HAL open science

Testing (ou méthode d'étude par correspondance)

Nicolas Jacquemet

To cite this version:

Nicolas Jacquemet. Testing (ou méthode d'étude par correspondance). LIEPP Fiche méthodologique n°5, 2023. hal-04094010

HAL Id: hal-04094010

https://sciencespo.hal.science/hal-04094010v1

Submitted on 10 May 2023

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

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

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

SciencesPo
LABORATORY FOR INTERDISCIPLINARY EVALUATION OF PUBLIC POLICIES

LIEPP FICHE MÉTHODOLOGIQUE n°5

2023

Cette fiche fait partie d'un ensemble de fiches méthodologiques publiées par le LIEPP. A ce titre, elle bénéficie du soutien apporté par l'ANR et l'État au titre du programme d'Investissements d'avenir dans le cadre de l'IdEx Université Paris Cité (ANR-18-IDEX-0001).

Testing (ou méthode d'étude par correspondance)

Nicolas JACQUEMET (Université Paris 1, CES, PSE)

nicolas.jacquemet@univ-paris1.fr

Partage selon les Conditions Initiales 4.0 International License

www.sciencespo.fr/liepp

Comment citer cette publication :
JACQUEMET, Nicolas, **Testing (ou méthode d'étude par correspondance)**, *LIEPP Fiche méthodologique* n°5, 2023-05

This publication is also available in English:
JACQUEMET, Nicolas, **Testing (or correspondence study method)**, *LIEPP Methods Brief* n°6, 2023-05

EN BREF :

Le testing, ou méthode d'étude par correspondance, est une méthode quantitative permettant de mesurer les discriminations. Elle passe par l'envoi de candidatures fictives en réponse à des offres réelles (par exemple, offres d'emploi). Par l'objectivation des comportements discriminatoires, cette méthode est très utile, sur le plan prospectif, pour la conception des politiques antidiscriminatoires.

Mots-clés : Méthode quantitative, correspondance, discriminations, politiques antidiscriminatoires, candidatures expérimentales

I. En quoi cette méthode est-elle utile pour l'évaluation des politiques publiques ?

Les discriminations correspondent à une inégalité de traitement sur la base de caractéristiques individuelles qui ne devraient pas être pertinentes pour la décision à prendre : favoriser une candidature masculine dont les compétences professionnelles sont supérieures n'est pas discriminatoire ; mais écarter une candidature féminine sur la base du soupçon que la disponibilité de cette candidate sera moindre que celle d'un candidat dont le profil est équivalent est bel et bien discriminatoire, car rien ne permet de penser que la candidate en question correspond à ce stéréotype. A ce titre, les discriminations sont une source d'inégalité majeure. Plus encore que d'autres types d'inégalités, elles sont à la fois très coûteuses sur le plan économique, en privant l'économie de certains de ses talents ; et persistantes, car l'anticipation de telles inégalités de traitement découragent les personnes discriminées et les conduisent à faire des choix (de niveau et de filière d'éducation, de parcours professionnel) qui ne font qu'amplifier ces inégalités de départ.

Malgré l'importance des enjeux, l'élaboration de politiques publiques destinées à lutter contre les discriminations souffre du manque d'éléments de diagnostic dû à la très grande difficulté de la mesurer. C'est l'objectif poursuivi par la méthode d'étude par correspondance (« *testing* »).

II. En quoi consiste cette méthode ?

Bien que cette méthode puisse être appliquée à de très nombreux secteurs de l'activité économique (recherche de logement, locations saisonnières, candidature des étudiants en master) et à de nombreuses sources différentes de discrimination (la religion, les préférences sexuelles, les origines socio-économiques, le lieu de résidence, le handicap) cette présentation se concentre par simplicité sur l'application de cette méthode à la mesure des discriminations à l'embauche fondées sur le sexe et / ou l'origine.

Cette méthode quantitative est conçue de manière à fournir une mesure du succès de différents types de candidats en fonction de leurs caractéristiques socio-démographiques, mais tout en neutralisant l'effet de la qualité intrinsèque des candidatures. Chacun de ces deux objectifs a des implications méthodologiques qui lui sont propres.

Constitution des candidatures fictives

Le succès de différents types de candidats est observé grâce au recours à des candidatures artificielles, envoyées en réponse à des offres d'emploi réelles circulant sur le marché du travail. La méthode combine pour ce faire trois ingrédients : des identités, des candidatures, et des offres d'emploi.

Les caractéristiques socio-démographiques dont on cherche à mesurer l'effet sont véhiculées par l'identité du candidat. Pour tester à la fois les discriminations liées au sexe et les discriminations qui

affectent les candidats issus, par exemple, de l'immigration maghrébine, il s'agira donc de créer une liste de quatre identités fictives (ou quatre catégories d'identités différentes) : deux patronymes à consonnance française, l'un associé à un prénom masculin et l'autre à un prénom féminin, et deux patronymes qui suggèrent que la personne est issue de l'immigration maghrébine, associées aux mêmes variations du prénom. Chacune de ces identités se voit dotée d'un numéro de téléphone unique et d'une adresse électronique permettant d'entrer en contact avec les candidats. Ces noms et prénoms ainsi que ces informations de contact correspondent au **bloc identité** des candidatures.

Pour répondre à des offres d'emploi, ces identités sont portées sur des candidatures qui combinent le plus souvent un CV et une lettre de motivation. L'objectif est de construire des candidatures qui soient aussi crédibles que possible, et permettront de distinguer le succès de différentes identités. Il faut donc que le processus de construction des candidatures conduise à une qualité qui ne soit ni trop élevée ni trop faible en comparaison des candidatures réelles qui seront reçues, car toute candidature qui conduit à un traitement indifférencié des candidats, qu'il soit positif ou négatif, rend impossible l'identification des caractéristiques favorisant le succès des candidatures expérimentales.

La construction du CV nécessite de choisir le contenu de la section formation ainsi qu'une liste d'expériences qui soient toutes deux réalistes et compatibles avec le métier pour lequel la candidature est envoyée, ainsi qu'une section dédiée aux activités extra-professionnelles. Afin de garantir la correspondance entre ces éléments de CV et les métiers visés, la plupart des études collectent de vrais CV (disponibles, par exemple, en ligne), mélangent les informations issues de plusieurs CV pour construire un CV unique puis modifient à la marge les sections « expérience », « formation », et « activités extra-professionnelles » ainsi obtenues. Le contenu du CV est complété par un bloc contenant des informations personnelles permettant d'indiquer *a minima* l'adresse postale, à laquelle peuvent éventuellement être associés le statut marital, la présence d'enfants, l'âge ou encore la date de naissance. La mise en forme de ces informations nécessite de choisir autant de modèles prédéfinis que de CV différents, qui détermineront l'ordre des sections, la police de caractère utilisée ainsi que l'organisation des différentes informations (de nombreux modèles dans différents formats de fichier peuvent être facilement trouvés en ligne). Les lettres de motivation sont construites de la même manière, en combinant le contenu de lettres de motivation existantes. Les accords de genre seront adaptés en fonction du sexe de l'identité portée sur la candidature (si le sexe fait partie des caractéristiques testées, il est souhaitable de choisir les formulations qui permettent de multiplier les accords de genre). Le couple formé par un CV et une lettre de motivation correspond à **une candidature**.

Les **offres d'emploi** auxquelles ces candidatures seront envoyées sont collectées sur des sites publics d'information (il s'agit souvent du site de Pôle emploi qui est le service public de l'emploi en France, mais selon les métiers visés il est parfois nécessaire de recourir à des sites spécialisés). Ces offres sont filtrées pour vérifier qu'elles correspondent aux critères d'inclusion prédéfinis, qui concernent au premier chef le métier et la localisation de l'emploi, mais aussi par exemple l'exigence d'expériences ou de compétences spécifiques. Les offres pour lesquelles il n'est pas possible d'envoyer une candidature selon les modalités préalablement définies (souvent, par mail, mais aussi lorsque par exemple le dépôt d'une candidature nécessite de compléter un questionnaire en ligne) sont systématiquement écartées. Pour les offres d'emploi restantes, qui seront incluses dans l'étude, l'ensemble des caractéristiques de l'offre (durée, type de contrat, salaire, etc.) sont soigneusement enregistrées afin de constituer une base permettant de documenter l'hétérogénéité observée des offres d'emploi.

Le **nombre de candidatures** expérimentales qui seront envoyées en réponse à une offre d'emploi donnée (qui va de pair avec le nombre de candidatures différentes qu'il est nécessaire de construire) est un choix délicat. D'un point de vue statistique, il est très avantageux de pouvoir comparer le succès de différentes candidatures en réponse à une offre d'emploi donnée (i.e., comparaisons « intra-offre »), car une telle comparaison élimine l'effet de tous les éléments inobservés qui sont spécifiques à l'offre d'emploi

et améliore donc la précision statistique des mesures. L'envoi de plusieurs candidatures correspondant à un groupe socio-démographique donné permet également de mesurer plus finement les caractéristiques de la distribution de la discrimination entre les différentes annonces (voir les résultats présentés dans Kline et al. 2020). S'il est par conséquent souhaitable d'envoyer plusieurs candidatures en réponse à chaque offre, le nombre maximum est limité par deux facteurs. D'une part, la multiplication des candidatures augmente les perturbations induites par la réalisation de l'étude sur le fonctionnement du marché du travail et, surtout, le risque de détection. Ce risque peut être contenu en prenant soin de laisser un délai suffisant entre l'envoi de deux candidatures, mais ce délai va de pair avec une diminution de la probabilité de succès pour les candidatures les plus tardives et ce d'autant plus que le métier attire un nombre important de candidatures. D'autre part, certains travaux récents (Philips, 2019) montrent que le portefeuille de candidatures envoyées en réponse à une offre donnée est susceptible d'affecter le succès relatif des candidatures expérimentales. L'augmentation du nombre de candidatures augmente les risques de tels biais.

Ces deux facteurs conduisent à être d'autant plus restrictif quant au nombre de candidatures envoyées que le métier est en tension. La combinaison de ces différents facteurs conduit la plupart des études à se limiter à l'envoi de 4 candidatures au maximum en réponse à chaque offre d'emploi, envoyées jusqu'à 24h au plus tard après leur publication. Pour ce faire, chaque identité est associée à une candidature unique (un CV et une lettre de motivation), conduisant à autant de candidatures expérimentales uniques et distinctes que le nombre d'envois en réponse à chaque offre.

La mesure du **succès** des candidatures expérimentales nécessite de conserver une trace précise, et limitée dans le temps (en ignorant par exemple les réponses reçues plus de 3 mois après leur envoi), de la communication des employeurs avec les candidats en archivant toute correspondance écrite et en retranscrivant le contenu des messages téléphoniques reçus. Ces réponses sont ensuite classifiées afin de distinguer les refus, les non-réponses, les demandes d'informations complémentaires et les convocations à un entretien d'embauche (qui sont parfois regroupées sous le terme de 'manifestations d'intérêt'). Pour des raisons éthiques évidentes, il est impératif de décliner toute manifestation d'intérêt aussi vite que possible, préférentiellement selon les mêmes modalités de contact et en suivant un script prédéfini (qui prétexte le plus souvent l'acceptation antérieure d'une offre d'embauche).

Protocole d'assemblage

La combinaison de l'ensemble de ces ingrédients fournit une mesure du succès de candidatures fictives qui se distinguent, entre autres, par le groupe socio-démographique auquel est associé l'identité portée sur la candidature. Bien évidemment, de telles différences de succès peuvent également être liées au contenu de la candidature elle-même, ce qui est d'autant plus probable que les candidatures sont nettement différentes les unes des autres. Une solution pourrait donc être de s'assurer que les candidatures sont aussi proches que possible les unes des autres. Mais outre que toute différence, même infime, entre les candidatures conduit à la même conclusion, il est particulièrement difficile de distinguer des différences négligeables de différences plus importantes, car les différences qui sont pertinentes concernent les variations subjectives de la qualité des candidatures telle que perçues par les personnes en charge du recrutement.

Le protocole qui permet aux études par correspondance de neutraliser l'effet de toutes les caractéristiques des candidatures expérimentales qui pourraient être confondantes (i.e., dont l'impact sur le taux de succès conduirait à des conclusions erronées quant à l'effet des caractéristiques d'intérêt) consiste à organiser une rotation systématique de l'association entre identités d'une part et candidatures d'autre part. Si, par exemple, l'identité *a* est portée sur la candidature *A* et l'identité *b* sur la candidature *B* lors du premier envoi, ces associations seront inversées lors de l'envoi suivant (l'identité *a* apparaissant sur la candidature *B*) avant de revenir à l'association initiale lors du troisième envoi, etc. Cette rotation

n'élimine pas l'effet de la qualité perçue de la candidature : si la candidature *A* se trouve être de meilleure qualité, le succès de l'identité qui lui est associée en sera affecté. Mais la rotation permet de s'assurer que toute différence systématique de succès associé à l'identité sur l'ensemble des envois ne puisse plus être attribuée au contenu de la candidature. D'un point de vue statistique, toute caractéristique pour laquelle une rotation systématique est organisée devient une source de bruit dans la mesure de la discrimination liée aux caractéristiques d'intérêt, c'est à dire une source de variation du taux de succès qui distingue les candidatures appartenant à différentes catégories sans toutefois être imputables à la discrimination. Par construction, ce bruit est indépendant des caractéristiques dont on cherche à mesurer l'effet et ne nuit donc pas à la capacité de la méthode de mesurer la discrimination, mais rend sa détection plus difficile. Ces conséquences du bruit dans les mesures peuvent être réduites en adaptant l'analyse statistique en conséquence (sous la forme d'effets fixes offres), mais une telle modélisation suppose un effet homogène de la qualité des candidatures sur l'ensemble des employeurs.

Au total, la méthode d'étude par correspondance repose donc sur trois principes : démultiplier le nombre de candidatures expérimentales afin de mesurer l'effet des caractéristiques socio-démographiques par lesquelles elles se distinguent, s'assurer de la plus grande homogénéité possible de ces candidatures afin de réduire le bruit qui affectera la mesure de leur effet, et organiser une rotation systématique de l'association entre les profils socio-démographiques et toute autre caractéristique susceptible d'en affecter le succès. Ces trois principes constituent une boîte à outils qui peut être déclinée à de très nombreux aspects du fonctionnement du marché du travail. On peut ainsi mesurer l'effet, par exemple, de la présence d'épisodes de chômage dans le parcours professionnel en modifiant expérimentalement la section « expérience » des candidatures, de la distance entre le domicile et le travail ou du lieu de résidence en manipulant l'adresse d'habitation, ou de la situation familiale en faisant varier le bloc identité selon la présence d'enfants ou le statut marital.

III. Un exemple d'utilisation de cette méthode

Une étude récente réalisée conjointement par l'Institut des Politiques Publiques et ISM Corum sous l'égide de la DARES est l'une des premières études de grande ampleur permettant de dresser un panorama des inégalités d'accès à l'emploi selon le sexe et l'origine sur le marché du travail français (Dares IPP et ISM Corum, 2021a et 2021b). Ces résultats confirment l'existence de discriminations liées à l'origine qui sont à la fois fortes et transversales à l'ensemble des métiers étudiés, conduisant à un handicap de l'ordre 30% dans les chances de recevoir une réponse positive. Cette étude met également en évidence l'absence de discrimination liée au sexe du candidat, suggérant que, contrairement à une idée reçue persistante, les fortes inégalités de carrière qui existent sur le marché du travail entre les hommes et les femmes ne peuvent pas être attribuées aux décisions d'embauche.

IV. Quels sont les critères permettant de juger de la qualité de la mobilisation de cette méthode ?

Le niveau du taux de rappel pour un type de candidature donné ne fournit que peu d'informations sur le fonctionnement du marché du travail. Les résultats issus de la méthode du *testing* proviennent des comparaisons de taux de rappel entre différents types de candidatures. Pour que ces comparaisons permettent de détecter l'écart de succès que rencontrent différents types de candidats, il est important que les taux de rappel des candidatures de référence soient suffisamment élevés.

Les variations des caractéristiques socio-démographiques sont introduites par l'intermédiaire de l'identité portée sur les candidatures, qui est supposée affecter les perceptions des personnes en charge du recrutement. Pour s'en assurer, il est de plus en plus fréquent dans les études de testing de faire précéder l'étude d'une enquête dans laquelle un échantillon de répondants doit associer un sexe et / ou une origine

à chacune des identités qui leur sont présentées. Cette enquête fournit une mesure empirique de la qualité des perceptions induites par les identités, et peut permettre de sélectionner les identités en conservant celles dont les perceptions sont les plus cohérentes avec le groupe d'appartenance souhaité. Une telle enquête peut également être l'occasion de collecter des informations supplémentaires sur le profil perçu des identités présentées : des travaux récents montrent en effet que les identités véhiculent de nombreux stéréotypes liés par exemple à la classe sociale ou à la zone d'habitation qui peuvent contribuer aux différences de succès observées des candidatures de différentes catégories (Gaddis, 2017).

Enfin, les différences observées de taux de rappel sont sujettes à la fameuse critique connue sous le nom de « critique d'Heckman », selon laquelle des différences perçues par les employeurs dans la variance des compétences à l'intérieur des différents groupes de population suffirait à produire des écarts systématiques en moyenne dans les taux de rappel, et serait interprétée à tort comme un biais systématique à l'encontre de ces groupes de population. L'existence de différences suffisantes de qualité entre les différentes candidatures expérimentales peut permettre de prendre en compte cet effet en menant des analyses statistiques qui autorisent des variances dans les éléments inobservés qui diffèrent entre groupes (Neumark, 2012).

V. Quels sont les atouts et les limites de cette méthode par rapport à d'autres ?

Le protocole de mesure fourni par les *testing* permet d'évaluer de manière précise et convaincante l'ampleur des pratiques discriminatoires et l'effet spécifique des caractéristiques socio-démographiques des candidats sur le succès de leur insertion sur le marché du travail. Il permet d'objectiver un phénomène que les approches qualitatives ont plus de difficultés à révéler : ces pratiques ne sont pas aisément verbalisées en entretien semi directif par exemple, car illégitimes et parfois non conscientes. Son principal atout est de garantir par construction l'indépendance entre ces caractéristiques et l'ensemble des autres éléments qui composent la candidature. La principale alternative est d'utiliser des données d'enquête pour étudier les différentiels de parcours sur le marché du travail entre différentes catégories de population. Mais de telles études nécessitent des hypothèses statistiques fortes, et souvent peu crédibles, destinées à neutraliser l'effet des différences d'éducation ou de parcours professionnel qui distinguent ces groupes de population et contribuent aux écarts de succès observés sur le marché du travail.

La portée des résultats produits par l'application de cette méthode est néanmoins limitée par deux facteurs importants.

Le premier est que la mesure du succès des candidatures repose entièrement sur la convocation, ou non, à un entretien d'embauche. Or cette étape n'est que le reflet imparfait du résultat final du processus de recrutement : l'existence d'une discrimination à cette étape du processus ne prédit une discrimination lors de l'embauche effective qu'à condition que tous les candidats convoqués soient traités sur un pied d'égalité. Si, au contraire, une discrimination supplémentaire s'exerce lors des entretiens d'embauche, les mesures de discrimination fournies par cette méthode sous-estiment le phénomène. S'il s'avère enfin que les populations discriminées dans le tri des candidatures sont favorisées dans des proportions strictement inverses lors du choix du candidat final, alors ces mesures distordent la réalité des discriminations. Les méthodes d'audit, qui consistent à faire appel à des acteurs jouant le rôle de candidats expérimentaux mais réels, permettent de dépasser cette limite, mais elles présentent l'inconvénient de faire intervenir un ensemble très large de facteurs (l'apparence physique, la voix) qui sont susceptibles d'influencer le processus de recrutement mais ne peuvent pas être distingués des caractéristiques socio-démographiques qui sont apparentes.

La seconde limite est commune à toute étude empirique mais s'avère particulièrement aigue dans le cas des *testings* : comme discuté plus haut, les mesures sont d'autant plus précises que les candidatures

sont homogènes. S'y ajoutent des raisons pratiques, liées au fait que le nombre et la spécificité des candidatures fictives augmente avec la diversité (géographique ou en termes de métiers) des offres d'emploi. En conséquence, les études de *testing* sont souvent circonscrites à un périmètre limité, et leurs résultats ne peuvent donc qu'être conditionnels au champ d'application de l'étude en termes de type d'emploi, de secteur d'activité, de zone géographique, de tranche d'âge des candidats, etc. La généralisation à l'ensemble du marché du travail des résultats observés dans le cadre de ce type d'étude repose donc sur l'hypothèse que le périmètre choisi ne présente pas de spécificités en termes de propension à discriminer (préférences des recruteurs, degré de concurrence dans le recrutement, etc.) ou, de façon plus convaincante, sur l'accumulation d'études concordantes portant sur différente sphères du marché du travail.

Quelques références bibliographiques pour aller plus loin

Adamovic, Mladen. « Analyzing Discrimination in Recruitment: A Guide and Best Practices for Resume Studies ». *International Journal of Selection and Assessment* 28, no 4 (2020): 445-64.

Adida, C.L., Laitin, D.D., Valfort, M.-A., 2010. Identifying barriers to Muslim integration in France. *Proceedings of the National Academy of Sciences* 107, 22384–22390.

Dares IPP et ISM Corum, 2021a, « Discrimination à l'embauche selon le sexe : les enseignements d'un testing de grande ampleur », *Dares Analyses* n°26/Note IPP n°67

Dares IPP et ISM Corum, 2021b, « Discrimination à l'embauche des personnes d'origine supposée maghrébine: quels enseignements d'une grande étude par testing? *Note IPP* n°76 / *Dares Analyses* n°67

Edo, A., Jacquemet. N. (2013). La discrimination à l'embauche, Sur le marché du travail Français. *Opuscule du CEPREMAP* n°31, Editions rue d'Ulm

Kline, Patrick M, et Christopher R Walters. « Reasonable doubt: Experimental detection of job-level employment discrimination ». *Econometrica* 89, no 2 (2020): 765-92.

du Parquet, Loïc, et Pascale Petit. « Discrimination à l'embauche : retour sur deux décennies de testings en France ». *Revue française d'economie* Vol. XXXIV, no 1 (2019): 91-132.

Gaddis, S Michael. « How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies ». *Sociological Science* 4 (2017): 469-89. https://doi.org/10.15195/v4.a19.

Neumark, David. « Detecting Discrimination in Audit and Correspondence Studies ». *Journal of Human Resources* 47, no 4 (2012): 1128-57.

Phillips, D.C., 2019. Do Comparisons of Fictional Applicants Measure Discrimination When Search Externalities Are Present? Evidence from Existing Experiments. *Economic Journal* 129, 2240–2264.

Fougère, Denis, Roland Rathelot, et Romain Aeberhardt. « Commentaire: Les méthodes de testing permettent-elles d'identifier et de mesurer l'ampleur des discriminations ? » *Economie et Statistique* 447, no 1 (2011): 97-101.