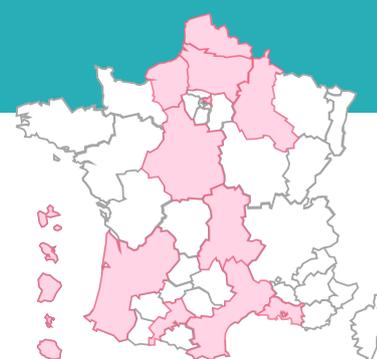


Note méthodologique pour l'évaluation contrefactuelle de l'initiative pour l'emploi des jeunes en 2018

Luc Behaghel - PSE, INRA
Mai 2016



Sommaire

Préambule

1. Qu'est-ce que l'évaluation contrefactuelle ? Application au cas de l'IEJ

- 1.1. Principes de l'évaluation d'impact contrefactuelle
- 1.2. Spécificités de l'IEJ au regard de l'évaluation d'impact contrefactuelle

2. Deux approches pour évaluer l'IEJ

- 2.1. Evaluer la plus-value globale de l'IEJ
- 2.2. Evaluer l'IEJ par les dispositifs qu'elle finance

3. Données et mise en œuvre

- 3.1. Données mobilisables
- 3.2. Mise en œuvre

Références

A partir de 2014, la France a déployé dans le cadre de l'Initiative pour l'emploi des jeunes (IEJ) un ensemble d'actions visant à repérer, accompagner et former des jeunes sortis du système scolaire mais qui ne sont ni en emploi ni en formation (jeunes "NEETs", *Not in Education, Employment or Training*). Cette note méthodologique propose un cadre et des pistes pour l'évaluation d'impact de l'IEJ qui doit être menée d'ici 2018, conformément aux obligations réglementaires de la France vis-à-vis de l'Union européenne. Spécifiquement, elle analyse comment répondre à la question contrefactuelle suivante : "Comment mesurer l'impact de l'IEJ sur les jeunes en France, entendu comme la différence entre les trajectoires de ces jeunes et celles qui auraient prévalu s'il n'y avait pas eu d'IEJ ?"

Cette question d'impact sur les jeunes s'inscrit dans l'ensemble plus large de questions évaluatives que pose l'IEJ, qui est présenté dans les recommandations de la Commission européenne (European Commission, 2015) et repris par les autorités de gestion chargées de l'IEJ en France (DGEFP, 2015). Ce questionnement évaluatif comporte cinq blocs : (i) une évaluation *ex ante* de la pertinence de la stratégie mise en œuvre au regard du contexte économique et de la finalité poursuivie (groupes cibles, objectifs) ; (ii) une évaluation de la mise en œuvre de l'IEJ (dispositif institutionnel, modalités de déploiement, implication des acteurs et des parties prenantes, articulation avec les autres dispositifs contemporains de l'IEJ) ; (iii) une analyse en termes d'efficacité, qui rapporte les résultats obtenus -- en particulier, les trajectoires des jeunes bénéficiaires -- aux objectifs de la politique ; (iv) une analyse en termes d'efficience, qui rapporte ces résultats aux coûts de la politiques ; (v) une analyse d'impact sur les participants. Les quatre premiers blocs de questions sont extrêmement importants pour mettre en évidence les possibilités d'amélioration de la politique, et pour en comprendre les effets. Ils sont largement abordés dans le rapport de synthèse de l'évaluation d'impact de l'IEJ au 31 décembre 2015 (Rapport national, 2015), et les réponses présentées constituent les éléments d'une évaluation d'impact "basée sur la théorie", qui permet un premier jugement sur les effets possibles du dispositif. Seule l'évaluation d'impact contrefactuelle, dont les principes sont rappelés dans cette note, permet cependant de quantifier la contribution de l'IEJ à l'objectif général qui lui a été fixé : l'intégration durable des NEETs sur le marché du travail.

Cette note procède en trois temps. Elle présente d'abord les attendus et le cadre d'une évaluation contrefactuelle, afin de souligner comment ce cadre doit être adapté pour prendre en compte les spécificités de l'IEJ. Dans un deuxième temps, elle analyse deux voies possibles d'évaluation de l'IEJ, l'une globale, l'autre à partir de l'évaluation d'impact de dispositifs particuliers financés par l'IEJ. La troisième partie propose des recommandations pour l'articulation et la mise en œuvre de ces deux approches.

1. Qu'est-ce qu'une évaluation d'impact contrefactuelle ? Application au cas de l'IEJ

1.1. Principes de l'évaluation d'impact contrefactuelle

Avant d'envisager l'évaluation de l'IEJ, il est utile de rappeler succinctement les principes de l'évaluation d'impact contrefactuelle¹. Le cadre utilisé est connu sous le nom de "modèle d'effet du traitement". Considérons pour cela le cas simple d'un jeune suivi en accompagnement renforcé au sein d'une mission locale. Cet accompagnement constitue le programme ou « traitement » à évaluer. **L'effet du traitement** se définit alors comme **l'écart entre ce qui arrive à ce jeune après le programme et ce qui lui serait arrivé en l'absence de ce programme**. Une conséquence immédiate de cette définition est que l'effet du traitement n'est jamais directement observable, puisque le deuxième terme de l'écart -- ce qui serait arrivé en l'absence de programme -- ne s'est pas réalisé. Un même jeune ne peut, à une date donnée, avoir bénéficié d'un accompagnement et ne pas en avoir bénéficié. C'est le problème fondamental de l'évaluation d'impact contrefactuelle : pour estimer l'effet d'un traitement, il ne suffit pas de suivre le jeune après ce traitement, il faut également reconstituer ce qu'aurait été **la trajectoire de ce même jeune aux mêmes dates dans une situation hypothétique** -- la situation contrefactuelle -- où il n'aurait pas bénéficié du traitement.

La reconstitution du contrefactuel repose, dans la très grande majorité des cas, sur la mobilisation d'un groupe témoin², constitué d'individus comparables en tout aux individus traités, et qui pourtant ne sont pas traités. Cette formulation met volontairement en avant la tension fondamentale inhérente à la reconstitution du contrefactuel : il faut trouver des individus qui, bien que comparables en tout aux individus traités, n'ont pas été traités. L'existence de tels individus ne va pas de soi : s'il existe un jeune exactement semblable au jeune suivi en accompagnement renforcé par une mission locale dans notre exemple, comment se fait-il qu'il ne soit pas lui-même en accompagnement renforcé ? Est-il vraiment semblable en tout, ou diffère-t-il en réalité par certaines caractéristiques inobservées qui l'ont conduit à ne pas pousser la porte de la mission locale, ou à refuser l'accompagnement renforcé qui lui a été proposé ? Et s'il diffère par ces caractéristiques inobservées -- moindre motivation par exemple --, dans quelle mesure les différences de trajectoires observées entre jeune traité et jeune témoin sont-elles dues à l'effet de l'accompagnement renforcé ou à l'effet de cette différence initiale de motivation ?

La qualité d'une évaluation d'impact repose donc en premier lieu sur l'identification de raisons, indépendantes des trajectoires futures des individus, qui font que certains bénéficient du traitement,

¹ Les principes et les outils de l'évaluation d'impact contrefactuelle sont présentés dans de nombreux documents. On pourra par exemple se reporter en particulier au guide écrit par J-PAL pour la DG Emploi, affaires sociales et inclusion (J-PAL Europe, 2011). Une présentation plus technique est proposée par Givord (2015).

² Ou groupe de contrôle.

et d'autres non. Cette raison est évidente lorsque groupe traité³ et groupe témoin ont été construits selon un **protocole expérimental** (aussi connu sous le nom d'évaluation par assignation aléatoire, ou évaluation randomisée). Ce protocole consiste à scinder en deux, par tirage au sort, une population initiale (par exemple, l'ensemble des jeunes suivis en mission locale dans une région donnée) et à traiter l'un des deux groupes seulement (par exemple, à offrir au groupe de traitement d'entrer en accompagnement renforcé, et à proposer un accompagnement plus léger au groupe témoin). A partir du moment où groupe test et groupe témoin sont de taille suffisante, le fait qu'ils soient tirés aléatoirement à partir de la même population fait qu'ils sont statistiquement semblables. Cela peut se vérifier en comparant les caractéristiques observées des deux groupes, et en constatant qu'elles ne diffèrent pas statistiquement. Et surtout cela reste vrai par construction, même si on ne peut le vérifier, d'autres caractéristiques inobservées susceptibles d'influencer les trajectoires des deux groupes, comme la motivation. Les différences de trajectoire alors observées après le traitement entre les deux groupes sont donc nécessairement dues au traitement.

Le modèle de l'évaluation randomisée peut être étendu à d'autres cas dits "**quasi-expérimentaux**" où un élément fortuit (comme une règle d'éligibilité comportant une part d'arbitraire) vient jouer le rôle du tirage au sort. Un bon exemple vient des critères discontinus qui conditionnent l'éligibilité à certains dispositifs. Par exemple, supposons que le dispositif d'accompagnement renforcé soit introduit le 1er mars 2015, mais que seuls les jeunes âgés de moins de 25 ans à cette date y soient éligibles. Cette règle introduit une discontinuité entre deux cohortes : celle née juste après le 1er mars 1990 -- éligible au traitement -- et celle née juste avant -- non éligible. Dans la mesure où être né le 28 février plutôt que le 1er mars 1990 relève de l'aléa naturel, la règle d'éligibilité au dispositif fait office de tirage au sort, assurant, par construction que les deux cohortes ainsi constituées sont comparables en tout, y compris dans des dimensions inobservées, mais ont, pour une raison bien comprise, accès à des modalités d'accompagnement différentes.

Pour certains dispositifs, il n'existe pas d'élément fortuit qui constitue ainsi deux groupes comparables. Il existe alors des techniques statistiques -- régression, modèles d'appariement (*matching*) -- qui permettent de reconstituer des groupes d'individus comparables dans toutes les dimensions observables (âge, sexe, milieu social, éducation, trajectoire antérieure sur le marché du travail) et qui pourtant diffèrent dans leur participation au dispositif qu'on veut évaluer. Par exemple, il est possible, à partir du vaste échantillon de l'ensemble des jeunes suivis en missions locales, de trouver plusieurs "jumeaux" du jeune suivi en accompagnement renforcé, c'est-à-dire des jeunes initialement semblables dans toutes les caractéristiques consignées dans les données administratives des missions locales, mais qui se trouvent ne pas avoir suivi d'accompagnement renforcé⁴. Ces "jumeaux" semblent donc pouvoir constituer un point de comparaison utile. Pour autant, il faut

³ Ou groupe test.

⁴ En pratique, il n'est pas nécessaire de trouver des jeunes qui aient exactement toutes les mêmes caractéristiques (appariement ou *matching* exact) ; il suffit de trouver des jeunes qui aient la même probabilité de suivre un accompagnement renforcé, au vu de leurs caractéristiques observées (appariement sur le score de propension).

noter qu'ils ne résolvent pas la contradiction que doit résoudre un bon groupe témoin : si rien ne permet de comprendre pourquoi deux "jumeaux" n'ont pas suivi le même type d'accompagnement, le risque subsiste que ce ne soient pas vraiment des jumeaux, et qu'ils diffèrent dans des caractéristiques inobservées qui expliquent leur traitement différent et dont les effets vont biaiser la comparaison (biais de sélection). Bien sûr, il ne s'agit que d'un risque. Il se pourrait bien après tout qu'à caractéristiques données, l'entrée en accompagnement renforcé tienne à des facteurs fortuits (passage au bon moment à la mission locale, au moment où de nouvelles places viennent de se libérer dans le portefeuille des conseillers) qui jouent le rôle du tirage au sort mais ne sont pas identifiés dans les données. L'évaluation par appariement souffre-t-elle de biais de sélection en pratique ? Plusieurs travaux de recherche ont analysé cette question empiriquement en tirant profit d'expériences randomisées dont les résultats fournissent le point de référence permettant de valider ou non les estimateurs non expérimentaux. Bléhaut et Rathelot (2014) étudient ainsi, sur données françaises, les effets d'un accompagnement renforcé sur de jeunes demandeurs d'emploi diplômés, fourni par des opérateurs privés. Seul un sous-groupe de jeunes tirés au sort s'était vu proposer cet accompagnement, selon un protocole randomisé, ce qui permet d'estimer l'effet du traitement à partir de la comparaison entre groupe test et groupe témoin. On conclut à un effet positif et significatif. Cependant, une partie des jeunes du groupe test n'avait pas suivi le programme, ce qui suggère une autre comparaison, au sein du groupe test, entre les jeunes effectivement entrés en accompagnement et les autres. Il s'avère possible de trouver, parmi les jeunes n'entrant pas en accompagnement, des "jumeaux" de ceux qui y sont eux entrés. Cela suggère de les utiliser comme contrefactuel. Mais ce contrefactuel s'avère extrêmement trompeur, puisqu'il conduit à conclure, à l'opposé de l'évaluation expérimentale, que l'accompagnement renforcé n'a pas eu d'effet sur l'accès à l'emploi, ou a même eu un effet négatif. Il semble bien, dans ce cas, que l'estimateur par appariement n'a pas résolu le problème de la sélection ; sans doute ces jeunes qui ne sont pas entrés en accompagnement renforcé sont-ils particuliers dans des dimensions inobservables, qui jouent un rôle important dans leur capacité à trouver un emploi hors accompagnement. Le résultat de Bléhaut et Rathelot (2014) se retrouve en fait dans la plupart des études de validation des évaluations par matching, dès lors qu'elles reposent sur la comparaison d'individus qui ont décidé à titre individuel d'entrer ou non dans un programme, ou qui ont été sélectionnés pour cela par des conseillers (programmes de formation, d'accompagnement, etc.) : il semble bien que la sélection individuelle dans ces programmes repose pour partie sur des dimensions non mesurées qui viennent biaiser l'analyse des effets (LaLonde, 1986 ; Friedlander et Robins, 1995 ; Heckman et al., 1998 ; Smith et Todd, 2005 ; Arceneaux, Gerber et Green, 2006).

1.2 Spécificités de l'IEJ au regard de l'évaluation d'impact contrefactuelle

Comment se présente l'IEJ face aux enjeux de l'évaluation d'impact ? On s'appuie ici sur la description de l'IEJ telle qu'elle ressort du rapport de synthèse de l'évaluation d'impact de l'IEJ au 31 décembre 2015 (Rapport national, 2015). L'ensemble des actions de l'IEJ ayant été programmées avant fin 2015, il a en effet été possible de mener une première évaluation de l'IEJ, qui porte sur les

stratégies retenues dans les Programmes opérationnels (national et régionaux) et leur déclinaison sur le terrain : mise en place des financements, sélection des dispositifs, montée en charge et premiers éléments de trajectoires des jeunes bénéficiaires. Une telle analyse nourrit la réflexion sur le type d'évaluation d'impact possible et opportun à l'horizon 2018, une fois que l'ensemble des actions programmées auront été mises en œuvre et qu'on disposera d'un recul suffisant pour mesurer les trajectoires des jeunes. On résume ici, parmi les résultats du Rapport de 2015, ceux qui sont directement pertinents pour l'évaluation de 2018 :

1. **Type de dispositifs.** L'IEJ est d'abord un outil de financement, qui a permis de maintenir des dispositifs existants, d'en étendre certains et, par le jeu d'appels à projets, de financer des dispositifs qui n'existaient pas jusque-là. Malgré cette diversité, l'examen des actions programmées montre la concentration des moyens financiers⁵ sur deux types d'actions : l'accompagnement renforcé, mis en œuvre surtout par Pôle emploi et les Missions locales et représentant près de la moitié de l'enveloppe totale programmée fin 2015 ; la formation (pré-qualifiante et qualifiante), qui représente elle 31% du total et passe principalement par les plans régionaux de formation. Les actions de repérage, présentes dans les Programmes opérationnels, sont moins investies en pratique ; les dispositifs nouveaux ou les actions transversales (par exemple, parcours intégrés) sont également rares. En résumé, si on se limite aux grandes masses, il apparaît que l'IEJ a eu pour orientations principales de :
 - s'inscrire dans le développement par le service public de l'emploi, depuis une dizaine d'années, d'un accompagnement renforcé, ciblé sur des publics éloignés de l'emploi et caractérisé par l'interaction régulière du demandeur d'emploi avec un conseiller attitré ;
 - marquer un effort particulier à destination des jeunes au sein des dispositifs de formation professionnelle, en englobant formation pré-qualifiante et qualifiante.
2. **Population bénéficiaire.** La population ciblée par l'IEJ est vaste et composite : on estime que les NEETs représentent 12,3% d'une classe d'âge, mais ce nombre recouvre des situations hétérogènes, allant du chômage frictionnel momentané de jeunes sortis récemment de formation à l'inactivité durable de jeunes décrocheurs du système scolaire. Les différents dispositifs financés par l'IEJ touchent rarement l'ensemble des NEETs, mais plus souvent l'un ou l'autre sous-groupe.
3. **Déploiement géographique.** L'IEJ a été déployée sur 16 territoires (13 régions, 3 départements) choisis selon un critère -- le taux de chômage des jeunes de 15-24 ans en 2012 -- qui s'avère comporter une part d'arbitraire, dans la mesure où le choix d'une autre année de référence aurait conduit à un ensemble de territoires partiellement différent. Par ailleurs, il faut noter que certains dispositifs financés par l'IEJ sont également présents, avec d'autres financements, dans des territoires non éligibles à l'IEJ. Par exemple, l'Accompagnement intensif des jeunes (AIJ) mis

⁵ Le total est de 730 millions d'euros, si on inclut la contrepartie nationale et les financements du Fond structurel européen associés à l'IEJ, dont 310,2 millions pour l'IEJ seule.

en œuvre par Pôle emploi est financé par l'IEJ dans les régions éligibles, mais est déployé également dans les régions non éligibles, avec un financement FSE.

4. **Montée en charge.** L'IEJ s'est inscrite dans une logique d'action rapide : toucher beaucoup de jeunes (365 000) en peu de temps ("effet masse"). Dans les faits, sur les 300 000 jeunes ciblés par le Programme opérationnel national, 108 000 (soit 36%) ont été touchés fin 2015. Si la tendance actuelle se prolonge, on peut penser que la cible sera atteinte en 2016. On ne dispose en revanche pas d'éléments à ce stade pour évaluer la montée en charge des Programmes opérationnels régionaux (62 000 jeunes ciblés).

D'un point de vue méthodologique, cette brève synthèse permet de souligner les difficultés spécifiques d'évaluation de l'IEJ, mais également d'esquisser des pistes qui sont explorées plus en détail dans les sections suivantes.

Tout d'abord, le modèle d'effet du traitement suppose en principe l'existence d'un traitement clair appliqué à une population aux contours bien définis. L'IEJ s'éloigne de cette référence tant par l'extrême diversité des dispositifs financés que par le caractère hétérogène de la population ciblée. Il reste techniquement possible de s'y ramener en considérant que l'IEJ est un "panier" (bundle) de dispositifs dont bénéficie potentiellement un ensemble de jeunes, par exemple, tous les jeunes de 16-25 ans des régions éligibles. La population cible est ici définie de façon volontairement large : en effet, tout jeune peut être "NEET" à un moment ou un autre de sa trajectoire, par exemple au moment de la transition entre études et premier emploi. La question qu'on pose est alors la suivante : que change en moyenne pour un jeune le fait d'être dans une région où a été déployé l'IEJ ? On voit que cette formulation respecte le cadre formel du traitement : la population est bien définie, sur un critère d'âge simple à vérifier, et le traitement également : être jeune dans une région IEJ. Mais on voit également que l'effet moyen ainsi défini mélange beaucoup de choses qu'on aimerait séparer ("unbundle") et attribuer aux sous-groupes de jeunes concernés : certains jeunes de 16-25 ans ne sont en pratique quasiment pas concernés par l'IEJ, d'autres sont potentiellement concernés mais ne participent pas aux dispositifs qui leur sont destinés (refus, manque d'information,...), d'autres enfin bénéficient de l'une ou l'autre composante. La littérature parle d'un "effet de l'intention de traiter" (intention-to-treat effect). Cet effet peut être utile pour avoir une vue d'ensemble. Mais ses enseignements pour les politiques publiques sont aussi limités. Supposons que l'effet obtenu soit faible. A quelle composante du programme l'attribuer : accompagnement renforcé ou formation professionnelle, ou autre ? S'agit-il d'un effet modéré sur beaucoup ou d'un fort impact sur quelques-uns ? Il sera difficile de répondre à ces questions sans analyses complémentaires.

La seconde difficulté d'évaluation de l'IEJ est l'identification de variations permettant de reconstituer un contrefactuel. Aucune composante de l'IEJ n'a été allouée individuellement aux jeunes sur la base d'un tirage au sort qui permettrait de constituer un groupe témoin. En revanche, le déploiement sur une partie du territoire permet d'envisager que les territoires non concernés puissent servir de territoires témoins. Néanmoins, les territoires concernés par l'IEJ n'ont pas été choisis au hasard, mais en raison de leur taux de chômage élevé, en 2012, chez les 15-24 ans. Il convient donc de

vérifier dans quelle mesure territoires IEJ et non IEJ sont comparables -- ce n'est pas exclu dans la mesure où il y a une part d'arbitraire (donc d'aléa) dans la date de mesure du taux de chômage, et dans le seuil d'éligibilité retenu. Reste alors à vérifier que le petit nombre de régions sur lesquelles reposerait la comparaison test - témoin fournit une précision statistique suffisante. Par ailleurs, la question des sources de variation peut s'étudier séparément pour les différentes composantes de l'IEJ : existe-t-il des particularités dans le déploiement des différents dispositifs (AIJ, GJ, Ecoles de la deuxième chance, etc.) qui permettent l'évaluation de leur impact, par la comparaison entre territoires (ou entre sous-populations) bénéficiant ou non du dispositif pour des raisons connues et qui présentent une part d'aléatoire ou d'arbitraire, révélant ainsi l'effet du dispositif lui-même ?

2. Deux approches pour évaluer l'IEJ

Cette section examine successivement deux approches possibles pour évaluer l'IEJ, tant du point de vue des questions que ces approches permettent de poser que du point de vue de leur faisabilité.

2.1 Evaluer la plus-value globale de l'IEJ

Evaluer globalement la plus-value de l'IEJ en comparant la situation des jeunes (ou de certains sous-groupes de jeunes) dans les territoires IEJ et les territoires non IEJ semble un point de départ naturel : qu'est-ce que change en moyenne pour un jeune d'être dans un territoire éligible aux financements IEJ plutôt que dans un territoire non éligible ? On examine ici les questions auxquelles une telle évaluation permet de répondre, et sa faisabilité.

Opportunité d'une évaluation globale

Une évaluation globale de l'IEJ présente l'avantage de sembler répondre à l'objectif de l'évaluation de l'IEJ "en tant que telle" : *in fine*, l'IEJ a-t-elle en moyenne amélioré la situation des jeunes face à l'emploi, conformément à l'objectif politique initial ?

Une telle évaluation combine, par construction, l'ensemble des effets de l'IEJ, c'est-à-dire non seulement l'effet des dispositifs de l'IEJ pris un à un, mais aussi les éventuels effets de complémentarité qui peuvent exister entre dispositifs, ou *a contrario* les effets de cannibalisation ; non seulement l'effet sur les bénéficiaires directs, mais aussi d'éventuels effets induits sur les non bénéficiaires du même territoire (externalité positive, par exemple si la mise en place d'un nouveau dispositif financé par l'IEJ a permis de réduire l'engorgement dans un dispositif préexistant ; externalité négative via les effets d'équilibre sur le marché local du travail si les jeunes accompagnés par un dispositif IEJ évincent d'autres jeunes sur les emplois disponibles localement).⁶ L'évaluation est d'autant plus globale que la population analysée s'approche de la population totale en âge de travailler : ainsi, si on limite l'analyse aux jeunes de 26 ans au plus, on ne pourra pas voir si la politique a eu un effet d'éviction sur les jeunes juste plus âgés. Bien sûr cependant, une perspective

⁶ Pour une mesure de ces effets sur des jeunes diplômés, voir Crépon et al. (2013).

trop globale (qui voudrait mesurer par exemple l'effet de l'IEJ sur le taux d'emploi tous âges et niveaux de qualification confondus, pour englober toute externalité du dispositif au-delà de sa cible) se heurterait à des problèmes de détection statistique (voir *infra*). Mais il est toujours possible de désagréger l'analyse par sous-population, et d'examiner groupe d'âge par groupe d'âge, niveau de qualification par niveau de qualification, si un effet est détectable.

La contrepartie de cette perspective englobante est qu'elle ne permet pas de décomposer les effets obtenus afin de les attribuer à telle ou telle composante de l'IEJ (ou à leur interaction). Du point de vue de la décision publique, l'évaluation globale est donc utile pour décider de la reconduction ou non de la politique. Mais elle ne permet pas d'identifier les points forts et les points faibles parmi ses composantes. Au vu du caractère hétérogène des mesures financées par l'IEJ, c'est une limite significative.

L'évaluation globale de l'IEJ doit comporter une analyse coûts - bénéfices. L'évaluation ne peut clairement pas être binaire (l'IEJ marche-t-elle ou non ?) : elle passe par une quantification des effets au regard des investissements consentis. Il est important de prendre conscience que l'évaluation des coûts requiert un travail d'évaluation propre. Le financement total de 730 millions d'euros programmé au titre de l'IEJ en France ne constitue en effet pas une bonne mesure des ressources *nettes* que l'IEJ a conduit à investir pour l'emploi des jeunes NEETs dans les régions IEJ. En effet, les observations de terrain interrogent sur l'existence de forts effets de substitution entre sources de financement. Ainsi, l'IEJ a permis à certaines régions de financer des composantes préexistantes de leur plan régional de formation. Il est donc probable que l'investissement public *additionnel* dans les régions concernées est sensiblement inférieur au financement IEJ. Mais il est difficile de dire de combien. Pour reprendre l'exemple des plans de formation régionaux, on peut envisager deux cas polaires : celui où l'IEJ n'aurait rien changé et aurait financé des investissements que la région aurait de toute façon consentis (l'effet de l'IEJ étant alors d'améliorer l'équilibre budgétaire de la région concernée) ; celui au contraire où, en l'absence d'IEJ, la région aurait choisi de couper les lignes concernées, si bien que les ressources de l'IEJ ont bien permis d'augmenter le niveau d'investissement dans la formation. En réalité, mesurer l'effet net de l'IEJ sur les dépenses d'emploi relève de l'évaluation contrefactuelle : il faut évaluer la dépense qu'auraient consentie les différentes autorités de gestion si elles n'avaient pas bénéficié de l'IEJ.

Pour résumer, une évaluation globale doit permettre, dans une même approche, d'identifier l'effet causal de l'IEJ sur deux types de variables : des variables de moyens / d'investissement (quel effet sur la dépense par jeune ?) et des variables de résultat final / de retour sur investissement (quel effet sur le taux d'emploi ?). Cette perspective globale permet une analyse coûts-bénéfices du dispositif dans son ensemble.

Faisabilité de l'approche globale

Est-il possible dans le cas de l'IEJ de mener une telle approche globale de façon crédible, et quelle précision (statistique) peut-on en attendre ?

Commençons par la crédibilité. On a vu que le choix opéré par la France de concentrer le financement de l'IEJ sur une partie du territoire ouvre la voie à une comparaison entre territoires IEJ et non IEJ. Néanmoins, le choix des territoires ne s'est clairement pas fait au hasard : les régions et départements bénéficiaires ont été choisis en raison d'un taux élevé de chômage des jeunes, mesuré en 2012. La comparaison naïve des taux de chômage en coupe en 2018 (ou de tout autre mesure reflétant la situation des jeunes sur le marché du travail) a toutes chances de biaiser vers le bas l'effet estimé de l'IEJ, dans la mesure où le chômage présente une certaine persistance. Une solution naturelle consisterait à comparer les taux de chômage des jeunes (ou d'autres mesures de leur situation sur le marché du travail, voir ci-dessous) non pas en coupe, mais en évolution (approche en "double différence"). L'hypothèse sous-jacente est que les territoires retenus pour l'IEJ, bien qu'ils se caractérisent par des niveaux de chômage élevés en 2012, auraient connu *en l'absence d'IEJ* entre 2015 et 2018 une évolution du chômage des jeunes parallèle aux territoires non IEJ. Cette hypothèse est connue sous le nom d'hypothèse des tendances parallèles. Elle est cruciale pour l'évaluation mais n'est pas directement testable. Un test indirect consiste néanmoins à regarder si, avant 2014, l'hypothèse des tendances parallèles était vérifiée. C'est ce que fait la figure 1 à partir des données de taux de chômage régionaux harmonisés publiées par Eurostat. Le graphique du milieu correspond à la sélection effective des régions IEJ⁷, qui se base sur les taux de chômage mesurés en 2012. Concentrons-nous pour commencer sur ce graphique. La ligne en trait plein mesure la différence entre le taux de chômage moyen dans les régions IEJ et le taux de chômage moyen des autres régions, pour les 15-24 ans et pour différentes années. En 2012, cet écart est d'environ 6 points de pourcentage (sur un taux de chômage de l'ordre de 25%). Ce n'est pas étonnant, dans la mesure où les régions IEJ ont précisément été choisies selon ce critère. Mais le graphique permet également de constater que l'année 2012 n'était pas exceptionnelle : dès le début des années 2000, les futures régions IEJ souffraient d'un chômage plus élevé chez les 15-24 ans (une différence d'un peu moins de 5 points de pourcentage). Cette différence est restée positive entre 2000 et 2012, même si elle a un peu fluctué (entre 1 et 6 points d'écart selon les années), et elle persiste ensuite. En d'autres termes, les régions IEJ et non IEJ ont suivi, sur la période qui précède le déploiement du dispositif (avant 2015) des tendances à peu près parallèles : l'hypothèse de tendances parallèles n'est donc pas rejetée sur le long terme.

Les graphiques supérieur et inférieur donnent une deuxième raison pour justifier la comparaison entre régions non IEJ et régions IEJ. Ils montrent ce qui se serait passé si la période de référence retenue pour décider de l'éligibilité à l'IEJ avait été différente, c'est-à-dire si on avait choisi les régions IEJ en fonction du taux chômage des jeunes en 2010 ou 2014, plutôt que 2012. Ainsi, le graphique du haut trace l'écart de chômage observé entre les régions qui avaient le plus fort taux de chômage des jeunes *en 2010* et les autres. L'écart reste positif, mais il ne suit pas exactement la même évolution que sur le graphique du milieu. De même pour une période de référence fixée à

⁷ Pour pouvoir utiliser des données régionales sur plusieurs années, on inclut dans les "régions IEJ" la région PACA, mais on inclut parmi les régions "non IEJ" l'Ile-de-France et la région Midi-Pyrénées.

2014. Ainsi, même si le critère retenu vise les régions qui ont "le plus besoin" du dispositif, la sélection obtenue comporte une part d'arbitraire, directement lié au fait de mesurer ce besoin par la situation *en 2012*⁸. Si 2010 avait été l'année de référence, la Bourgogne et la Lorraine auraient été des régions IEJ, mais pas l'Auvergne ni le Centre. Si les taux de chômage de 2014 avaient été disponibles, la liste aurait été encore plus différente puisqu'elle aurait comporté l'Alsace, le Limousin, le Poitou-Charentes, mais pas l'Aquitaine ni le Centre, ni la Champagne-Ardennes.⁹

En résumé, les régions IEJ sont spécifiques, car ce sont des régions où le taux de chômage des jeunes est élevé. Mais les écarts entre régions IEJ et non IEJ semblent relativement stables avant 2014, et pourraient être corrigés par une approche en double différence. Et il existe d'autres régions au taux de chômage des jeunes élevé (moins élevé en 2012, par définition, mais parfois plus élevé en 2010 ou en 2014) qui peuvent servir de groupe de comparaison. On peut donc considérer qu'il y a là une première source de variation quasi-expérimentale qui peut être exploitée par l'évaluation d'impact.

Il faut cependant être extrêmement prudent dans la façon d'utiliser cette source de variation : l'effet estimé risque d'être sensible aux choix d'estimation effectués. Pour s'en convaincre, on peut regarder le graphique du milieu. Imaginons, à titre d'exercice de pensée, qu'on veuille évaluer à partir de ce graphique les effets d'une politique qui serait intervenue en 2012. La conclusion n'est pas immédiate. On pourrait par exemple dire que la politique a été un succès dans la mesure où les écarts de taux de chômage augmentent fortement entre 2009 et 2012, puis se réduisent fortement en 2013 et 2014. Mais on pourrait aussi faire valoir que l'écart en 2013-14 est similaire à l'écart observé au début des années 2000, et conclure que la politique n'a eu aucun effet. Techniquement, la première évaluation qui prend en compte la divergence des taux de chômage entre 2009 et 2012 revient à lever l'hypothèse de tendances parallèles, et à corriger des trends préexistants ("random-growth model"). Ce faisant, elle risque néanmoins de sur-corriger, puisque la réduction de l'écart après 2013, période où l'IEJ n'était pas en place, relève bien sûr plus d'un effet de retour à la moyenne que d'une quelconque politique. Le *random-growth model* conduirait donc à surestimer fortement la réduction du chômage imputable à l'IEJ, là où une estimation en coupe conduirait à le sous-estimer, l'estimation en double différence se situant entre deux. Clairement, cet exercice suggère que le choix du modèle retenu n'est ni neutre ni indiscutable, et que le choix d'une spécification devrait être soigneusement justifié par l'évaluateur, et accompagné de nombreux tests de robustesse.

Par ailleurs, les graphiques de la figure 1 font figurer en pointillés des intervalles de confiance à 95%, c'est-à-dire la fourchette à l'intérieur de laquelle on peut estimer, avec moins de 5% de chances de se

⁸ En lui-même bien entendu, le taux de chômage serait un objectif insuffisant, dans la mesure où sa réduction pourrait provenir d'une baisse du taux d'activité des jeunes. Il est donc nécessaire, à tout le moins, de le combiner avec d'autres indicateurs (taux d'activité, taux de participation à des formations, par exemple).

⁹ En 2010, les neuf régions dont le taux de chômage des jeunes étaient le plus élevées étaient : Aquitaine, Bourgogne, Champagne-Ardenne, Haute-Normandie, Languedoc-Roussillon, Lorraine, Nord - Pas-de-Calais, Picardie, Provence-Alpes-Côte d'Azur ; en 2014, il s'agissait de Alsace, Haute-Normandie, Languedoc-Roussillon, Limousin, Lorraine, Nord - Pas-de-Calais, Picardie, Poitou-Charentes, Provence-Alpes-Côte d'Azur.

tromper, que se situe la différence véritable entre régions IEJ et non IEJ. Cette fourchette est très étendue (plus ou moins cinq points de pourcentage), ce qui traduit l'importance des fluctuations conjoncturelles de taux de chômage d'une région à une autre, dues à des chocs ou à de l'erreur de mesure. Supposons dans ce contexte qu'on estime ainsi un modèle corrigeant les différences de trends régionaux passées (random-growth model¹⁰) sur la période 1999-2014, et que l'IEJ entre en vigueur entre 2012 et 2013. On peut calculer que l'effet de l'IEJ est estimé avec une marge d'erreur de l'ordre de plus ou moins deux points de pourcentage. L'effet minimum détectable est alors de 2,6 points environ, ce qui signifie que, pour avoir des chances raisonnables de détecter un effet de l'IEJ sur le chômage des jeunes (puissance statistique de 80% pour des tests à 5%), il faudrait que l'IEJ ait fait baisser ce taux d'environ 10%, ce qui correspond à 16 000 emplois supplémentaires. Peut-on attendre de tels effets de l'IEJ ? Il faut rapporter ces emplois supplémentaires aux quelque 300 000 jeunes ciblés par l'IEJ : il faudrait que plus de 5% d'entre eux soient en emploi grâce à l'IEJ, à la date retenue pour l'évaluation. Il s'agit là d'effets élevés mais pas inimaginables si on pense par exemple aux effets estimés des programmes d'accompagnement renforcé (Behaghel et al., 2013).

¹⁰ La variable expliquée est le taux de chômage régional, mais on introduit des indicatrices de région, de dates, et des trends (linéaires) spécifiques à chaque région. L'effet de l'IEJ est le coefficient sur la variable indicatrice interagie 'région IEJ x année>2012'. Les écarts-types sont robustes à des effets de cluster au niveau de la région. Une telle spécification est clairement très imparfaite, en raison du faible nombre de régions. Elle permet cependant de se faire une première idée de la précision statistique qu'on peut espérer atteindre.

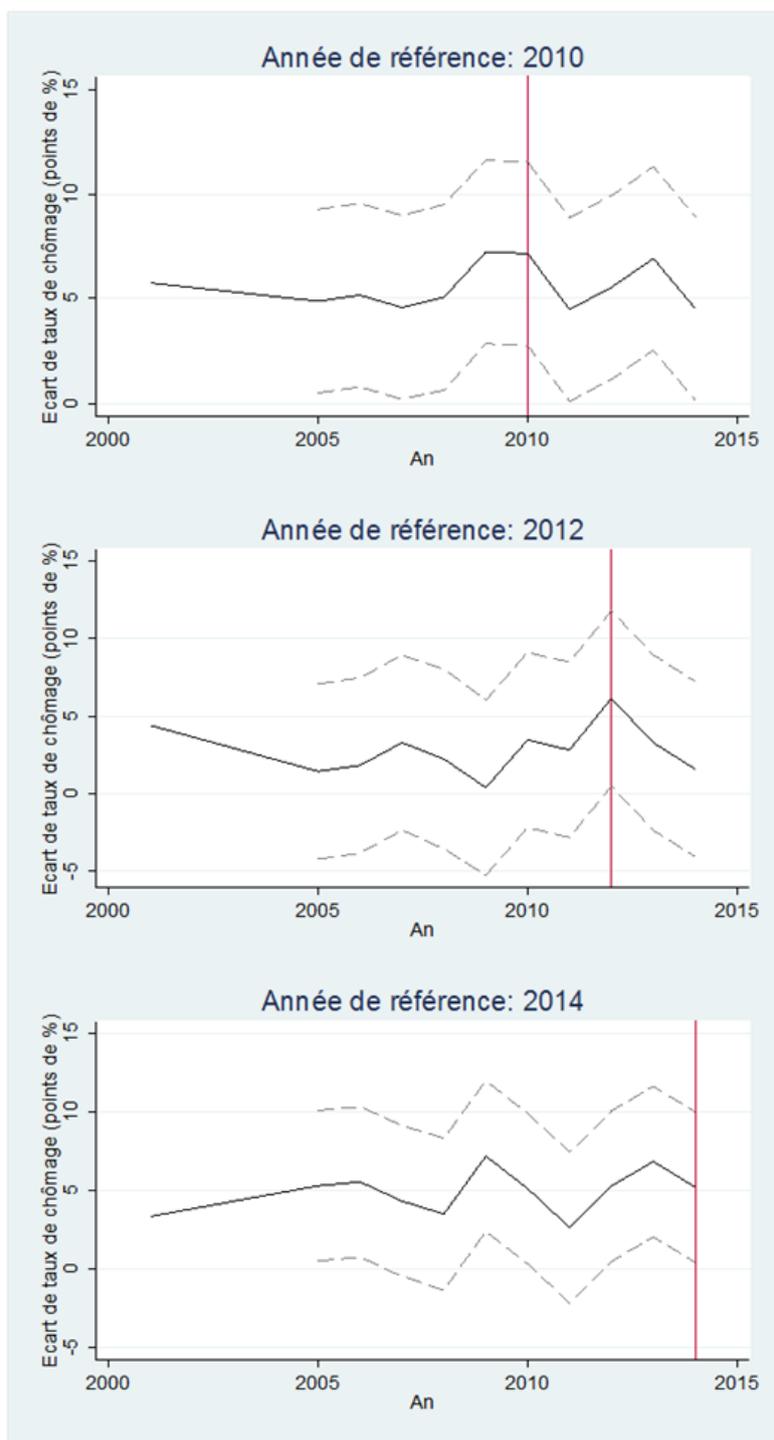


Figure 1 : Différence de taux de chômage entre régions à taux de chômage élevé et les régions à taux de chômage faible (pour trois années de référence : 2010, 2012 et 2014). Les régions à taux de chômage élevées en 2012 sont celles retenues pour l'IEJ. Les graphiques du haut et du bas illustrent les conséquences du choix d'une autre année de référence. Le chiffre porté pour 2001 correspond à la moyenne en 1999-2002, pour limiter le bruit lié aux observations manquantes. Les lignes en pointillés correspondent aux intervalles de confiance à 95%.

Sources : Taux de chômage régionaux des 15-24 ans, données Eurostat.

Amélioration de l'approche globale : désagrégation par groupes d'âge

Une autre source de variation de nature "quasi-expérimentale" envisageable est liée au critère d'âge fixé pour l'éligibilité à l'IEJ, qui est réservée aux jeunes de moins de 26 ans. Cette règle introduit une discontinuité entre cohortes au sein des régions IEJ : les cohortes qui ont plus de 26 ans au moment où l'IEJ commence à se déployer (en 2015) n'ont pas accès à l'IEJ, tandis que celle qui ont moins de 26 ans à la même date sont éligibles. La discontinuité n'est sans doute pas très nette, dans la mesure où l'IEJ s'est mise en place progressivement, si bien les cohortes approchant la limite d'âge au démarrage de l'IEJ ont pu n'y avoir que partiellement accès. Il n'en reste pas moins que les jeunes de plus de 26 ans des régions IEJ satisfont *a priori* les conditions permettant de constituer un bon groupe de contrôle : proche en tout des jeunes éligibles à l'IEJ (à l'effet de l'âge près), mais n'accédant pas à l'IEJ, pour une raison bien comprise, leur âge. Bien sûr, il ne suffit pas de comparer, en 2018, la situation des jeunes qui avaient moins de 26 ans en 2015-2016 avec celle des jeunes qui avaient plus de 26 ans à ces dates. Ces situations ne vont pas différer seulement en raison de l'accès à l'IEJ, mais aussi en raison de la différence d'âge. Il est cependant à nouveau possible de suivre une approche en double différence : on examine, à différentes dates, l'écart de taux d'emploi entre cohortes de jeunes adjacentes (par exemple, aux âges de 27-28 ans et 29-30 ans). Si l'écart se réduit de façon spécifique lorsque la cohorte âgée de 27-28 ans a eu accès à l'IEJ plus jeune, tandis que la cohorte âgée de 29-30 ans n'y a pas eu accès, alors on peut attribuer ce résultat favorable pour les plus jeunes à l'effet de l'IEJ.¹¹

La comparaison entre groupes d'âge présente également un avantage en termes de précision statistique : elle peut s'effectuer au sein des régions, ce qui neutralise les variations conjoncturelles propres aux régions, qui touchent également les groupes d'âge adjacents. Par ailleurs, les régions non IEJ permettent d'effectuer un "test placebo". L'effet attribué à l'IEJ ne doit pas apparaître dans les régions non IEJ : s'il apparaît, c'est que d'autres phénomènes sont à l'œuvre (effet différencié du cycle économique selon les âges, par exemple), et l'évaluation devient plus difficile. Mais si l'effet n'apparaît que dans les régions IEJ, alors on peut de façon plus sûre l'attribuer à l'IEJ.

Il reste cependant à noter une limite : en se focalisant sur la comparaison autour de 26 ans, l'analyse ne permet plus tout à fait d'effectuer une évaluation globale. S'il existe des effets de substitution entre groupes d'âge, l'effet relatif observé peut venir en partie d'un effet d'éviction, le groupe plus âgé, moins formé ou moins accompagné, "perdant" des emplois au profit du groupe plus jeune.

L'évaluation "globale" de l'IEJ peut donc combiner des comparaisons entre jeunes au-dessus et en dessous de limite d'âge, et entre régions IEJ et non IEJ. On peut s'interroger sur la possibilité d'exploiter également un troisième critère d'éligibilité à l'IEJ : la situation de NEET. Les "non NEETS"

¹¹ Une alternative serait d'effectuer une analyse de la discontinuité juste autour de 26 ans (regression discontinuity design, ou RDD), mais cette piste n'est pas poursuivie ici dans la mesure où elle suppose des données avec une mesure précise de l'âge (au jour ou mois près) et une politique dont le déploiement est quasi instantané. L'approche en double différence sur des cohortes adjacentes n'est cependant pas très éloignée en pratique d'une double différence, avec les tests supplémentaires de spécification que rend possible la dimension longitudinale.

peuvent-ils constituer un groupe de contrôle satisfaisant ? La réponse semble négative, pour une raison principale : le fait d'être en emploi, en formation ou en cours d'études, ou pas, constitue une différence importante, sans doute tirée par un ensemble de facteurs dont certains sont inobservables, et qui viendront biaiser toute comparaison qu'on pourra vouloir faire ensuite. On retrouve la mise en garde faite en section 1 : lorsque ce sont des comportements individuels qui déterminent (au moins pour partie) l'accès à un dispositif, alors la comparaison directe entre bénéficiaires et non bénéficiaires risque fort d'être biaisée.

2.2 Evaluer l'IEJ par les dispositifs qu'elle finance

Une approche complémentaire de la précédente consiste à évaluer l'IEJ en évaluant des dispositifs qu'elle finance. L'objectif n'est plus d'analyser si l'IEJ a fait globalement et en moyenne une différence (les dispositifs financés par l'IEJ auraient pu être financés par d'autres canaux), mais si le financement de l'IEJ a été utilisé sur des programmes efficaces. C'est une autre façon de répondre à la question de l'impact de l'IEJ : l'IEJ est efficace si elle finance des dispositifs efficaces.

Un premier intérêt de cette façon d'évaluer l'IEJ est de permettre une réponse plus fine : peut-être l'IEJ a-t-elle été efficace dans certaines composantes et pas dans d'autres. Evaluer les composantes de l'IEJ séparément permet de comparer leurs effets. Un second avantage est d'ouvrir le champ des sources de comparaison possibles. En effet, une part importante des dispositifs financés par l'IEJ existent sur une période plus longue. Il n'est du coup plus nécessaire de faire reposer l'évaluation sur la comparaison entre territoires IEJ et non IEJ, ou entre jeunes de plus ou moins de 26 ans au moment du déploiement de l'IEJ. Cette comparaison serait pour certains des dispositifs financés par l'IEJ inopérante : en particulier, on sait que la différence de traitement qui aurait pu être induite par le financement IEJ a été en pratique lissée entre territoires par des opérateurs nationaux qui ont utilisé l'IEJ comme une source de financement parmi d'autres. Par exemple, la formation des emplois d'avenir a été également financée dans les territoires non IEJ, mais par le FPSPP ; l'AIJ a été déployée selon les mêmes principes sur l'ensemble du territoire par Pôle emploi, alternativement sur financement IEJ ou FSE.

Quels dispositifs évaluer ?

La seconde approche d'évaluation de l'IEJ consiste donc à choisir, parmi les multiples dispositifs financés par l'IEJ, certains dispositifs à évaluer individuellement. Le choix des dispositifs à évaluer devrait être fait selon trois critères : (i) les gains attendus de l'évaluation en termes d'aide à la décision publique, qui seront d'autant plus grands que les effets du dispositif choisi sont mal connus et que ce dispositif s'applique ou peut s'appliquer à grande échelle ; (ii) dans une perspective d'évaluation de l'IEJ, la représentativité du dispositif ou son adéquation aux objectifs de la politique ; (iii) la faisabilité d'une évaluation de qualité.

Commençons par les deux premiers critères, et examinons dans quelle mesure les dispositifs dont la part dans le financement IEJ est significative posent des questions d'évaluation dont les réponses sont susceptibles d'avoir un effet important sur les décisions publiques. On a vu ci-dessus que l'IEJ comporte deux accents principaux, sur l'accompagnement et sur la formation qualifiante et pré-qualifiante.

L'accent porté sur l'accompagnement renforcé est cohérent avec les résultats globalement positifs des évaluations de programmes d'accompagnement renforcé (Behaghel et al., 2013, pour la France ; Card et al., 2010 et 2015, pour des revues internationales). Pour autant, les résultats de l'accompagnement renforcé paraissent plus mitigés lorsqu'on considère spécifiquement les jeunes non qualifiés (Zamora, 2016). L'évaluation du programme expérimental du Revenu contractualisé d'autonomie suggère des difficultés spécifiques pour mobiliser ces jeunes dans la recherche d'emploi (Aeberhardt et al., 2014). D'importantes questions restent donc ouvertes : l'accompagnement renforcé bénéficie-t-il aux jeunes non qualifiés, et à quelles conditions ? L'IEJ finance différentes déclinaisons de l'accompagnement renforcé sur des groupes de jeunes différents : Accompagnement intensif des jeunes (AIJ) de Pôle emploi, accompagnement renforcé mis en œuvre spécifiquement par les Missions locales au titre de l'IEJ et Garantie jeunes (GJ). Il serait donc éclairant de comparer les effets de ces différents programmes sur les trajectoires d'emploi des jeunes, et d'essayer de comprendre les mécanismes sous-jacents par l'analyse des réponses comportementales des jeunes (effort de recherche, etc.).

L'autre accent de la programmation de l'IEJ -- qui concerne la formation pré-qualifiante et qualifiante -- pose des questions légèrement différentes. Les revues de littératures existantes, comme celle de Card et al. (2011) montrent des effets moyens faibles des programmes de formation professionnelle à destination des demandeurs d'emploi, voire des effets négatifs à court et moyen terme via l'effet de *lock-in* (la personne formée n'étant plus disponible pour prendre un emploi). Néanmoins, les difficultés croissantes rencontrées par les travailleurs non qualifiés sur le marché du travail suggèrent qu'il pourrait être extrêmement coûteux de laisser s'engager durablement dans la vie active des jeunes sans qualification. Certains parcours mis en place par les régions pour articuler formation pré-qualifiante et qualifiante peuvent constituer une réponse. Se sont par ailleurs développées des formules de formation innovantes qui repartent des savoirs de base et engagent un suivi intensif du jeune (Ecole de la deuxième chance, Défense deuxième chance). Ces dispositifs n'ont pas été évalués en France, alors que la littérature internationale montre que de telles interventions intensives permettent d'obtenir des effets forts, comme les *Job Corps* aux Etats-Unis, même si l'analyse coûts-bénéfices ne va pas de soi.

Par ailleurs, au-delà de "l'effet masse" (obtenu par la mobilisation de dispositifs de formation et d'accompagnement à forte capacité), l'IEJ s'est donné pour objectif de promouvoir des actions innovantes, ce qui s'est traduit en particulier en France par le jeu d'appels à projet visant des acteurs ou des alliances d'acteurs nouveaux. Enfin, même s'il ne semble pas que cet objectif ait été suivi de changements majeurs sur le terrain, en raison de difficultés organisationnelles et de contraintes de gestion, il faut noter l'accent mis par les différents programmes sur l'articulation des différents

dispositifs, afin de mettre en place des parcours cohérents. Cela pose la question des effets d'interaction entre dispositifs, question rarement abordées dans l'évaluation d'impact contrefactuelle des politiques de l'emploi. Dans les faits, le rapport au 31 décembre 2015 montre que ces dispositifs sont assez rares, et que l'innovation y est souvent de nature incrémentale.

Faisabilité des évaluations

Un certain nombre de dispositifs financés par l'IEJ se prête à des évaluations "quasi-expérimentales" et certaines d'entre elles, comme celles de la Garantie jeunes, sont déjà en cours. Les évaluations peuvent en effet exploiter le même type d'approches, combinant des comparaisons entre régions (pas nécessairement IEJ vs. non IEJ) et entre cohortes.

Parmi les évaluations possibles, considérons à titre d'exemple le cas des Ecoles de la deuxième chance (E2C). La comparaison, à une date donnée, entre jeunes entrés en E2C et autres jeunes risque de souffrir de biais de sélection persistants, dans la mesure où le processus d'entrée en E2C fait particulièrement appel à la motivation (inobservable) du jeune. En revanche, des comparaisons entre territoires peuvent être plus pertinentes, en tirant parti du déploiement progressif des E2C. En particulier, l'essor des Ecoles de la deuxième chance entre 2009 et 2012, qui passent de 49 à 107 sites et de 5 700 à 13 000 jeunes accueillis, induit des évolutions contrastées d'un territoire à l'autre.

Les dispositifs suscités localement par appels à projet sont en revanche plus difficiles à évaluer. Dans la mesure où ces dispositifs sont locaux, il n'est pas possible de s'appuyer sur des comparaisons entre territoires. L'évaluation expérimentale selon un protocole randomisé semble ici incontournable. Un des objectifs dans la préparation de cette note méthodologique était de repérer des dispositifs locaux se prêtant à de telles évaluations. Deux conditions sont en effet nécessaires : (i) la possibilité de randomiser, qui se justifie en général soit comme moyen de rationnement dans un contexte d'excès de demande pour le dispositif, soit comme moyen de lever une interrogation sur les meilleures modalités de mise en œuvre du dispositif ; (ii) des échantillons suffisants (plusieurs centaines de jeunes), ce qui suppose que le porteur du dispositif ait déjà atteint une certaine taille. Cependant, cette recherche est restée infructueuse, malgré l'interrogation systématique des acteurs de terrain et même si certains dispositifs s'approchaient des conditions requises. C'est le cas d'un parcours intégré d'accompagnement, d'insertion par l'activité économique et de formation en Haute-Normandie, mis en place par le GRAIN et AGEFOS Haute-Normandie, en partenariat avec les missions locales. L'impact attendu de ce dispositif se pense en plus-value par rapport à une double référence: l'insertion par l'activité économique seule, mise en œuvre par les associations d'insertion membres du GRAIN, dont la composante d'accompagnement reste modeste et pas spécifiquement adaptée aux jeunes et dont l'articulation avec les autres acteurs (ML, organismes de formation) ; l'accompagnement mené par les ML, moins intensif et mobilisant moins facilement l'immersion en milieu professionnel. Le programme est mis en place par un groupement d'associations, qui disposent d'un nombre limité de places et font face à une demande importante. Le rationnement qui en résulte pourrait être régulé par assignation aléatoire : pour des groupes de 12, les listes remontées par les partenaires sont montées jusqu'à 67 jeunes. Néanmoins, la cible totale est de 144 jeunes seulement. Une évaluation randomisée aurait du sens si on pouvait attendre des effets

massifs, détectables même avec de petits échantillons. Mais il s'agit là d'un programme dont la valeur ajoutée reste incrémentale et dont les effets ne sont sans doute pas détectables avec de petits échantillons.

3. Données et mise en œuvre

Cette dernière section présente des recommandations pour la mise en œuvre de l'évaluation de l'IEJ en 2018 et présente les données qui pourraient être mobilisées ; pour faciliter la lecture, ces recommandations sont isolées en tant que telles dans le texte.

La discussion précédente conduit à recommander de combiner deux approches pour évaluer l'IEJ : une approche globale, qui prenne appui sur les conditions spécifiques de déploiement de l'IEJ (régions et groupes d'âge), et une approche par dispositifs, pris en tant que tels sans se restreindre à la période et aux régions pour lesquelles l'IEJ les a financés.

Recommandation 1 : *Combiner l'évaluation globale de l'IEJ avec l'évaluation d'un ou deux plusieurs dispositifs spécifiques financés par l'IEJ.*

Recommandation 2 : *Evaluer des dispositifs financés par l'IEJ pour eux-mêmes, et non pour savoir si l'IEJ a influé sur leur mise en œuvre.*

3.1 Données mobilisables

Quatre types de données peuvent être mobilisés :

Données de suivi des jeunes pris en charge dans le cadre de l'IEJ

Un premier type de données est collecté à travers l'outil « Ma démarche FSE », pour enregistrer, caractériser et suivre les bénéficiaires de l'IEJ. Ces données sont indispensables pour décrire le public touché par l'IEJ, les différentes composantes du programme, et pour mesurer des résultats. Elles ne permettent cependant pas d'évaluer des impacts, dans la mesure où l'absence de jeunes "témoins", non éligibles au dispositif, empêche de reconstituer un contrefactuel.

Données d'enquête ou de recensement sur la situation des jeunes

L'analyse d'impact globale de l'IEJ proposée en section 2.1 repose sur des comparaisons de statut d'emploi entre cohortes de naissance et entre régions. Une information individuelle sur le passage par l'IEJ n'est pas nécessaire, dans la mesure où on vise ici un effet global, sur les bénéficiaires et les non bénéficiaires. Un véritable panel n'est pas non plus nécessaire, dans la mesure où on peut constituer un « quasi-panel », c'est-à-dire avoir, à différentes dates, des mesures représentatives de la situation des différents groupes (cohorte de naissance x région). En revanche, on a besoin, pour appliquer une approche en doubles différences, de disposer de données répétées ce qui, sur la période d'analyse limitée dont on disposera après la mise en place de l'IEJ, suppose des enquêtes de fréquence élevées (annuelle ou pluri-annuelle).

Les deux principales bases qu'on peut alors mobiliser sont les Enquêtes emploi (EE), enquête de référence sur l'emploi en France, et les enquêtes annuelles de recensement. Les Enquêtes emploi sont incontournables par la richesse des informations disponibles sur l'activité des jeunes (halo autour du chômage, par exemple) et sur leur formation. L'autre grand avantage est que ces données

sont collectées tous les trimestres et rapidement disponibles. Il devrait être ainsi possible d'analyser des effets à un ou deux ans de l'IEJ, en s'appuyant en 2018 sur les données collectées en 2017. Le principal inconvénient est la taille des échantillons, qui conduit à une mesure bruitée dès lors qu'on désagrège les résultats, comme on l'a illustré section 2.1.

Les Enquêtes annuelles de recensement (EAR) sont issues du recensement en continu de la population. Leur principal avantage est la taille des échantillons, qui permet des analyses fines par groupe d'âges (voir par exemple l'exploitation de la discontinuité d'éligibilité au RMI et au RSA autour de 25 ans par Bargain et Vicard, 2014)¹². Une première limite tient au nombre limité de variables disponibles : dans la perspective de l'analyse proposée, on pourra essentiellement analyser le statut d'emploi, en désagrégeant par niveau de diplôme. Une seconde limite tient aux délais de mise à disposition : les enquêtes annuelles de recensement sont réalisées au premier trimestre de l'année, mais l'INSEE ne les met pas rapidement à disposition des chercheurs. Un enjeu sera de vérifier s'il est possible d'obtenir une mise à disposition dans l'année qui suit la collecte.

Données physiques et financières sur les programmes d'activation sur le marché du travail

Comme on l'a vu en section 2.1, les impacts éventuels mesurés sur l'emploi des jeunes dans les régions IEJ doivent être rapportés aux ressources mises en œuvre. Le reporting financier à la Commission européenne n'est pas suffisant pour cela, dans la mesure où il ne permet pas de rendre compte d'éventuels effets de substitution entre sources de financement. Une approche globale exhaustive de l'ensemble des ressources dédiées aux programmes d'activation sur le marché du travail par région et par groupe d'âge est sans doute hors d'atteinte. Notre recommandation est plutôt de s'appuyer sur les données de gestion des principaux opérateurs et commanditaires de politiques actives à destination des jeunes, en considérant à la fois des données physiques (nombre de jeunes pris en charge, mesures d'intensité de la prise en charge, ressources humaines -- ETP dédiés) et financières. Il est important, pour une analyse contrefactuelle de l'impact de l'IEJ sur les ressources dédiées aux jeunes, que ces données soient disponibles de façon homogène dans les régions IEJ et non IEJ. On pourra ainsi considérer trois types de sources au moins : les missions locales (systèmes d'information ICARE et Parcours 3/i-Milo), Pôle emploi (fichiers de suivi des demandeurs d'emploi – fichier historique – et données de gestion sur les conseillers), ainsi que les données sur les formations collectées dans le cadre du CNEFOP.

¹² Cet article est par ailleurs intéressant dans la mesure où il propose un usage des EAR en coupe, année par année, alors que l'INSEE en propose plutôt un usage par groupes de 5 années (par exemple, le recensement de population « de 2012 » est basé sur les EAR de 2010 à 2014). Ce regroupement des années est nécessaire pour assurer la représentativité des données au niveau local. Il n'est cependant pas nécessaire pour des comparaisons sur des blocs de région,

Données appariées sur les mesures et sur les trajectoires des jeunes : la base TRAJAM (Trajectoire des Jeunes Appariées aux Mesures actives du marché du travail)

Pour suivre la seconde approche d'évaluation présentée en section 2.2, il faut mobiliser des données combinant l'information sur les dispositifs dont ont bénéficié ou non des jeunes et sur leurs trajectoires ultérieures. Le projet TRAJAM vise à construire une base de données statistiques permanente qui permette de suivre les trajectoires professionnelles des jeunes et leur participation à des mesures actives du marché du travail. Sa construction répond en particulier aux exigences de suivi des jeunes dans le cadre du plan de « Garantie européenne pour la jeunesse ». Elle s'inscrit aussi dans un mouvement plus large de mobilisation des sources administratives appariées en vue de l'évaluation des politiques publiques, comme l'appariement FH-DADS qui combine les données de Pôle emploi (Fichier historique) et les Déclarations annuelles de données sociales (DADS) remontées par les employeurs. Le projet TRAJAM va au-delà en combinant le suivi des jeunes sur le marché du travail avec l'information sur un ensemble de programmes de repérage, d'accompagnement, d'emplois subventionnés, de stage de formation professionnelle dont ces jeunes ont bénéficié entre 16 et 25 ans. La plupart de ces programmes ont bénéficié d'un financement IEJ. La base est en cours de construction mais le calendrier de construction paraît compatible avec le calendrier d'évaluation contrefactuelle de l'IEJ de 2018, avec une première version de la base prévue au printemps 2017, et une seconde version prévue au printemps 2018. On peut résumer la structure de la base TRAJAM de la façon suivante :

- **Périmètre** : Les jeunes sont inclus dans la base à partir du moment où ils sont repérés dans les fichiers de l'une ou l'autre des mesures actives couvertes par la base. L'un de ces fichiers s'approche d'un recensement : c'est le fichier des Journées de Défense et de Citoyenneté (JDC), dont le taux de couverture parmi les jeunes de nationalité française est estimé à 97%. Dans la première version de la base (printemps 2017), les cohortes couvertes sont les cohortes âgées de 16 à 25 ans entre 2010 et 2015.¹³ Les données sur les JDC sont disponibles à partir de 2007, ce qui permet de disposer d'une base de population pour l'ensemble des cohortes nées en 1990 et après.
- **L'information sur les programmes suivis** est issue des données administratives des différentes structures portant ou finançant les mesures couvertes par la base) : en particulier, missions locales, Pôle emploi, DGEFP, Régions (voir tableau 1). La période couverte dans la première version prévue de la base est 2010-2015.
- Le suivi des **trajectoires professionnelles** se fait au travers du « Panel tous salariés » extrait de sources administratives : les déclarations annuelles de données sociales (DADS) pour le secteur privé, et le système d'information des agents des services publics (SIASP) produit par l'Insee, qui permet d'homogénéiser le traitement des fichiers de paie de la Fonction publique d'Etat et les DADS des Fonctions publique hospitalière et territoriale. Il est complété par les

¹³ La version suivante, au printemps 2018, couvrira les cohortes âgées de 16 à 25 ans de 2010 à 2016.

données de suivi des demandeurs d'emploi de Pôle emploi (Fichier historique). Le Panel tous salariés suit un échantillon au 1/12^e défini par le jour de naissance (octobre). Le suivi des trajectoires est donc limité à cet échantillon. Les délais de construction du « Panel tous salariés » font que le suivi porte, dans la première version de la base, sur les années 2010 à 2014 (incluses).

Tableau 1 : Dispositifs dont la couverture est visée par la base TRAJAM
(source : DARES)

Nature de la mesure	Dispositifs et programmes	Propriétaire des données administratives source	Flux annuel d'entrées (ordre de grandeur)
Accompagnement	CIVIS (2010-),	DGEFP	170 000
	ANI Jeunes (2010-2015)	DGEFP	20 000
	Garantie jeunes (2013-)	DGEFP	25 000
	E2C(2010-)	Régions	14 000
	EPIDE (2010-)	DGEFP	3 500
	Programme Personnalisé d'Accompagnement vers l'Emploi (PPAE - Pôle emploi)	Pôle emploi	700 000
	Mission Lutte Décrochage Scolaire (2010-)	DGESCO (MEN)	A vérifier
Emplois subventionnés	CUI-CAE (2010-)	DGEFP	31 000
	CUI-CIE (2010-)	DGEFP	10 000
	Emplois d'avenir (2012-)	DGEFP	70 000
	Insertion par l'Activité Economique (2010-)	DGEFP	45 000
Stages de formation	Stages Etat (2010-),	DGEFP	250 000
	Stages Régions (2010-)	Régions	
	Stages Pôle Emploi	Pôle Emploi	
	Apprentissage (2010-)	DGEFP	270 000
	Contrats de professionnalisation (2010-)	DGEFP	130 000
Repérage	Journées de Défense et de Citoyenneté	Ministère de la Défense	750 000
Autres	Service civique (?-)	Agence du Service Civique	35 00

3.2 Mise en œuvre

L'évaluation d'impact contrefactuelle telle qu'elle a été présentée dans cette note requiert deux types de compétences en particulier : la maîtrise des méthodes statistiques et économétriques, et la connaissance des données. Ces compétences sont particulièrement présentes dans des équipes de recherche universitaires et les services statistiques ministériels. Deux types de mise en œuvre de l'évaluation sont donc possibles :

- une évaluation externe par appel à projet auprès d'équipes de recherche

- une évaluation interne par les services de l'administration, éventuellement en lien avec des équipes universitaires (en particulier grâce au recrutement de stagiaires ou de doctorants en bourse CIFRE, par exemple).

L'évaluation externe peut paraître préférable pour répondre aux exigences d'une évaluation indépendante. Un atout important de l'appel à projet, pour attirer les équipes universitaires, est la mise à disposition de données nouvelles, sur de larges échantillons.

L'appel à projet pourrait porter sur les deux approches d'évaluation envisagées dans cette note. Une question importante est de savoir si les deux approches doivent être exigées, et dans quelle mesure l'évaluateur peut disposer de marges de manœuvre dans le choix des dispositifs à évaluer au titre de la seconde approche. On plaide ici pour la combinaison dans le même appel des deux approches, éventuellement en lots séparés mais avec une prime aux réponses couvrant les deux lots. Le choix des dispositifs pourrait être laissé à l'appréciation des évaluateurs, dans la mesure où une part importante de la réponse à l'appel est de proposer une approche empirique crédible pour évaluer tel ou tel dispositif, approche qui peut venir de la connaissance spécifique qu'ont les chercheurs des différents dispositifs. Il est possible cependant d'orienter les réponses en demandant, par exemple, que soit évalué un dispositif de formation.

Recommandation 3 : Lancer un appel à évaluation, portant à la fois sur l'évaluation globale de l'IEJ et sur l'évaluation de dispositifs financés par l'IEJ. Laisser les évaluateurs être force de proposition dans le choix des dispositifs spécifiques évalués.

Une vigilance particulière doit être portée au calendrier de mise à disposition des données. Il serait dommage de ne pas saisir dans l'évaluation de 2018 l'opportunité d'utiliser la base TRAJAM. Mais ce sera une base récente, sans doute encore en rodage. L'analyse nécessitera des interactions importantes entre équipes de recherche et la DARES qui coordonne la construction de la base, voire avec les propriétaires des données source. L'appel à projets nécessiterait donc une concertation étroite entre la DGEFP et la DARES.

Recommandation 4 : Prendre en compte le calendrier de production de TRAJAM dans le calendrier de l'évaluation d'impact de 2018 ; le cas échéant, mobiliser une partie des financements destinés à l'évaluation d'impact de 2018 pour contribuer à une mise en place rapide de TRAJAM.

Un dernier point important porte sur le niveau géographique d'analyse. On l'a vu, les comparaisons entre territoires (notamment IEJ et non IEJ) sont un élément important des évaluations à mener. De telles comparaisons supposent de mobiliser une échelle géographique suffisante, de façon à neutraliser les chocs idiosyncratiques locaux en faisant la moyenne sur un nombre suffisants de territoires. L'évaluation à un niveau régional ou infrarégional est alors difficile. Mais les régions peuvent être fortement impliquées dans une évaluation nationale des effets de la formation

professionnelle, par exemple. Et si la présente note a porté sur l'IEJ telle qu'elle s'est déployée en 2015-2016, avec les contraintes que cela implique sur les évaluations d'impact contrefactuelles possibles, elle a aussi permis d'esquisser les possibilités offertes par l'évaluation randomisée pour l'analyse de dispositifs ou de modalités de mise en œuvre innovants, qui peuvent avoir un caractère plus local. De telles méthodes pourraient être mobilisées dans le cadre d'évaluations d'impact d'actions FSE, en sortant du calendrier d'évaluation de l'IEJ à l'horizon 2018.

Recommandation 5 : *Privilégier l'évaluation au niveau national. Mais y associer les différentes autorités de gestion, et examiner, en dehors du calendrier d'évaluation de l'IEJ, la possibilité d'évaluations d'impact randomisées de dispositifs plus locaux.*

Références

- Abadie, A., A. Diamond, and J. Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (2010): 493-505.
- Aeberhardt, R., V. Chiodi, B. Crépon, M. Gaini et A. Vicard (2014), « Du revenu contractualisé d'autonomie à la garantie jeunes : les enseignements d'une expérimentation », in [De l'éducation à l'insertion : dix résultats du Fonds d'expérimentation pour la jeunesse](#) (FEJ), Rapport INJEP / Documentation française.
- Arceneaux, K. A., S. Gerber et D. Green (2006), "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment", *Political Analysis*, 14, pp. 37-62.
- Bargain, O. et A. Vicard (2014), « Le RMI et son successeur le RSA découragent-ils certains jeunes de travailler ? Une analyse sur les jeunes autour de 25 ans », *Economie et statistiques*, n°467-468.
- Behaghel, L., B. Crépon, M. Gurgand, T. Kamionka, L. Lequien, R. Rathelot et P. Zamora (2013), "L'accompagnement personnalisé des demandeurs d'emploi: les enseignements de trois expériences contrôlées menées en France," *Revue française d'économie*, XXVIII, pages 123-58.
- Bléhaut, M., et R. Rathelot (2014), "Expérimentation contrôlée contre appariement : le cas d'un dispositif d'accompagnement de jeunes diplômés demandeurs d'emploi," *Economie et Prévision*, 204-205(1), pp.163-181.
- Card, D., J. Kluve et A. Weber (2010), "Active Labour Market Policy Evaluations: A Meta-Analysis", *The Economic Journal*, 120 (548), pages F452–F477.
- Card, D., J. Kluve et A. Weber (2015), "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations", NBER Working Paper No. 21431.
- Crépon, B., E. Duflo, M. Gurgand, R. Rathelot et Ph. Zamora, "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment", *Quarterly Journal of Economics*, vol. 128, n°2, 531-580, May 2013.
- ¹ DGEFP (2015), *Cahier des clauses techniques particulières pour l'évaluation de l'impact du programme européen "Initiative pour l'emploi des jeunes" en 2015*.
- European Commission (2015), *Guidance on evaluation of the youth employment initiative*, DG Employment, Social Affairs and Inclusion, septembre 2015.
- Friedlander, D. et P. K. Robins (1995), "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods", *American Economic Review*, 85(4), pp. 923-37.
- Givord, P. (2015) « Méthodes économétriques pour l'évaluation de politiques publiques », *Economie et Prévision*, 2015, n°204-205.
- Heckman J. et al. (1998) "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66(5), pp. 1017-1098.

J-PAL Europe (2011) *Social Experimentation, A Methodological Guide for Policy Makers*, <http://ec.europa.eu/social/BlobServlet?docId=7102&langId=en>).

LaLonde, R. J. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 73(1), pp. 31-43.

Rapport national (2015), *Evaluation de l'impact du programme européen "Initiative pour l'emploi des jeunes en 2015"*.

Smith, J. et P. Todd (2005), "Does Matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, 125 (1-2), pp. 305-353.

Zamora, P. (2016), "Les programmes d'accompagnement des jeunes", présentation au Conseil d'orientation de l'emploi le 16 février 2016, http://www.coe.gouv.fr/download.php?file_url=IMG/pdf/presentation_coe_v210216_Ph-_Zamora.pdf.