

Qu'est-ce qu'évaluer ? La carte du territoire

Quels sont les principes *invariants* de tout acte évaluatif, quelle que soit la forme empirique sous laquelle il est réalisé sur le terrain ? Le modèle présenté ci-dessous a pour fonction de rassembler ces invariants *in a nutshell*, de manière synthétique, afin de faciliter l'identification des opérations propres à tout acte évaluatif. En quoi se distinguent-elles les unes des autres ? Quelle est la cohérence globale de cet acte ?

Le modèle « archaïque » de l'évaluation... et son détournement

L'invariant premier se présente sous la forme d'une dualité : tout acte évaluatif est caractérisé, au minimum, par une « performance » et un « verdict ».

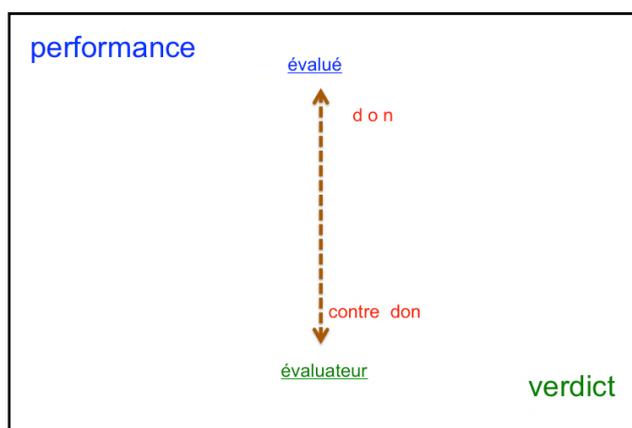


Figure 1 : les composantes du contrat évaluatif duel

La performance (au sens du verbe anglais *to perform* : accomplir une action impliquant une certaine habileté, un entraînement) est ce qui est évalué¹ : elle doit être observable, c'est à dire perceptible auditivement, visuellement, gustativement (évaluation d'un plat), tactilement (un tissu) ou olfactivement (un parfum). Elle peut être envisagée du point de vue de son aboutissement, le *produit* fini, ou du point de vue des *processus* ayant conduit à sa réalisation.

Le « verdict » est l'énoncé public d'une vérité sur la performance évaluée. Le terme évoque non sans raison le système judiciaire et son étymologie renvoie à un « dire vrai ». Que cette vérité soit souvent discutable est une évidence mais ce qui importe, en réalité, c'est la confiance *a priori* dans l'expertise *professionnelle* de celui qui prononce le verdict, au nom d'une *institution* socialement reconnue comme ayant compétence à le faire. Cette dernière fixe les « règles du jeu² » en fonction desquelles ses agents sont censés évaluer.

Or cette confiance implicite peut disparaître si l'institution se révèle incapable d'instaurer des règles du jeu professionnelles communes : les verdicts sont alors ressentis comme in-justes, c'est à dire dépourvus de *justesse* ; c'est aussi le cas s'ils semblent favoriser de manière récurrente certaines personnes ou catégories sociales

¹ C'est la performance qui est évaluée et non, contrairement à ce que laisse entendre un raccourci de langage tenant peu compte de la déontologie de l'enseignement, la personne de l'évalué. On n'évalue pas les élèves mais leurs apprentissages, leurs productions...

² C'est le titre d'un ouvrage de Charles Hadji : *L'évaluation, règles du jeu* (ESF éditeurs)

Introduction à la problématique de l'évaluation

ou culturelles et en défavoriser d'autres : ils sont alors ressentis comme injustes au sens de contraires à la *justice*. Le défaut de justesse et le défaut de justice sont les deux faiblesses majeures du système d'évaluation scolaire actuel en France.

La confiance ne tient pas fondamentalement aux qualités ou aux défauts des individus ni à leurs traits de caractère (sévérité, bienveillance etc.), bien qu'ils influent sur la relation réelle. La raison en est que cette relation n'est pas fondamentalement psychologique mais repose sur un contrat social. Il faut entendre par « contrat » un système d'attentes, le plus souvent implicites, lié aux places et rôles attribués aux « contractants », l'évaluateur et l'évalué, par les représentations sociales et culturelles en vigueur dans une communauté. Ce contrat s'apparente aux structures élémentaires des sociétés dites « archaïques » étudiées par l'anthropologie. Marcel Mauss, dans son *Essai sur le don*³, montre que, dans ces sociétés, le système d'échange de dons et contre-dons (donner, recevoir, rendre) constitue le fondement du lien social initial⁴. L'évaluation scolaire reproduit cette structure élémentaire : le *don* par un élève d'un travail (devoir, récitation, interrogation écrite etc.) implique la réciprocité d'un *contre-don* par l'enseignant sous la forme d'une note, d'une appréciation⁵ ou autre.

Le contrat évaluatif prend aujourd'hui la forme d'une relation asymétrique illustrée par ce dessin de Charb paru dans les *Cahiers Pédagogiques*. Entre défiance réciproque et affrontement ouvert, l'évaluation apparaît en partie détournée de son sens didactique originel vers un système de contrainte fondé sur une pression notatoire continue, censée « motiver » les élèves à fournir un travail minimum. Le rapport de réciprocité entre don et contre-don s'inverse à 180° : le contre-don tend à devenir la motivation et la justification principales du don.



Introduire une dose supplémentaire de bienveillance ne suffira pas à rompre avec cette relation duelle fossilisée. Il faut choisir une focale bien plus large pour révéler les autres composantes de l'acte évaluatif, masquées par l'étroitesse de la focalisation actuelle. Il devient alors possible de concevoir des alternatives à ce contrat évaluatif détourné. Une nouvelle « carte du territoire » de l'évaluation se dessine alors.

³ *Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques*. Marcel Mauss, 1923-1924.

⁴ Le contrat social de J.J. Rousseau

⁵ On trouve ce lien entre évaluation et don entre autres chez Y. Chevallard : *Le moment de l'évaluation, ses objets, ses fonctions : déplacements, ruptures, refondation*.

http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Le_moment_de_l_evaluation_ses_objets_ses_fonctions.pdf

La carte du territoire de l'évaluation

La relation « archaïque » initiale ne disparaît pas de ce nouveau territoire mais elle est relativisée par son insertion dans un réseau plus complexe de notions et de relations, qui en modifie le sens. On peut représenter ce réseau sous une forme inspirée du « carré sémiotique⁶ » : les termes sur chaque côté et à chaque angle correspondent aux principales opérations de l'acte évaluatif et les relations entre elles marquent leur interdépendance. A chacune des étapes l'évaluateur doit opérer des choix entre des positions et des principes en tension les uns par rapport aux autres.

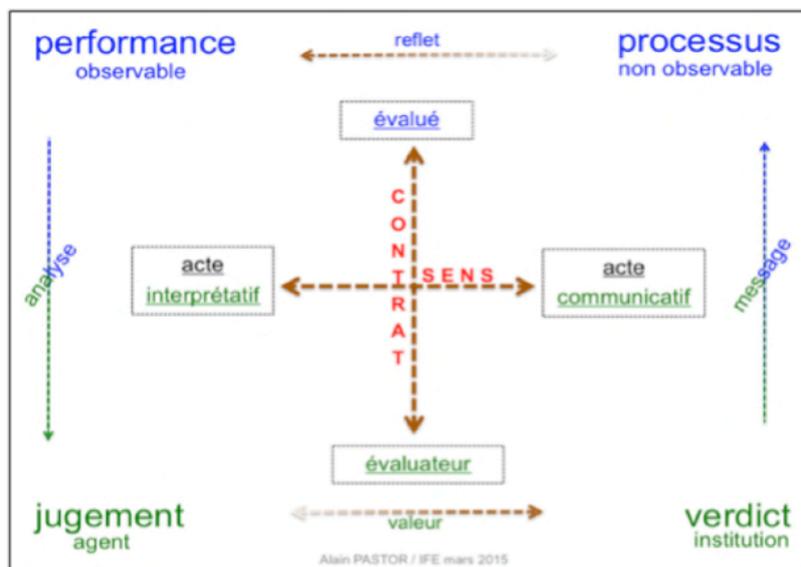


Figure 2 : les composantes du territoire de l'évaluation

La performance reflète-t-elle les apprentissages à évaluer ? (haut du carré)

Pour ce qui est de l'usage d'une langue, une performance est une action langagière, audible (écoute, parole) et/ou visible (lecture, écriture). Cette performance est le reflet extériorisé des processus mentaux (intériorisés) requis par sa réalisation : perceptions, représentations personnelles ou collectives, opérations psychomotrices (articulatoires, kinésiques etc.) et mentales (énonciatives, logico-mathématiques, déductives, inductives etc.), schèmes (ensembles organisés d'opérations intégrées, mobilisés selon le type de situation). Une compétence (figure 3) est l'association d'une performance avec les processus mentaux qui en sont à l'origine. Or, ces processus étant peu accessibles à l'observation comme à l'introspection, il n'est possible ni d'en attester à coup sûr la réalité ni de les modifier par une intervention externe *directe*, comme celle d'un enseignant. Ce dernier ne peut intervenir qu'*indirectement*, en créant des situations propices à leur développement, leur modification ou leur renforcement.

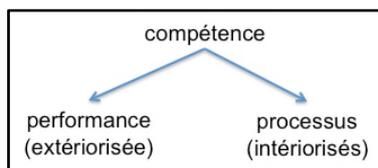


Figure 3 : compétence, performance, processus

⁶ Utilisé par des sémioticiens comme A.J. Greimas (schéma actanciel du conte) ou F. Rastié.

Or, ce sont justement ces processus qu'il est intéressant de connaître dans l'acte évaluatif. Alors que, dans la vie professionnelle, on évalue la compétence d'un travailleur principalement à partir de ses performances (ses résultats, sa productivité), dans un système éducatif ce n'est normalement pas le cas : la performance n'est intéressante que par les indices qu'elle fournit sur les processus mentaux mis en œuvre par les élèves. Mais, à défaut de pouvoir les observer directement, l'enseignant en est réduit à des conjectures qui, avec l'expérience, peuvent devenir plus fiables mais sans jamais perdre leur caractère conjectural.

Cela rend la conception des tâches d'évaluation (première phase de l'acte évaluatif) problématique. Telle tâche reflète-t-elle réellement les compétences visées ? Permet-elle d'évaluer de façon pertinente ce qu'on veut évaluer ? Cette question cruciale implique pour l'évaluateur une analyse précise à la fois des compétences visées et des caractéristiques de la tâche censée les attester : on n'évalue pas la même chose avec une tâche ouverte ou fermée, complexe ou simple, « authentique » ou académique, avec aides ou sans etc. Le choix de la tâche pertinente s'effectue parmi une gamme étendue d'activités langagières, issues de contextes socioculturels différents, pas uniquement scolaire. Par ailleurs, l'évaluateur ne peut pas totalement anticiper les facteurs externes ou internes susceptibles d'influer sur la manifestation de la compétence, sur la réussite ou l'échec à accomplir la tâche. Bref, il demeure toujours une incertitude quant à ce que reflète réellement la performance d'un élève. Aussi considère-t-on que l'acquisition d'une compétence ne peut être attestée de façon plausible que sous certaines conditions : par exemple, si elle se manifeste à l'occasion d'au moins deux ou trois performances distinctes, si l'évalué est capable d'expliquer ou de justifier sa performance etc.

Le jugement évaluatif (1) : l'interprétation de la performance (côté gauche du carré)

La performance une fois réalisée, l'évaluateur porte sur elle un jugement évaluatif. C'est la seconde phase. Comment ce jugement se construit-il ?

Certaines tâches, comme les tests dits « objectifs » de type QCM, vrai/faux etc., réduisent le jugement à un décompte : nombre d'items réussis ou d'erreurs commises, somme de micro-capacités etc.. Le jugement évaluatif est principalement quantitatif, il est fondé sur une conception de l'évaluation comme *mesure*, censée garantir son objectivité. Mais cette prétendue objectivité se paie au prix fort par une perte de validité de l'évaluation : les décomptes et mesures ne permettent de rendre compte que d'une infime partie des composantes de l'activité langagière. Etant donné la complexité et la diversité de cette dernière, la validité du jugement évaluatif ne peut résulter que d'une démarche intellectuelle *qualitative*, qui conjugue des points de vue analytique et synthétique (holistique) et tienne compte de l'explicite et de l'implicite.

L'analyse qualitative fait plus appel aux ressources propres de l'évaluateur que les procédures quantitatives : son expérience mais aussi ses valeurs (dimension éthique) et sa conception de l'activité langagière (dimension didactique). La conception la plus répandue, inspirée du cognitivisme⁷, considère l'activité langagière comme *traitement* de messages —en réception ou en production— envisagés surtout du point de vue de leurs composants linguistiques (phonèmes ou graphèmes, structures, lexique, phrases). Une seconde conception intègre la précédente dans une perspective plus large, celle de l'*usage* du langage par les « acteurs sociaux » : elle considère les

⁷ Courant de pensée principalement nord-américain qui étudie les phénomènes qui se produisent à l'intérieur de la boîte noire mentale (« inside the black box »).

activités langagières sous l'angle de leurs dimensions textuelle (genres de textes) et discursive (types de discours) et de leur cohérence (pragmatique, sociolinguistique) avec les contextes sociaux et culturels⁸. L'analyse de la performance par l'évaluateur dépend pour une large part de la conception à laquelle il se réfère implicitement ou explicitement (son « référent »). Ces divergences s'ajoutent à la complexité *per se* de toute activité langagière pour aboutir à des *interprétations* différentes d'une même performance. Le jugement évaluatif est un acte interprétatif.

Le jugement évaluatif (2) : la confrontation aux échelles de valeur (bas du carré)

Les inévitables divergences d'interprétation d'une même performance sont une des raisons pour lesquelles il est souhaitable de créer les conditions d'un consensus par rapport aux caractéristiques des performances évaluées. A l'échelle de tout un système éducatif, voire des systèmes éducatifs d'un groupe de pays comme l'U.E., c'est une entreprise extrêmement délicate, tant divergent les vues sur les dimensions de l'activité langagière qu'il importe de privilégier. Cette visée d'harmonisation conduit à mettre à la disposition des enseignants des *référentiels* communs, composés d'inventaires plus ou moins structurés (taxonomies) de descripteurs et de critères de compétences, et à déterminer des *standards* : tel palier de compétence est censé être atteint à tel âge scolaire. Le *CECRL*, le socle commun ou le *Portfolio européen* des langues sont des résultats tangibles de ces démarches à différents niveaux.

Que fait l'évaluateur de ces standards ? C'est la seconde phase du jugement évaluatif, après celle de l'analyse et de l'interprétation, et la troisième de l'acte évaluatif. Il ne peut se contenter d'analyser la performance : en effet, l'analyse est une opération intellectuelle qui constitue une étape nécessaire mais non suffisante du jugement évaluatif. E/**valuer**, étymologiquement, c'est déterminer la valeur de ce qu'on évalue et, ce faisant, prononcer le verdict dont il a déjà été question. A l'échelle d'un système éducatif, il serait incohérent et injuste que chaque évaluateur utilise sa propre échelle : on ne peut déterminer la valeur d'une performance que par rapport à une échelle de valeurs comparables, C'est la fonction des standards que d'harmoniser les échelles de valeurs, en les formulant par rapport aux attentes de la société, sachant que ces dernières peuvent changer selon l'époque et le lieu⁹. La valeur d'une performance n'est donc ni intrinsèque ni absolue : elle lui est attribuée¹⁰ par l'évaluateur relativement aux standards institutionnels, qui reflètent les conceptions sociales en vigueur de l'activité langagière. Le verdict est toujours relatif, mais l'utilisation des échelles communes empêche de le considérer comme arbitraire. La distinction entre jugement relatif et jugement arbitraire clarifie le débat sur « l'objectivité ». La distinction entre les phases, souvent confondues, du jugement évaluatif —analyse de la performance et détermination de sa valeur— clarifie la démarche de l'évaluateur.

Il faut être conscient, cependant, du rapport problématique des standards avec les compétences réelles. Les compétences telles qu'elles sont décrites et formulées dans les référentiels et les standards ne sont pas les compétences réellement mises en œuvre, ce sont seulement des représentations de ces compétences sous forme de

⁸ Cette seconde conception de l'activité langagière, comme usage social et culturel d'une langue, est celle exposée par le *CECRL*, dénommée « perspective actionnelle ». Mais, pour diverses raisons qu'il est impossible de développer dans le cadre de cette introduction, l'interprétation du *CECRL* par la majorité des enseignants est clairement la première.

⁹ C'est ce qui est en train de se produire à propos du socle commun, qui date seulement de 2005/2006 et dont une nouvelle mouture vient d'être proposée à la ministre par le CSP.

¹⁰ Y. Chevillard, *ibid.*, p. 4.

descripteurs. Un référentiel est aux compétences réelles ce que ce tableau de Magritte est aux pipes utilisées par les fumeurs dans la « vraie » vie : « ceci n'est pas une pipe », ce n'est qu'un tableau représentant une pipe, on ne peut pas le fumer.



Magritte. *Ceci n'est pas une pipe*

Les descripteurs sont déclaratifs (descriptions verbales) alors que les compétences sont opératoires, elles se manifestent dans des situations d'activité. Les descripteurs sont figés dans des tableaux et des grilles alors que les compétences manifestent la capacité des locuteurs à s'adapter à des situations différentes. Les descripteurs sont discrets (ils « découpent » les compétences en composantes distinctes juxtaposées dans des inventaires) alors qu'une compétence réelle est complexe : ses composantes sont intégrées de façon fluide et continue dans une action globale. L'évaluateur doit veiller à ne pas confondre ces deux registres, à éviter l'erreur de considérer les compétences qu'il évalue comme se réduisant à celles formulées dans les standards.

Faut-il, par conséquent, rejeter les standards ? Evidemment non puisqu'ils servent à réduire les écarts d'interprétation entre évaluateurs —sans pour autant les abolir totalement— en établissant des normes communes. Cette harmonisation est essentielle pour réduire l'aléatoire et l'arbitraire dont pâtit l'évaluation. Mais c'est une erreur de réduire le jugement évaluatif à l'utilisation des standards. Loin de constituer le défaut de la cuirasse par lequel se réintroduirait l'arbitraire, les inévitables différences de nature entre standards et compétences réelles (cf. § précédent) permettent à l'évaluateur d'opérer des choix assumés, en toute connaissance de cause, selon les situations d'évaluation. Évalue-t-il par référence aux normes institutionnelles, se réfère-t-il à son jugement propre, au « style » plus ou moins personnel de la production évaluée, ou encore à sa connaissance des apprenants et à leurs progrès ? Selon les caractéristiques et les enjeux de chaque situation d'évaluation, le référent qu'il choisit peut être différent. L'évaluateur compétent n'est pas celui qui se réfère invariablement au même référent, au même point de vue, c'est celui qui est capable de décider par lui-même, en fonction de la situation d'évaluation dans laquelle il intervient, quel est ou quels sont les référents à privilégier.

Il convient avant tout d'éviter la dérive bureaucratique qui consiste à ne plus considérer les tâches réalisées par les élèves que comme des occasions de poser un verdict, de « valider » des compétences, au point d'affirmer que le terme « évaluation » serait désormais remplacé par celui de « validation ». Si c'est le cas, *on n'enseigne plus que ce qu'on évalue (valide), au lieu d'évaluer ce qu'on a enseigné*. Cette inversion des priorités professionnelles et pédagogiques est totalement contre-

productive pour les apprentissages car elle prive l'élève du temps d'apprendre, qui doit être un temps de « silence évaluatif¹¹ ».

La communication du verdict : le message évaluatif (côté droit du carré)

La quatrième phase de l'acte évaluatif est la communication du verdict, sous une forme appropriée, aux divers destinataires concernés, en premier lieu l'évalué lui-même mais aussi ses ayants droits, par exemple ses parents, ou encore l'administration de l'établissement. L'évaluateur produit des messages évaluatifs qui ont une double fonction : ils *informent* le destinataire sur la valeur de la performance et ils *orientent* les décisions à prendre en indiquant des perspectives.

En ce qui concerne l'information, le mieux est souvent l'ennemi du bien : il n'est pas utile que l'évaluateur surcharge le message en rendant compte de l'acte évaluatif dans sa totalité. Il n'a à communiquer que ce qui peut être *compris* des destinataires. En ce qui concerne la décision il n'a à indiquer que ce qui peut leur être *utile*. Mais le message évaluatif est aussi l'occasion d'engager, par une démarche ouverte d'explicitation, une interaction avec les destinataires au cours de laquelle soit ils apportent à l'évaluateur des données complémentaires, soit ils comprennent mieux en quoi consiste (s')évaluer. Le développement de la compétence des élèves à (s')évaluer et à comprendre en fonction de quoi ils sont évalués est un processus *métacognitif* qui influe en retour sur les processus d'apprentissage, précisément ceux sur lesquels l'enseignant ne peut pas intervenir directement : l'évaluation a donc une dimension *métacognitive* qui contribue à la prise de conscience par l'élève des attentes et des moyens pour progresser. Elle permet de sortir du cercle vicieux de la relation duelle figée et d'amorcer le cercle vertueux de la progression.

Le message évaluatif peut revêtir, si nécessaire, une forme synthétique chiffrée. Mais ce chiffrage procède d'une démarche méthodologique totalement inverse à la notation traditionnelle. Le degré de compétence atteint par l'élève n'est pas estimé de façon *différentielle*, en comparant sa note à une moyenne absolue (10/20), repère purement abstrait, ni à la moyenne d'un groupe classe à la composition aléatoire : dans ces deux cas, l'étalonnage est fondé sur des sables aussi mouvants que la moyenne. Pour être cohérent, il doit être établi de façon *référentielle*, en confrontant la performance aux échelles de compétences communes. Dans un second temps, il est alors possible de lui donner une traduction chiffrée selon une échelle de correspondance avec les standards institutionnels correspondant à l'année scolaire suivie par l'élève, indépendamment des autres élèves de son groupe classe, voire de son établissement.

Acte interprétatif et acte communicatif : effets sur le contrat évaluatif

Entre l'évalué et l'évaluateur nous savons qu'il existe un contrat de nature sociale (axe vertical au centre de la carte). Mais quel est le rapport entre l'acte interprétatif et l'acte communicatif, de part et d'autre de l'axe horizontal ? C'est un rapport caractéristique de tout acte de langage : l'acte interprétatif est de nature « propositionnelle », il rend compte de manière aussi rationnelle et argumentée que possible de la compétence de l'élève ; l'acte communicatif est de nature « actionnelle », il contribue, potentiellement, à modifier les rapports entre l'évalué et le savoir qu'il doit construire, mais aussi ceux entre les partenaires du contrat évaluatif. Il lui donne un autre sens : au lieu d'un face à face entre deux rôles sociaux

¹¹ Y. Chevallard, *ibid.*.

figés, le contrat évaluatif se transforme progressivement en coopération entre partenaires.

Cette évolution s'appuie sur la place accordée au langage, à ses fonctions d'explicitation et de négociation, dans les différentes phases de l'acte évaluatif. Le contrat évaluatif archaïque est fondé sur des rapports *implicites*, entre évalué et évaluateur : l'implicite est à l'origine de malentendus incessants. Dès lors qu'on élargit la focale à l'ensemble du champ, qu'on prend en compte d'autres actes évaluatifs, qu'on élucide le rôle des acteurs qui y sont associés, le contrat évaluatif se décale par rapport à sa centration duelle initiale. D'implicite, il devient *explicite*, chacun étant amené à préciser son rôle en donnant son avis sur les exigences et les critères.

Ce processus n'a rien de spontané : les élèves doivent être préparés à cette participation à l'évaluation, en premier lieu la leur ou celle de leurs pairs, mais pas uniquement. C'est un apprentissage de longue haleine, qui devrait être initié le plus tôt possible dans la scolarité afin de ne pas laisser les représentations des élèves se fossiliser autour des diverses formes du contrat évaluatif archaïque et de surmonter la perte de confiance dont il est l'objet. La capacité des élèves à devenir responsables de leurs apprentissages, à développer les compétences leur permettant de s'émanciper (*empowerment*) progressivement de la tutelle de l'enseignant, dépend de leur participation à l'acte évaluatif. Plutôt que de regretter leur manque d'autonomie et de se contenter de leur enjoindre de l'être plus¹², l'évolution des pratiques d'évaluation offre une voie à explorer pour les mettre en situation de devenir progressivement plus autonomes.

De la carte de l'évaluation à la formation IFE

Aucun évaluateur ne se reconnaîtra pleinement dans la caricature de Charb. Aucun non plus, sans doute, ne prétendra évaluer de façon idéale. Chacun est en chemin entre le contrat figé et un contrat évaluatif plus dynamique, avec un curseur personnel plus ou moins proche d'un pôle ou de l'autre. La carte du territoire est un modèle pour continuer son cheminement, non pas au sens où elle décrirait une pratique évaluative idéale mais au sens où sa fonction est d'indiquer les leviers sur lesquels il est possible d'agir, dans la situation professionnelle où on se trouve, pour poursuivre son cheminement personnel ou collectif. L'objet de la formation IFE à laquelle vous avez participé était d'étudier plus précisément certains de ces leviers.

¹² La plus typique des injonctions paradoxales est « sois autonome ».