoluer la situation.	Les organisateurs de et les bénéficiaires de la formation et les parties prenantes ayant des attentes vis-à-vis des personnes formées.	Légitimité et pertinence des acteurs interrogés.
2. Design	Identifier les sources de données en lien avec le résultat de l'étape d'analyse. Établir un plan de construction de dispositif de recolte de données, incluant son agencement dans le temps ^[1] .	Déterminer les sources de données en tenant compte des données existantes (E), des données à prélever (P), des données internes (i) et des données externes (e) ce qui donne quatre binômes : Ei, Ee, Pi, Pe	Bases de données internes/externes ^[2]	Validité et fidélité des bases de données

Tableau 1 (suite)
Échancrure de référentiel lié au cycle CGQEIFF élaboré lors du SIR CORTEEV

Buts poursuivis		Opération et produits finaux	Ressources	Contrôles qualité
3. Opérationnalisation	Concevoir et élaborer les instruments de récolte et/ou les bases de données à créer, ou encore les liaisons entre bases de données existantes.	Créer des questionnaires et des instruments de récolte de données. Établir des <i>query</i> de bases de données existantes.	Créateurs d'instruments de récolte de données. Acteurs détenteurs de bases de données. Gestionnaires de données.	Expertise des créateurs. Structuration des bases de données. Intégrité des données.
4. Information	Provoquer l'adhésion des personnes qui seront prendront part au dispositif de récolte des données.	Créer un plan de communication pour ces personnes.	Responsable des communications. Responsables des bases de données existantes. Personnes interrogées dans le cadre de la création de bases de données.	Qualité du plan de communication: pertinence des canaux, lisibilité, clarté et transparence des informations livrées et des buts poursuivis.
5. Collecte des données	Rassembler les données nécessaires aux étapes suivantes.	Activer les instruments de l'étape 3 selon le plan de l'étape 2.	Bases de données. Gestionnaires de données FORDIF. Personnes interrogées.	Conformité avec le plan (étape 2). Utilisation adéquate des outils.
6. Traitement	Synthétiser les données afin d'informer sur les impacts et sur leurs différentes dimensions définies à l'étape d'analyse.	Réaliser les traitements utiles pour les rétroactions (étape 7).	Gestionnaires du dispositif d'évaluation de l'impact ^[3] .	Validité et fidélité des résultats obtenus par les traitements.

Tableau 1 (suite)
Ébauche de référentiel lié au cycle CGQEIf élaboré lors du SIR CORTEV

Buts poursuivis	Opération et produits finaux	Ressources	Contrôles qualité
7. Rétroaction	Donner un retour aux différents acteurs participant à l'évaluation de l'impact et identifiés à l'étape d'analyse.	Traduire les résultats en informations compréhensibles par ces différents acteurs et utilisables à des fins d'amélioration de la formation.	Gestionnaires du dispositif d'évaluation de l'impact. Validité et diagnosticité des rétroactions.
8. Macrorégulation	Recul critique sur les opérations et produits des étapes précédentes en vue d'améliorer un prochain cycle d'évaluation d'impact.	Interroger les acteurs concernés par les différentes étapes. Méthode KISS ^[4] . Représentants d'acteurs qui participent au dispositif d'évaluation de l'impact.	Gestionnaires du dispositif d'évaluation de l'impact. Prise en compte systématique des critères de qualité en évaluation : validité, fidélité, sensibilité, diagnosticité, équité, praticabilité, diagnosticité et authenticité.

^[1]Cela permettrait d'éliminer certaines variables confondantes identifiées à l'étape 1.

^[2]Si les bases de données n'existent pas, il faut convaincre la hiérarchie de l'organisation d'en créer (ce qui est plus simple si ces bases de données peuvent servir un autre but que la mesure de l'impact).

^[3]Personnes responsables de la gestion centrale d'une organisation, qui connaissent l'objet social de l'organisation, qui se trouvent souvent dans des structures de coordination au niveau de la direction. Ces personnes auraient suivi le processus depuis sa conception.

^[4]KISS = $K \rightarrow KEEP; I \rightarrow IMPROVE; S \rightarrow START; S \rightarrow STOP$.

Tableau 2
Synthèse des résultats de l'application de la méthode KISS aux résultats de la mise à l'essai du cycle CGQEIf dans le contexte FORDIF

	Keep	Improve	Start	Stop
0. Enjeux	Cette réflexion préalable sur les enjeux est particulièrement pertinente.		Utiliser le modèle RUFDATA à cette étape (Saunders, 2000).	
1. Analyse	Coconstruire la matrice acteurs/attentes avec les responsables.	Améliorer l'approche micro-méso-macro.	Créer un comité CGQEIf avec des représentants micro-méso-macro.	S'il y a un comité CGQEIf, il n'est plus nécessaire de chercher des personnes ressources.
2. Design	Validation du plan du dispositif par les parties prenantes			
3. Opérationnalisation	Banque d'items structurée en fonction de la matrice acteurs/attentes			
4. Information	Sensibilisation des parties prenantes aux enjeux et aux modalités de l'évaluation de l'impact de la formation	Meilleure participation des répondants si information par les responsables	Sensibiliser les participants en signalant en amont de la formation qu'il y aura une évaluation de l'impact.	L'équipe d'évaluation (des inconnus) ne doit plus prendre en charge l'information.

Tableau 2 (suite)
Synthèse des résultats de l'application de la méthode KISS aux résultats de la mise à l'essai du cycle CGQEIF dans le contexte FORDIF

		<i>Keep</i>	<i>Improve</i>	<i>Start</i>	<i>Stop</i>
5. Collecte des données	Praticabilité de la récolte de données	Stratégies de propagation des questionnaires et des <i>focus-groups</i>	Inclure l'idée de participation à la collecte dans un contrat de formation.		
6. Traitement	Complémentarité des analyses quanti et quali	Mieux tenir compte des contextes des parties prenantes.			
7. Rétroaction	Visualisation des attentes qui ont été satisfaites	Modalités de diffusion auprès des parties prenantes			
8. Macrorégulation	Méthode KISS	Mise en œuvre systématisque	Traçabilité des propositions d'amélioration		

LISTE DE RÉFÉRENCES

- AERA-APA-ACME (1999). *Standards for International and Psychological Testing*. Washington: American International Research Association.
- AERA APA NCME (2014). *Standards for International and Psychological Testing*. Washington: American International Research Association.
- Alliger, G. M. & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, 42(2), 331
- Alsalalah, A. & Callinan, C. (2022). The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959-2020). *Industrial and Commercial Training*, 54(1), p. 36-63.
- Ammons, D. N. & Niedzielski-Eichner, P. A. (1985). Evaluating Supervisory Training in Local Government: Moving Beyond Concept to a Practical Framework. *Public Personnel Management*, 14(3), 211-30.
- Best, M. & Neuhauser, D. (2006). Walter A Shewhart, 1924, and the Hawthorne factory. *BMJ quality & safety*, 15(2), 142-143.
- Boutillier, C., Jeanrenaud, L., Gilles, J.-L., Bouche, L. & Quillet Cotting, J. (2020). Healthcare Students on Placements: a Cyclical Quality Method for Satisfaction Assessments. *Medical Science Educator*, 30, 1427-1435. <http://hdl.handle.net/20.500.12162/4096>
- Chochard, Y. (2012). *Les variables influençant le rendement des formations managériales*. [Thèse de doctorat en Sciences économiques et sociales, Université de Fribourg].
- Chochard, Y., De Souza, G. & Gilles, J.-L. (2017). Évaluation du transfert des acquis de formation en milieu de travail: une nouvelle méthode. Bulletin de l'Observatoire compétences-emploi - OCE, 7, 1-6. <http://hdl.handle.net/20.500.12162/127>
- Chochard, Y. & Gilles, J.-L. (2016). Optimiser l'impact de la formation continue universitaire: développement d'un nouvel instrument d'évaluation du transfert des acquis sur le terrain professionnel. Dans O. Adıgüzel (dir.), *Teaching and Training Today for Tomorrow*. Eskisehir, Turquie: Anadolu University. <http://hdl.handle.net/20.500.12162/697>
- Deming, E. (1986). *Out of the Crisis*. MIT, Center for Advanced Engineering Study.
- Dos Santos Paulo, M. (2023). Mise en Place et Amélioration du Cycle de Construction et de Gestion Qualité d'Évaluations de l'Impact de Formations (cycle CGEIF) à la Formation en Direction d'Institutions de Formation (FORDIF) [Mémoire de maîtrise inédit, Haute école pédagogique de Lausanne].
- Dunberry, A. & Péchard, C. (2007). L'évaluation de la formation dans l'entreprise : état de la question et perspectives. CIRDEP-UQAM.
- Espace Compétences (2025). <https://espace-competences.ch/>
- FORDIF (2023). Le pilotage d'un établissement de formation à l'ère du numérique. <https://www.fordif.ch/formation/pilotage-numerique/>

- Gascon, H. & Germain, M. P. (2017). La recherche-développement, une méthode centrée sur l'élaboration d'un produit. Dans P. Beaupré, R. Laroui & M.-H. Hébert (dir.), *Le chercheur face aux défis méthodologiques de la recherche. Freins et leviers* (p. 121- 134). Presses de l'Université du Québec.
- Gérard, F. M. (2003). L'évaluation de l'efficacité d'une formation. Gestion 2000.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires - Mise au point d'indices édumétriques d'analyse de la qualité spectrale des évaluations des acquis des étudiants universitaires et application aux épreuves MOHICAN check up '99 [Thèse de doctorat, Université de Liège]. <https://orbi.uliege.be/handle/2268/2217>
- Gilles, J.-L. (2023). *Ingénierie des évaluations en formation. Cycle N4 CAS IEF*. Support de cours non publié, Haute école pédagogique, Vaud.
- Gilles, J.-L. & Charlier, B. (2020). Dispositifs d'évaluation à distance à correction automatisée versus non automatisée : analyse comparative de deux formes emblématiques. *Évaluer: journal international de recherche en éducation et formation - e-JIREF*, Numéro Hors-série (1), 143-154. <http://hdl.handle.net/20.500.12162/3865>
- Gilles, J.-L. & Chochard, Y. (2022). Construction Cycle and Quality Controls for Training Transfer Evaluations in Lifelong Learning Programs in Quebec and Switzerland. *Creative Education*, 13, 3533-3558. <http://hdl.handle.net/20.500.12162/5920>
- Gilles, J.-L., Chochard, Y., Berset, T., Bieri, S., El Guermai, J., Graveline, A., ... Rupp-Nantel, K. (2017). *Cycle de construction et de gestion qualité d'évaluation du transfert (CGQET)*. Université du Québec à Montréal. <http://hdl.handle.net/20.500.12162/122>
- Gilles, J.-L. & Leclercq, D. (1995). Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG. Communication présentée à Symposium International sur la Rénovation Didactique en Biologie, Tunis, Tunisie. <http://hdl.handle.net/20.500.12162/2709>
- Gilles, J.-L. & Rigo, J.-M. (2022). Conception et mise en œuvre d'un cycle de Construction et de gestion qualité d'évaluation de l'impact de formations (cycle CGQEIf). Communication présentée au Congrès ACFAS 2022, Symposium 526 - Formation professionnelle en milieu de travail: qualité des dispositifs offerts et effets pour les travailleurs formés, les formateurs et les employeurs, Montréal, Canada.
- Goldwasser, D. (2001). Beyond ROI. *Training*, 38(1).
- Harvey, S. & Loiselle, J. (2009). Proposition d'un modèle de recherche développement. *Recherches qualitatives*, 28(2), 95-117. <https://doi.org/10.7202/1085274ar>
- Hyatt, M. (2020). Achieving together: The full focus system for teams. https://michaelhyatt.com/wp-content/uploads/dlm_uploads/2020/12/Achieving_Together_Interactive_Workbook.pdf
- Juran, J. M. (2003). *Juran on leadership for quality*. Simon and Schuster.
- Kaeser Boutillier, C., Jeanrenaud, L., Gilles, J.-L., Bouche, L., & Quillet Cotting, J. (2022). Évaluer et réguler la formation pratique en milieu hospitalier à l'aide d'un cycle de Construction et de gestion qualité des évaluations de satisfaction (cycle CGQES): enjeu et impact d'une expérience pilote au CHUV de Lausanne. Dans P. Detroz, L. Mottier Lopez et E. Issaieva (dir.), *Livret des résumés*. <http://hdl.handle.net/20.500.12162/5916>
- Kirkpatrick, D. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13(11), 3-26.

- Kirkpatrick, D. L. (1978). Evaluating In-House Training Programs. *Training and Development Journal*, 32(9), 6-9.
- Kirkpatrick, D. (1996). Great Ideas Revisited. Techniques for Evaluating Training Programs, Revisiting Kirkpatrick's Four-Level Model. *Training & Development*. 50(1), p. 54-59.
- Kirkpatrick, D. & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publishers.
- Loiselle, J. (2001). La recherche développement en éducation : sa nature et ses caractéristiques. Dans M. Anadón (dir.), *Nouvelles dynamiques de recherche en éducation* (p. 77-98). Presses Université de Laval.
- Loiselle, J. & Harvey, S. (2007). La recherche développement en éducation : fondements, apports et limites. *Recherches qualitatives*, 27(1), 40-59.
- Meignant (2014). *Manager la formation* (9^e éd.). Éditions Liaisons.
- Monnot, A. (2014). Les pratiques d'évaluation de la formation professionnelle et leurs déterminants dans les entreprises du SBF 120. *Management & Avenir*, 68, 92-111. <https://doi.org/10.3917/mav.068.0092>
- Peters, T. J. & Waterman Jr, R. H. (1982). *In Search of. Of Excellence*. Harper and Row.
- Phillips, J. J. (1997). *Handbook of training evaluation and measurement methods* (3^e éd.). Gulf Publishing Company
- Phillips & Stone (2002). How to measure Training Results, tracking 6 key indicators. McGraw-Hill.
- Plan d'études et moyens d'enseignement romand (PER) (2023). Plan d'études romand. <https://portail.ciip.ch/per/domains>.
- Pottiez, (2019). Nouveau modèle de Kirkpatrick : quels apports pour la fonction formation ? Le blog de C-Campus. <https://www.blog-formation-entreprise.fr/nouveau-modele-kirkpatrick-apports-fonction-formation/>
- Pottiez, J. (2021). L'évaluation de la formation : Pilotez et maximisez l'efficacité de vos formations. Dunod.
- Rigo, J.-M., Ef Jaafari, S., El Kamoun, N. & Gilles, J.-L. (2022, mai). De CORETEV à GIRIEF : un réseau international de recherches et d'expertises en ingénierie des évaluations. Communication présentée à Séminaire international de recherche CORETEV 2022 (SIR CORETEV 2022), Lausanne, Suisse. <http://hdl.handle.net/20.500.12162/5893>
- Saunders, M. (2000). Beginning an evaluation with RUFDATA: theorizing a practical approach to evaluation planning. *Evaluation*, 6(1), 7-21.
- Shewhart, W. A. (1931). Statistical method from an engineering viewpoint. *Journal of the American Statistical Association*, 26(175), 262-269.
- Shewhart, W. A. (1958). Nature and origin of standards of quality. *The Bell System Technical Journal*, 37(1), 1-22.
- Suarez, J. G. (1992). Three Experts on Quality Management: Philip B. Crosby, W. Edwards Deming, Joseph M. Juran (Vol. 11). Department of the Navy TQL Office.
- Sugrue, B. & Kim, K. (2004). The 2004 ASTD state of the industry report.
- Twitchell, S., Holton, E. F. & Trott Jr, J. W. (2000). Technical training evaluation practices in the United States. *Performance Improvement Quarterly*, 13(3), 84-109.

- Van Buren, M. E. & Erskine, W. (2002). *The 2002 ASTD state of the industry report*. American Society of Training and Development.
- Van der Maren, J.-M. (2014). La recherche appliquée pour les professionnels - Éducation, (para)médical, travail social. Éditions de Boeck.
- Wang, G. G., Dou, Z. & Li, N. (2002). A system approach to measuring return on investment for HRD interventions. *Human Resource Development Quarterly*, 13(2), 203-224.
- Warner, M. H. (2020). Project lessons learned, or: Asking the team to “KISS”. *The Project Management Blueprint*. <https://www.theprojectmanagementblueprint.com/blog/project-overview/project-lessons-learned-via-kiss-method>.