

CONSÉQUENCES À LONG TERME DES CHÈQUES D'ÉDUCATION SECONDAIRE : DONNÉES TIRÉES DE DOSSIERS ADMINISTRATIFS EN COLOMBIE*

Joshua Angrist¹, Eric Bettinger², and Michael Kremer³

RÉSUMÉ

Bien que l'utilisation de chèques-éducation devienne de plus en plus fréquente dans les pays en voie de développement, l'effet de ces programmes reste à déterminer. Les évaluations des effets à long terme sont particulièrement rares. Le présent article décrit l'estimation de l'effet à long terme du programme PACES de la Colombie, qui a fourni à plus de 125 000 élèves de quartiers pauvres des chèques dont le montant couvrait la moitié du coût des études secondaires en établissement privé. Ce programme offre une occasion inaccoutumée d'évaluer l'effet du financement de la demande de services éducatifs dans un pays d'Amérique latine où les écoles privées assurent la formation d'une part importante des élèves. Le programme PACES présente un intérêt particulier, car nombre de chèques ont été attribués par tirage au sort, de sorte que l'on peut en évaluer les effets de façon fiable. Nous nous servons de dossiers administratifs pour déterminer l'effet à long terme de l'obtention de chèques du PACES sur la situation d'obtention du diplôme d'études secondaires et sur les notes d'examen. Les avantages principaux des dossiers administratifs sont qu'il n'y a aucune perte due à l'érosion de l'échantillon et que les données reviennent nettement moins cher que celles recueillies au moyen d'une enquête coûteuse et éventuellement dangereuse. En revanche, les numéros d'identification personnels pourraient être consignés incorrectement, compliquant ainsi le couplage des enregistrements, et un biais de sélection contamine l'échantillon d'élèves qui passent l'examen d'entrée au collège ou à l'université. Nous discutons de diverses solutions pour résoudre ces problèmes. Les résultats semblent indiquer que le programme de chèques-éducation fait augmenter le taux d'obtention du diplôme d'études secondaires et que les notes de l'examen d'entrée au collège sont plus élevées pour les demandeurs gagnants lors du tirage au sort des chèques que pour les perdants.

1. INTRODUCTION

Les pays d'Amérique latine et d'autres pays en développement sont de plus en plus intéressés par la mise en place de programmes de subventions de la demande d'enseignement primaire et secondaire. Mentionnons, à titre d'exemples, le Chili et la Colombie, qui ont offert des chèques-éducation permettant la fréquentation des écoles secondaires privées, ainsi que le Brésil, l'Inde, Israël et le Mexique, qui ont introduit le versement d'allocations aux élèves pour récompenser la fréquentation et les bons résultats scolaires. L'intérêt pour les subventions de la demande dans les pays en développement fait pendant à celui manifesté aux États-Unis, où, par exemple, des chèques émis par l'État ont été distribués au Milwaukee, à Cleveland et en Floride.

Jusqu'à présent, la plupart des données sur la distribution de chèques-éducation ont trait principalement aux effets à court terme, tels que mesurés par les notes d'examen au primaire et au secondaire. Ces données laissent entendre que les chèques bénéficient aux élèves qui les obtiennent, quoique la mesure dans laquelle les notes d'examen de ces élèves augmentent continue de faire l'objet de débats⁴. La plupart des études sur les effets des chèques-éducation ne comporte pas d'évaluation de leur impact sur les résultats à long terme, comme le taux d'élèves qui obtiennent un diplôme d'études secondaires, qui sont corrélés de façon plus évidente à la réussite économique. La recherche sur les chèques-éducation se concentre naturellement sur les résultats à court terme, car les sujets observés sont habituellement jeunes. De surcroît, les efforts de suivi à long terme sont ardues et coûteux, particulièrement dans les

¹ MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142

² Department of Economics, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio 44106-7206

³ Department of Economics Littauer Center, 207 Cambridge, MA 02138

⁴ Consulter, p. ex., Rouse (1998), Howell et coll. (2000) et Krueger (2002)

pays en développement. Les familles déménagent souvent, les adresses sont inconnues ou inexactes, et les conditions de travail sur le terrain sont difficiles pour les enquêteurs.

Le présent article décrit les résultats d'une étude des effets à long terme de l'une des plus importantes initiatives de distribution de chèques-éducation jamais entreprises, à savoir le PACES mis en place en Colombie⁵. Ce programme offre une occasion inaccoutumée d'évaluer l'effet du financement de la demande de services éducatifs dans un pays d'Amérique latine où les écoles privées assurent la formation d'une part importante des élèves. De 1991 à 1997, le PACES a octroyé presque 125 000 chèques à des élèves du secondaire vivant dans une famille à faible revenu. À Bogota, ainsi que dans plusieurs autres grandes villes, la demande a excédé l'offre, si bien que les chèques du PACES ont été attribués par tirage au sort. Cette assignation aléatoire des chèques facilite l'adoption d'un plan d'expérience dans les conditions naturelles, où les perdants représentent un groupe témoin pour les gagnants.

Lors d'une étude antérieure (Angrist, Bettinger, Bloom, King et Kremer, 2002), nous nous sommes fondés sur l'attribution des chèques du PACES par tirage au sort pour estimer l'effet sur le nombre d'années d'études achevées entre la 6^e et la 8^e années, le redoublement au cycle secondaire, les notes d'examen et la probabilité de trouver du travail. Les résultats donnent fortement à penser que les élèves qui gagnent un chèque-éducation consacrent plus de temps à leurs études et apprennent davantage à l'école que s'ils n'avaient pas reçu le chèque. La plupart des données de notre étude antérieure portent sur des élèves observés environ trois ans après qu'ils aient gagné le chèque, mais auxquels il restait encore trois années d'études avant d'obtenir le diplôme d'études secondaires. Elles indiquent manifestement que les chèques ont un effet bénéfique à court terme sur les études et le rendement scolaire, mais elles ne nous disent pas s'ils améliorent les résultats ayant une signification économique comme le taux d'obtention d'un diplôme secondaire et le rendement obtenu à la fin des études.

Ici, nous entreprenons une évaluation des effets à plus long terme en nous appuyant sur des données administratives provenant du service centralisé d'examen d'entrée au collège de la Colombie, appelé ICFES⁶, qui organise un examen auquel s'inscrivent la plupart des diplômés du secondaire. Pour appliquer cette méthode, nous avons procédé à l'appariement des données provenant des listes d'élèves ayant fait une demande au PACES à celles des dossiers d'examen de l'ICFES au moyen des numéros d'identification d'élève. Comme lors de notre première étude, nous avons utilisé l'assignation aléatoire des chèques-éducation par tirage au sort pour préciser les relations de cause à effet.

Les principaux avantages des dossiers administratifs comparativement aux méthodes d'enquête sont qu'il n'y a aucune perte par érosion de l'échantillon et que les données administratives reviennent nettement moins cher que celles recueillies au moyen d'une enquête coûteuse (et éventuellement dangereuse). Cependant, les études fondées sur des données administratives posent aussi des problèmes. Alors que les dossiers de demande de chèque du PACES constituent une population cible complète pour l'appariement, en pratique, ils pourraient contenir des données inexactes. Ces inexactitudes peuvent compliquer le couplage des enregistrements et donner lieu à un taux d'appariement qui favorise les gagnants du tirage au sort. En outre, un biais de sélection contamine l'échantillon d'élèves qui font l'examen de l'ICFES, ce qui rend plus difficile les comparaisons des notes d'examen des gagnants et des perdants au tirage au sort. Nous examinons divers moyens de résoudre ces problèmes. Dans l'ensemble, les résultats de notre effort de suivi indiquent que les chèques ont des effets durables sur les gagnants qui affichent un taux d'obtention du diplôme d'études secondaires nettement plus élevé et, parmi ceux qui font l'examen ICFES, des notes d'examen plus élevées, après correction pour tenir compte du biais de sélection.

La section II fournit des renseignements contextuels sur le PACES et sur la stratégie d'évaluation. Nous décrivons ensuite le cadre d'estimation. À la section III, nous présentons les mesures de l'effet des chèques du PACES sur le taux d'obtention du diplôme d'études secondaires. À la section IV, nous discutons du problème que pose le biais de sélection dans l'analyse des notes d'examen et nous présentons des estimations des effets sur les notes obtenues au moyen de diverses méthodes de correction du biais de sélection. À la section V, nous présentons nos conclusions.

⁵ PACES est l'acronyme de Programa de Ampliación de Cobertura de la Educación Secundaria.

⁶ ICFES est l'acronyme de Instituto Colombiano Para El Fomento De La Educación Superior.

2. CONTEXTE

2.1 Le programme PACES

Le gouvernement colombien a établi le PACES à la fin de 1991 dans le cadre d'une démarche générale de décentralisation et d'un effort en vue d'augmenter rapidement la capacité scolaire et d'accroître le taux d'inscription à l'école secondaire (King, Orazem et Wolgemuth, 1998). En 1993, alors que 89 % des enfants colombiens en âge d'école primaire étaient inscrits à l'école, 75 % seulement de ceux admissibles au cycle secondaire étaient inscrits. Parmi les enfants admissibles du quintile le plus pauvre de la population, 78 % étaient inscrits à l'école primaire, mais 55 % seulement étaient inscrits à l'école secondaire (Sanchez et Mendez, 1995).

Le PACES, qui visait les familles à faible revenu, offrait des chèques-éducation aux enfants des quartiers considérés comme faisant partie des deux strates de niveau socio-économique le plus faible (sur six strates possibles). Pour avoir droit à un chèque, le demandeur devait être prêt à entrer au cycle d'études secondaires de la Colombie qui débute à la 6e année et avoir 15 ans ou moins. Avant de faire la demande, les élèves devaient avoir déjà été admis par une école secondaire participant au programme (c.-à-d. une école qui accepterait le chèque). Les écoles participantes devaient être situées dans les villes participant au programme, c'est-à-dire toutes les grandes villes colombiennes. Un peu moins de la moitié des écoles privées des 10 plus grandes villes acceptaient les chèques en 1993.

Les bénéficiaires des chèques avaient droit au renouvellement automatique jusqu'à la onzième année, qui est la dernière du cycle secondaire en Colombie, à condition que leurs résultats scolaires garantissent le passage à l'année suivante. En principe, les élèves transférés d'une école privée participante dans une autre pouvaient transférer le chèque à la nouvelle école. Toutefois, en pratique, selon notre enquête, nombre d'élèves qui ont changé d'école après avoir gagné le chèque ont perdu ce dernier. Les grandes et les petites villes utilisaient un système de tirage au sort pour répartir les chèques lorsque la demande excédait l'offre. Dans la présente étude, nous nous concentrons sur l'une des cohortes les plus importantes de demandeurs, c'est-à-dire les élèves de Bogota qui ont fait une demande de chèques-éducation en 1995. En fait, les membres de la cohorte de demandeurs de 1995 ont fait la demande au PACES en 1994 afin d'entrer dans une école privée en 6e année en 1995. En supposant qu'il n'y a pas eu de redoublement, les élèves de cette cohorte se sont retrouvés en 11e année, c'est-à-dire la dernière année du cycle secondaire, en 2000. Tout élève de 11e année ou tout diplômé de l'école secondaire peut passer l'examen de l'ICFES, qui est exigé pour l'admission aux collèges et universités de la Colombie.

2.2 Données et statistiques descriptives

L'étude est fondée sur l'appariement des dossiers informatisés des membres de la cohorte de demandeurs de chèques du PACES de Bogota de 1995 aux dossiers de l'ICFES indiquant si ces élèves se sont inscrits à l'examen et, le cas échéant, la note qu'ils ont obtenue. La description détaillée de la méthode d'appariement et de la base de données sur les notes d'examen figure à l'annexe sur les données. Le tableau 1 donne les statistiques descriptives pour la cohorte du PACES de 1995 pour Bogota qui comptait 4 044 demandeurs. Les caractéristiques démographiques présentées dans le tableau sont extraites des formulaires de demande soumis au PACES. Environ 59 % des demandeurs ont obtenu le chèque-éducation. Les demandeurs avaient presque 13 ans, en moyenne, et comptaient des proportions à peu près égales de garçons et de filles. Environ 88 % faisaient partie d'un ménage ayant le téléphone ou l'accès à un téléphone.

Nous avons apparié les demandeurs du PACES aux dossiers correspondants de l'ICFES au moyen du numéro d'identification national colombien, qui comprend 11 chiffres. Les six premiers correspondent à la date de naissance. Les cinq autres n'ont aucun contenu démographique, mais leur validité peut être contrôlée au moyen du dernier « chiffre de vérification » du numéro d'identification, qui présente une relation mathématique avec les autres chiffres. Nous avons utilisé le chiffre de vérification intégré et la date de naissance pour déterminer si les numéros d'identification étaient valides. La date de naissance était incorrecte pour environ 9,5 % des demandeurs. Parmi les 3 661 demandeurs pour lesquels la date de naissance était plausible, 97 % ont déclaré un numéro d'identification valide.

Les caractéristiques démographiques des gagnants des chèques-éducation étaient comparables à celles des perdants, à part, peut-être, une légère différence d'âge. Ces observations sont illustrées par les colonnes 3 à 6 du tableau 1 qui montrent les différences selon la situation d'obtention d'un chèque. La différence d'âge selon la situation d'obtention d'un chèque semble être le résultat de quelques observations atypiques probablement dues au codage incorrect du numéro d'identification parmi les demandeurs perdants, puisque l'âge du demandeur est établi d'après le numéro d'identification. La différence d'âge est plus faible quand on limite l'échantillon aux demandeurs dont le numéro d'identification était valide, mais elle demeure marginalement significative. Par conséquent, nous tenons compte de l'effet de l'âge lors de l'estimation des effets des chèques.

Le premier résultat étudié est une variable nominale indiquant si le candidat au PACES s'est inscrit ou non à l'examen de l'ICFES. En plus de témoigner d'un intérêt pour des études plus poussées, l'inscription à l'examen de l'ICFES est une excellente approximation de la situation d'obtention du diplôme d'études secondaires, puisque 90 % des diplômés du secondaire s'inscrivent à l'examen de l'ICFES (Banque mondiale, 1993) et que tous ceux qui font l'examen doivent être en 11e année ou avoir obtenu leur diplôme d'études secondaires (principalement le premier cas). En revanche, l'évaluation incorrecte des numéros d'identification rend l'appariement plus difficile et complique l'interprétation des résultats. À cet égard, la possibilité que des numéros d'identification invalides de gagnants puissent avoir été corrigés au cours d'interactions bureaucratiques permanentes avec le programme PACES, alors que les numéros d'identification incorrects des perdants sont moins susceptibles d'avoir été corrigés, est un élément particulièrement préoccupant. Une différence de probabilité de correction selon la situation de gagnant ou de perdant pourrait donner lieu à un écart illusoire entre les taux d'appariement en faveur des gagnants du PACES.

En pratique, nos données fournissent peu de preuves qu'une différence d'exactitude du numéro d'identification selon la situation d'obtention du chèque soit un problème dans le cas de la cohorte de demandeurs de 1995 à Bogota, comme le montre le tableau 2, qui donne les différences selon la situation d'obtention du chèque et les estimations par probabilité linéaire de la relation entre la validité du numéro d'identification selon la situation de gagnant ou de perdant. En particulier, le tableau donne les estimations de α_0 dans le modèle de régression :

$$y_i = X_i' \beta_0 + \alpha_0 D_i + \epsilon_i, \quad (1)$$

où y_i est une variable nominale précisant la validité du numéro d'identification du demandeur i , D_i est un indicateur précisant si le demandeur i a gagné un chèque, et X_i est un vecteur de variables de contrôle pour l'âge et le sexe. Les erreurs-types présentées dans le tableau sont corrigées pour l'hétéroscédasticité.

La colonne 1 du tableau 2 montre qu'environ 88 % des candidats avaient un numéro d'identification valide. La probabilité qu'un gagnant du chèque ait un numéro d'identification valide était inférieure d'un point de pourcentage à celle observée pour les perdants, mais cette différence n'est pas significative. Limiter l'échantillon aux élèves dont la date de naissance intégrée dans le numéro d'identification était valide réduit encore davantage l'effet du chèque, que l'on introduise ou non des variables de contrôle démographiques dans le modèle. Les résultats sont similaires si l'on produit des estimations distinctes selon le sexe. Ces résultats donnent vivement à penser que les gagnants d'un chèque ne sont pas plus susceptibles que les perdants d'être appariés à un dossier de l'ICFES simplement parce que leur numéro d'identification est enregistré de façon plus exacte⁷.

3. EFFETS SUR L'OBTENTION DU DIPLÔME D'ÉTUDES SECONDAIRES

Nous avons produit les estimations des effets sur l'obtention du diplôme d'études secondaires au moyen d'une équation comparable à (1), en remplaçant la variable dépendante par un indicateur de l'inscription à l'examen de l'ICFES. Les élèves de la cohorte de demandeurs de 1995 qui ont été promus dans les délais prévus devraient s'être inscrits à l'examen de l'ICFES à l'une des deux occasions offertes durant l'année scolaire 2000. Comme certains élèves pourraient avoir sauté ou redoublé certaines années, nous avons recherché les demandeurs du chèque du PACES parmi l'ensemble des inscrits à tous les examens de l'ICFES offerts en 2000 et en 2001. Pour les élèves qui avaient passé l'examen plus d'une fois, nous avons retenu le premier ensemble de notes. En outre, nous avons expérimenté plusieurs stratégies d'appariement plus contraignantes utilisant la ville de résidence et le nom de l'élève aux fins de la validation⁸.

Dans l'ensemble, environ 35 % de numéros d'identification de demandeurs du PACES ont été appariés aux dossiers de l'ICFES. Ces résultats sont présentés au tableau 3, qui donne les résultats d'appariement pour l'échantillon de demandeurs dont le numéro d'identification était valide, tel que déterminé par le chiffre de vérification du numéro d'identification et les données valides sur l'âge intégrées dans le numéro d'identification. Une définition plus stricte, qui nécessite à la fois le numéro d'identification et la ville de résidence pour établir un appariement a donné un taux global d'appariement de 34 %, tandis que l'appariement sur les numéros d'identification et les sept premières lettres du nom de famille de l'élève a donné un taux d'appariement de 33 %. Enfin, l'appariement d'après la ville ainsi que les sept premières lettres du nom de famille donne un taux d'appariement de 32 %.

L'effet de l'obtention d'un chèque-éducation sur l'appariement d'après le numéro d'identification, sans variable de contrôle, qui est présenté dans la colonne 1, est d'environ 7 %. Cette proportion baisse pour s'établir à 6 % environ si l'on introduit des variables de contrôle démographiques, mais revient à presque 7 % quand on ajoute une variable nominale pour l'abonnement au téléphone. Les effets estimés sont un peu plus importants pour les garçons que pour

⁷ Nous avons obtenu certaines preuves de différences de tenue des dossiers pour la cohorte de demandeurs de Bogota de 1992, c'est-à-dire la toute première cohorte de Bogota. Pour cette raison et d'autres problèmes de données, nous avons exclu cette cohorte de l'étude.

⁸ Nous avons considéré qu'il y avait appariement concernant la ville quand un demandeur du PACES s'était inscrit à l'examen de l'ICFES à Bogota. Naturellement, en pratique, certains demandeurs de chèque-éducation de Bogota pourraient avoir passé l'examen ailleurs. Un petit nombre d'appariements apparents pour l'examen de 1999 sont inclus également. Voir les données en annexe pour plus de précision.

les filles, et le taux de base est plus faible pour les garçons. L'utilisation de la ville de résidence pour valider les appariements produit un effet de traitement légèrement plus faible pour les filles et dans l'ensemble, mais la variation n'est pas importante. La validation au moyen du nom produit un effet de traitement presque identique à celui observé sans validation, et la validation au moyen de la ville et du nom produit des estimations comparables à celles obtenues en utilisant la ville uniquement. Dans l'ensemble, les effets estimés de l'obtention d'un chèque-éducation sur l'inscription à l'examen de l'ICFES sont étonnamment robustes aux variations de l'échantillon et des spécifications. Donc, il semble raisonnable de déclarer que les chèques du PACES augmentent la probabilité d'inscription à l'examen de l'ICFES et, probablement, d'obtention du diplôme d'études secondaires de 5 à 7 points de pourcentage, sans qu'il existe de profil cohérent de différence entre les garçons et les filles.

Une augmentation de 5 à 7 points de pourcentage des niveaux correspond à une augmentation de 15 % à 20 % de la probabilité que les élèves s'inscrivent à l'examen de l'ICFES. Cet effet, qui semble important, confirme nos résultats antérieurs pour la cohorte de Bogota de 1995 selon lesquels la proportion d'élèves bénéficiaires d'un chèque-éducation terminant la 8e année en 1998 excédait de 10 points de pourcentage environ celle des élèves n'ayant pas obtenu le chèque. L'augmentation de 5 à 7 points de pourcentage du taux d'obtention du diplôme d'études secondaires parmi les gagnants du PACES pourrait tenir au fait que de la moitié aux trois quarts de ces élèves supplémentaires terminant la 8e année aient poursuivi leurs études et terminé l'école secondaire trois ans plus tard.

4. EFFETS SUR LES NOTES À L'EXAMEN D'ENTRÉE AU COLLÈGE

4.1 Le problème de sélection

À la présente section, nous évaluons les effets des chèques du PACES sur le rendement scolaire tel que mesuré par les notes à l'examen de l'ICFES. La principale difficulté économétrique que pose une évaluation des effets sur les notes tient au fait que les gagnants du chèque étaient plus susceptibles de s'inscrire à l'examen que les perdants. Donc, les notes des gagnants et des perdants ne sont plus comparables. Pour examiner les conséquences probables de cette sélection différentielle sur les comparaisons des notes des participants à l'examen, représentons par Y_{1i} la note d'examen qu'obtiendrait l'élève i après avoir obtenu le chèque-éducation et par Y_{0i} , la note d'examen qu'obtiendrait l'élève i autrement. Nous supposons que ces deux résultats possibles sont bien définis pour tous les élèves, que ceux-ci aient effectivement passé ou non l'examen. L'effet causal moyen de l'obtention du chèque-éducation sur les notes pour l'ensemble des gagnants est donné par $E[Y_{1i} - Y_{0i} | D_i = 1]$. Naturellement, en pratique, nous n'observons que les notes des élèves qui ont fait l'examen. En outre, parmi ces derniers, nous n'observons que Y_{1i} pour les gagnants et Y_{0i} pour les perdants.

Si nous posons que la cote des élèves qui ne font pas l'examen est nulle, alors nous pouvons représenter la comparaison entre les notes des gagnants et des perdants parmi les élèves qui ont fait l'examen sous la forme

$$\begin{aligned} E[Y_i | Y_i > 0, D_i = 1] - E[Y_i | Y_i > 0, D_i = 0] &= E[Y_{1i} | Y_{1i} > 0, D_i = 1] - E[Y_{0i} | Y_{0i} > 0, D_i = 0] \\ &= E[Y_{1i} | Y_{1i} > 0] - E[Y_{0i} | Y_{0i} > 0], \end{aligned}$$

où la deuxième égalité est due au fait que D_i est assigné aléatoirement. Cependant, cette différence n'as pas d'interprétation causale, puisque les élèves pour lesquels $Y_{1i} > 0$ et $Y_{0i} > 0$ ne sont pas tirés de la même population. En fait, nous pouvons développer davantage cette expression pour obtenir

$$\begin{aligned} E[Y_i | Y_i > 0, D_i = 1] - E[Y_i | Y_i > 0, D_i = 0] &= E[Y_{1i} - Y_{0i} | Y_{0i} > 0] \\ &\quad + \{E[Y_{1i} | Y_{1i} > 0] - E[Y_{1i} | Y_{0i} > 0]\} \end{aligned} \quad (2)$$

Donc, la différence entre les gagnants et les perdants parmi les élèves inscrits à l'examen est égale à l'effet causal moyen sur ceux qui auraient passé l'examen de toute façon, $E[Y_{1i} - Y_{0i} | Y_{0i} > 0]$, auquel s'ajoute un terme reflétant le biais de sélection dû au conditionnement sur un résultat endogène d'inscription à l'examen. Dans l'équation (2), le biais est vraisemblablement négatif si les chèques font augmenter la note à l'examen. Supposons, par exemple, que $Y_{1i} = Y_{0i} + \alpha$, où $\alpha > 0$. Alors, le biais de sélection est

$$E[Y_{1i} | Y_{1i} > 0] - E[Y_{1i} | Y_{1i} > \alpha],$$

qui est clairement négatif.

Nous appliquons plusieurs stratégies pour essayer d'atténuer le biais de sélection inhérent aux comparaisons conditionnellement à l'inscription à l'examen. Une méthode simple consiste à traiter les élèves qui ne s'inscrivent pas à l'examen comme si leur note était nulle ou égale à toute autre valeur faible, disons une note correspondant au quantile 0,01 ou 0,1, puis à inclure ces valeurs dans l'analyse de l'échantillon complet. Cette méthode permet d'éviter le biais de sélection, mais peut donner lieu à des conclusions erronées si certains élèves qui ne s'inscrivent pas à l'examen auraient obtenu de meilleurs résultats. Nous pouvons aussi supposer que les observations sur les élèves non inscrits à l'examen représentent la réalisation censurée de la distribution normale sous-jacente des notes. Nous obtenons ainsi des estimations tobit. Pour rendre cette approche plus plausible et pour tester certaines hypothèses sous-jacentes, nous essayons divers points de censure.

Une deuxième méthode de résolution du problème de sélection exploite le fait que, puisque le biais de sélection est le plus probablement négatif, l'analyse des résultats conditionnellement à l'inscription à l'examen est prudente. Autrement dit, les estimations contaminées par la sélection fournissent une borne inférieure de l'effet probable des chèques sur le rendement scolaire. De surcroît, nous pouvons obtenir une borne supérieure des effets sur les notes, valide sous des hypothèses non paramétriques faibles, en adaptant un résultat théorique de notre étude antérieure (Angrist et coll., 2002). Plus précisément, supposons que les chèques n'ont jamais d'effet négatif, c.-à-d. que $Y_{1i} \geq Y_{0i}$ pour tout i . Cette hypothèse semble raisonnable, puisque les gagnants du chèque sont, naturellement, libres de ne pas utiliser ou de cesser d'utiliser les chèques s'ils perçoivent des effets indésirables. Si nous abandonnons l'indice inférieur i pour simplifier la notation et que nous représentons par $q_0(\theta)$ le quantile θ de la distribution de Y_0 et par $q_1(\theta)$ le quantile θ de la distribution de Y_1 , alors, pour tout quantile, nous obtenons le résultat suivant sur les effets du traitement en cas de biais de sélection de l'échantillon :

Proposition 1. Supposons que $Y_1 > Y_0$. Alors

$$\begin{aligned} E[Y | D=1, Y > q_1(\theta)] - E[Y | D=0, Y > q_0(\theta)] &\geq E[Y_1 - Y_0 | Y_0 > q_0(\theta)] \\ &\geq E[Y | D=1, Y > q_0(\theta)] - E[Y | D=0, Y > q_0(\theta)] \end{aligned}$$

En outre, si $Y_1 = h(Y_0)$ pour toute fonction faiblement croissante $h(\bullet)$, alors l'inégalité de gauche est une égalité. Preuve. $E[Y | D=1, Y > q_1(\theta)] = E[Y_1 | Y_1 > q_1(\theta)]$ et $E[Y | D=0, Y > q_0(\theta)] = E[Y_0 | Y_0 > q_0(\theta)]$ par assignation aléatoire. De surcroît,

$$\begin{aligned} E[Y_1 | Y_1 > q_1(\theta)] - E[Y_0 | Y_0 > q_0(\theta)] &= E[Y_1 - Y_0 | Y_0 > q_0(\theta)] + \{E[Y_1 | Y_1 > q_1(\theta)] - E[Y_1 | Y_0 > q_0(\theta)]\} \\ &\equiv E[Y_1 - Y_0 | Y_0 > q_0(\theta)] + \tau_0 \end{aligned}$$

Si $Y_1 = h(Y_0)$, h préserve le rang et $\tau_0=0$, si bien que la deuxième partie est prouvée. Sinon, nous devons montrer que $\tau_0 > 0$. Notons que

$$\tau_0 = E[Y_1 1(Y_1 > q_1) - Y_1 1(Y_0 > q_0)] / P(Y_0 > q_0) \text{ puisque } P(Y_1 > q_1) = P(Y_0 > q_0),$$

$$\tau_0 \geq E[Y_1 (1(Y_1 > q_1) - 1(Y_0 > q_0))] = E[Y_1 | Y_1 > q_1, Y_0 < q_0] p_1 - E[Y_1 | Y_1 < q_1, Y_0 > q_0] p_0$$

où $p_1 \equiv \Pr[Y_1 > q_1, Y_0 < q_0]$ et $p_0 \equiv \Pr[Y_1 < q_1, Y_0 > q_0]$. Manifestement, $E[Y_1 | Y_1 > q_1, Y_0 < q_0] \geq E[Y_1 | Y_1 < q_1, Y_0 > q_0]$. En outre, $p_1 = p_0$ car

$$\Pr[Y_1 > q_1, Y_0 < q_0] + \Pr[Y_1 > q_1, Y_0 > q_0] = \theta = \Pr[Y_1 > q_1, Y_0 < q_0] + (\theta - \Pr[Y_1 < q_1, Y_0 > q_0]).$$

Ceci établit la borne supérieure. La borne inférieure découle du fait que

$$E[Y | D=1, Y > q_0(\theta)] - E[Y | D=0, Y > q_0(\theta)] = E[Y_1 - Y_0 | Y_0 > q_0(\theta)] + \{E[Y_1 | Y_1 > q_0(\theta)] - E[Y_1 | Y_0 > q_0(\theta)]\}$$

et $E[Y_1 | Y_0 > q_0(\theta)] = E[Y_1 | Y_1 \geq Y_0 > q_0(\theta)] \geq E[Y_1 | Y_1 > q_0(\theta)]$.

Notons que nous pouvons choisir un quantile, θ_0 , tel que $q_0(\theta_0) = 0$, puis éliminer le quantile inférieur θ_0 de la distribution de Y_1 pour obtenir une borne supérieure sur $E[Y_1 - Y_0 | Y_0 > 0]$. Parallèlement, la différence entre les notes, conditionnellement à une valeur positive de l'effet, fournit une borne inférieure. En outre, si Y_1 et Y_0 sont liés de façon déterministe, la borne supérieure donne une estimation de $E[Y_1 | Y_0 > 0]$ ⁹. Nous exploitons cette stratégie simple pour estimer les bornes des effets du traitement en plusieurs points de la distribution des notes.

4.2 Estimations

Étant donné le taux élevé de validation des appariements fondés sur le numéro d'identification, il semble raisonnable de se concentrer sur les notes d'examen pour l'ensemble des élèves pour lesquels un appariement d'après le numéro d'identification a été obtenu. Parmi les élèves qui ont fait l'examen, la note moyenne en langue était d'environ 47 points avec un écart-type de 5,6. L'effet sur la note en langue des élèves qui ont fait l'examen est de 0,7 avec une erreur-type de 0,33. Ces résultats figurent dans la première colonne du tableau 4a, qui donne aussi les effets estimés du traitement calculés en utilisant des résultats winsorisés et des corrections tobit par censure. Donner une note nulle aux élèves qui n'ont pas fait l'examen et inclure ces notes dans l'analyse, de sorte qu'il n'y ait aucun biais de sélection d'échantillon, donne un effet global de traitement de 3,1, avec une erreur-type de 0,69. Toutefois, ces résultats sont dus principalement à la probabilité plus forte de faire l'examen chez les gagnants du chèque-éducation. Une variante de cette approche plus facile à motiver consiste à ramener les notes du 1 % inférieur de la distribution jusqu'au premier percentile d'élèves qui ont fait l'examen et d'ajouter ceux qui n'ont pas fait l'examen à ce groupe. Cela revient à supposer que les élèves qui ne font pas l'examen obtiendraient une note comprise dans le premier percentile et à analyser les effets sur le résultat winsorisé. Ceci donne un effet de traitement de 1,1, avec une erreur-type de 0,24, qui est présenté à la troisième colonne.

Les quatrième à sixième colonnes donnent les coefficients tobit estimés sous diverses hypothèses. Ici, un modèle tobit type revient à supposer que les élèves qui ne font pas l'examen obtiendraient une note inférieure ou égale à 0 s'ils faisaient l'examen; naturellement, cette hypothèse engendre un effet de traitement implausible de 9,8. Pour rendre l'hypothèse tobit plus plausible, nous censurons artificiellement les données à des points supérieurs à zéro. Pour montrer pourquoi il est valide de procéder ainsi, supposons que le résultat observé est généré par censure de type tobit au-delà d'un certain seuil positif, c :

$$y_i(c) = 1[y_i^* > c]y_i^* ;$$

$$y_i^* = X_i' \beta + \alpha D_i + \eta_i ,$$
(3)

où η_i est une erreur normalement distribuée. Alors, il est également vrai que

$$y_i(c + \delta) \equiv 1[y_i(c) > c + \delta]y_i(c) = 1[y_i^* > c + \delta]y_i^* ,$$
(4)

de sorte que les estimations tobit de (3) pour toute valeur positive de δ sont convergentes pour α à condition que c soit choisi de sorte que les élèves n'ayant pas fait l'examen obtiennent une note inférieure à ce seuil s'ils faisaient l'examen. Bien que le choix de $c=0$ ne marchera vraisemblablement pas, censurer à un quantile plus élevé semble plausible. En outre, les estimations de α peuvent être comparées pour divers choix de la valeur de δ à titre de test de suridentification pour les hypothèses combinées de sélection de type tobit et d'erreurs latentes suivant la loi normale.

Déplacer le point de censure jusqu'au premier percentile, puis jusqu'au 10e percentile chez les élèves qui font l'examen, c.-à-d. supposer que ceux qui ne font pas l'examen seraient situés dans ces parties de la distribution des notes s'ils le faisaient, donne des estimations significatives de l'ordre de 2 à 4 points. Les effets sur les notes de

⁹ Krueger et Whitmore (2001) utilisent cette idée pour estimer $E[Y_1 - Y_0 | Y_0 > 0]$ dans une étude de la taille des classes. Voir aussi Lee (2002).

mathématiques, présentés au tableau 4b, sont plus faibles que ceux sur les notes en langue, bien qu'ils soient encore significatifs après censure au premier percentile et au 10e percentile. Une analyse selon le sexe produit des estimations un peu plus importantes pour les garçons que pour les filles.

Les estimations tobit de α en fixant c au premier et au 10e percentiles de la distribution des notes chez les élèves qui passent l'examen sont présentées aux cinquième et sixième colonnes. Les estimations tobit de α sont égales à 3,3 et 2,1, valeurs toutes deux raisonnablement précises. Si l'on suppose que le modèle tobit s'applique avec censure au premier percentile, les estimations devraient être les mêmes au 10e percentile. La diminution des estimations donne à penser que la censure au premier percentile donne un seuil trop faible pour les élèves qui ne s'inscrivent pas à l'examen. Par contre, les estimations tobit de α sont étonnamment stables lorsque la distribution est censurée artificiellement en choisissant un seuil qui élimine les 20 % à 80 % inférieurs des notes, ce qu'illustrent les figures 1a et 1b, qui représentent graphiquement l'estimation des coefficients tobit et des bandes de confiance pour divers choix de la valeur de c . Les effets estimés du traitement varient de 1 à 2 points, baissant légèrement lorsqu'on censure les 90 % inférieurs des notes chez les élèves qui s'inscrivent à l'examen. Il convient toutefois de souligner que les intervalles de confiance s'élargissent à ce point. En outre, l'hypothèse d'une distribution normale des notes pourrait être une moins bonne approximation pour la queue supérieure. Dans l'ensemble, le modèle résumé dans (2) et (3) semble refléter de façon raisonnablement convergente l'effet des chèques-éducation sur les notes latentes. Les effets du traitement de l'ordre de 1 à 2 équivalent à une augmentation de la note de 0,15 à 0,33 écart-type dans la distribution des élèves qui s'inscrivent à l'examen. De nouveau, ces résultats confirment nos estimations antérieures des effets du traitement sur le rendement scolaire pour un échantillon aléatoire d'élèves de 8e année de Bogota en 1998.

Estimations non paramétriques

La comparaison de la distribution complète des notes d'examen pour les gagnants des chèques-éducation et pour les perdants soutient l'idée selon laquelle l'obtention du chèque se traduit par une augmentation du rendement scolaire chez les gagnants. Les figures 2a et 2b, qui donnent le tracé des estimations de la densité par la méthode du noyau dans l'échantillon de l'ensemble des élèves qui se sont inscrits à l'examen, indiquent que la distribution est légèrement aplatie et décalée vers la droite pour les gagnants. Toutefois, comme pour les comparaisons de moyennes, cette différence est contaminée par le biais de sélection, en particulier l'introduction probable d'élèves obtenant une note faible dans l'échantillon des gagnants du chèque qui ont fait l'examen. La correction pour le biais de sélection de l'échantillon correspondant à la proposition 1 donne une nettement meilleure impression du déplacement. Celui-ci peut être observé aux figures 3a et 3b, qui donnent le tracé des distributions des notes après limitation de la distribution des gagnants du chèque aux 28 % supérieurs de la distribution des notes (y compris les valeurs nulles), c'est-à-dire la même proportion de la population que celle ayant fait l'examen parmi les élèves n'ayant pas obtenu le chèque. Donc, la figure 3 représente les notes conditionnellement à $Y_0 > q_0(0,28)$, où $q_0(0,28) = 0$ et $Y_1 > q_1(0,28)$. Le chiffre corrigé montre un déplacement net vers la droite de la distribution pour les gagnants.

Les bornes non paramétriques sur les effets du traitement pour divers quantiles, c.-à-d. les bornes estimées pour $E[Y_1 - Y_0 | Y_0 > q_0(\theta)]$, sont présentées aux tableaux 5a et 5b pour les notes en langue et en mathématiques, pour θ_0 tel que $q_0(\theta_0) = 0$ et pour $\theta = 0,75, 0,85$ et $0,95$. Les effets les plus importants sont ceux observés pour la valeur nulle, c.-à-d. les effets sur tous les élèves qui se seraient inscrits à l'examen même s'ils n'avaient pas gagné au tirage au sort. La borne inférieure pour les effets sur les notes en langue dans cette population est égale à 0,68 (e.-t.=0,33), tandis que la borne supérieure est égale à 2,8 (e.-t.=0,31). Les effets tombent à une borne inférieure non significative de 0,35 et une borne supérieure de 1,4 (e.-t.=0,34) pour $\theta = 0,95$. Donc, le profil des estimations par quantile laisse entendre que le décalage des notes est plus important pour les élèves pour lesquels Y_0 est proche de l'extrémité inférieure de la distribution des notes des élèves qui s'inscrivent à l'examen que chez ceux se trouvant dans la queue supérieure de cette distribution. Les estimations tobit avec censure comprises entre 0,2 et 0,8 sont plus proches de la borne supérieure que de la borne inférieure, ce qui donne à penser que ces estimations pourraient être plus exactes.

Le calcul d'estimations distinctes pour les garçons et les filles ne donne aucune preuve d'une variation marquée des effets du traitement selon le sexe. Cependant, les effets estimés conditionnellement à $Y_0 > 0$ sont plus importants pour les garçons que pour les filles. Enfin, les effets sur la note en mathématiques ont tendance à être un peu plus faibles que sur la note en langue. Les bornes inférieures pour les effets estimés sur la note en mathématiques ne diffèrent pas significativement de zéro, mais, étant donné leur similarité avec les estimations corrigées pour le biais de sélection calculées par d'autres méthodes, les bornes supérieures semblent plus exactes. Un renversement intéressant

comparativement à nos résultats paramétriques antérieurs est l'obtention de bornes supérieures légèrement plus élevées pour les effets sur la note de mathématiques comparativement aux effets sur la note en langue, lorsqu'on observe la queue supérieure de la distribution des notes.

La dernière méthode non paramétrique utilisée pour estimer les effets du traitement sur les notes d'examen sans biais de sélection est une régression par quantile. Si l'on suppose que les élèves non inscrits à l'examen obtiendraient une note inférieure au quantile choisi, les écarts entre les quantiles conditionnellement à la situation d'obtention du chèque (gagnant ou perdant) ne sont pas modifiés par le biais de sélection et ont une interprétation causale. Les estimations résultantes, présentées au tableau 6, montrent un effet du traitement de 5,5 points (e.-t.=0,67) sur la note en langue et de 4,4 points (e.-t.=0,27) sur la note en mathématiques au 75e quantile, et un effet beaucoup plus faible, mais encore significatif, au 85e quantile. L'effet sur le 95e quantile n'est significatif que pour la note en langue dans l'échantillon regroupé, mais les effets sur la queue supérieure de la distribution sont plus significatifs pour les deux notes pour l'échantillon de garçons. Néanmoins, dans l'ensemble, les résultats de la régression par quantile soutiennent aussi l'idée selon laquelle le déplacement de la distribution est concentré à la partie inférieure de la distribution des notes pour les élèves qui s'inscrivent à l'examen.

5. SOMMAIRE ET CONCLUSIONS

Dans le présent article, nous donnons des preuves de l'effet des chèques-éducation sur les résultats à long terme des élèves qui ont soumis une demande au programme PACES à Bogota. L'étude du PACES présente un intérêt particulier, parce que les établissements d'enseignement secondaire privés représentent plus de la moitié des élèves du secondaire de Bogota et que l'assignation aléatoire des chèques-éducation facilite la comparaison des relations de cause à effet. Les données administratives sur l'examen d'entrée au collège nous permettent d'estimer l'effet des chèques-éducation sur le taux d'obtention du diplôme d'études secondaires et sur le rendement scolaire. Deux problèmes empiriques que pose une étude de ce genre sont le taux élevé d'appariement pour les élèves qui obtiennent le chèque-éducation et le biais de sélection dans les comparaisons entre les élèves qui ont fait l'examen. Nous montrons que la qualité de la tenue des dossiers semble être la même pour les élèves qui ont obtenu et ceux qui n'ont pas obtenu le chèque-éducation, si bien que les faux appariements ne posent probablement pas de problème. Nous utilisons diverses stratégies simples pour corriger le biais de sélection.

Tous les résultats indiquent que le taux d'obtention du diplôme d'études secondaires augmente de 5 à 7 points de pourcentage comparativement à un taux de base de 25 % à 30 %, ce qui corrobore nos résultats antérieurs montrant une augmentation de l'ordre de 10 points de pourcentage du taux d'élèves qui achèvent la 8e année parmi ceux qui ont gagné un chèque-éducation. L'importance de l'augmentation estimée des notes d'examen dépend en partie de la méthode utilisée pour tenir compte du biais de sélection. Les estimations tobit avec censure artificielle des données donnent un effet de traitement de 1 à 2 points, soit un gain approximatif de 0,15 à 0,33 écart-type dans l'échantillon des élèves qui ont passé l'examen. Des bornes non paramétriques encadrent ce nombre, la borne inférieure étant significativement différente de zéro pour les effets dans l'échantillon complet, sous la condition d'une valeur positive. Puisque la borne supérieure est exacte sous l'hypothèse d'un décalage monotone et que le profil des estimations tobit avec censure artificielle est convergent, un effet de traitement approchant de 2 points semble probable. Pour les quantiles plus élevés de la distribution des notes et pour les notes de mathématiques, les effets sont, dans la plupart des cas, un peu plus faibles.

Toutes les méthodes d'estimation utilisées ici semblent indiquer que le programme de chèques-éducation donne lieu à une augmentation considérable du taux d'élèves qui obtiennent un diplôme d'études secondaires, ainsi que du rendement scolaire. L'importance et la persistance de cet effet portent à conclure que le PACES était un programme rentable. Bien qu'il soit difficile de quantifier les avantages de l'augmentation du rendement scolaire proprement dit, le rendement économique de l'obtention du diplôme d'études secondaires est considérable. Au minimum, les résultats laissent entendre que les efforts de financement de la demande de services éducatifs par des moyens comparables au programme PACES méritent une étude approfondie.

REMERCIEMENTS

*Nous remercions tout spécialement Cristina Estrada, Claudia Gonzalez, Marcela Monsalvo et Ana Gomez de leur aide durant les travaux de recherche. Nous sommes également reconnaissants à Jorge Estrada de son assistance lors de l'interprétation des numéros d'identification colombiens et aux employés de l'ICFES de nous avoir fourni les données. Nous remercions les National Institutes of Health et la Banque mondiale d'avoir financé l'étude.

ANNEXE - DESCRIPTION DES DONNÉES

Précisions sur l'appariement aux dossiers de l'ICFES

La population à appairer comprend les dossiers de 4 044 demandeurs de chèques-éducation du PACES faisant partie de la cohorte de 1995 à Bogota. Pour exécuter l'appariement, on a procédé à une recherche sur les fichiers informatisés dans les bureaux de l'ICFES de Bogota. La recherche a été réalisée par nos assistants de recherche avec l'aide d'employés de l'ICFES. Les examens de l'ICFES ont lieu deux fois par an. Nous avons recherché les appariements parmi tous les élèves inscrits à l'examen en 1999, en 2000 et en 2001. Un appariement est défini comme étant un appariement exact des numéros d'identification. En outre, pour certaines analyses dont les résultats sont présentés au tableau 3, nous avons validé les appariements en déterminant si l'examen associé à un appariement probable avait eu lieu à Bogota et en appariant les sept premières lettres du nom de famille du candidat.

Dans l'échantillon contenant les appariements exacts du numéro d'identification, 27 élèves se sont inscrits à l'examen mais ne l'ont jamais passé. Nous codons cet événement comme un appariement, mais nous lui attribuons une note nulle. En outre, les dossiers de 44 inscrits contenaient des notes d'examen passé en 1999, avant que l'ICFES soit remanié et qu'une nouvelle échelle de notation soit adoptée. Nous avons attribué à ces sujets la note moyenne calculée pour leur cellule âge-sexe parmi ceux qui ont passé l'examen en 2000 et en 2001. Les résultats sont comparables lorsqu'on traite les appariements de 1999 comme des non-appariements.

Matières de l'examen de l'ICFES et distribution des notes

Les notes de l'ICFES utilisées ici sont issues du système de notation remanié qui est entré en vigueur en mars 2000. Les examens sont ceux du tronc commun de compétences élémentaires (Nucleo Comun Competencias Basicas), qui inclut des modules de biologie, de chimie, de physique, de mathématiques, de langue, d'histoire et de géographie, ainsi qu'un examen portant sur une langue étrangère choisie par l'élève. En plus de ces matières obligatoires, les élèves choisissent d'être évalués dans des domaines de spécialisation dont le nombre peut aller jusqu'à quatre. Les notes que nous analysons sont celles des volets de mathématiques et de langue du tronc commun. L'examen de l'ICFES se déroule sur une période de deux jours comportant deux séances en matinée et une séance l'après-midi le premier jour.

Les volets de mathématique et de langue du tronc commun prennent chacun une heure et sont notés sur un total de 35 points. Les notes d'examen sont diffusées sur une échelle de 0 à 100 et la distribution des notes est fortement concentrée dans l'intervalle de 30 à 70. Les distributions des notes en mathématiques et en langue pour l'ensemble des élèves ayant passé l'examen à Bogota en mars 2000 sont présentées dans les figures en annexe (pour 6 868 participants à l'examen).

RÉFÉRENCES

- Angrist, Joshua. "Conditional Independence in Sample Selection Models." *Economics Letters*, February 1997, 54(2), 103-112.
- Angrist, Joshua; Bettinger, Eric; Bloom, Erik; King, Elizabeth and Kremer, Michael, "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment," *American Economic Review*, forthcoming 2002.
- Behrman, Jere; Sengupta, Piyali and Todd, Petra. "Progressing through PROGRESA: An Impact Assessment of Mexico's School Subsidy Experiment." Draft 2000.
- Bettinger, Eric. "Do Private School Vouchers Affect Test Scores and Why? Evidence from a Private School Scholarship Program." Case Western Reserve, Department of Economics, January 2001a, mimeo.
- Bettinger, Eric. "The Effect of Charter Schools on Charter Students and Public Schools." Case Western Reserve, Department of Economics, March 2001b, mimeo.
- Chamberlain, Gary. "Asymptotic Efficiency in Semi-Parametric Models with Censoring." *Journal of Econometrics*, July 1986, 32(2), 189-218.
- DNP, *Sistema de Indicadores Sociodemograficos para Colombia (SISD) 1980-1997* Boletin No. 21, p. 58, Bogota : Departamento Nacional de Planeacion, June 1999.
- Dynarski, Susan M. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion , " John F. Kennedy School of Government, Faculty Research Working Paper RWP01-034, September 2001.
- Evans, William N. and Schwab, Robert M. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics*, November 1995, 110(4), 941-974.
- Howell, William G.; Wolf, Patrick, J.; Peterson, Paul E. and Campbell, David E. "Test-Score Effects of School Vouchers in Dayton, New York, and Washington: Evidence from Randomized Field Trials." Paper presented at the annual meeting of the American Political Science Association, Washington, D.C., September 2000.
- King, Elizabeth; Orazem, Peter and Wolgemuth, Darin. "Central Mandates and Local Incentives: The Colombia Education Voucher Program." Working Paper No. 6, Series on Impact Evaluation of Education Reforms, Development Economics Research Group, The World Bank, February 1998.
- King, Elizabeth; Rawlings, Laura; Gutierrez, Marybell; Pardo, Carlos and Torres, Carlos. "Colombia's Targeted Education Voucher Program: Features, Coverage and Participation." Working Paper No. 3, Series on Impact Evaluation of Education Reforms, Development Economics Research Group, The World Bank, September 1997.
- Krueger, Alan and Whitmore, Diane. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *Economic Journal*, January 2001, 111(468).
- Lee, David, "Trimming for Bounds on Treatment Effects with Missing Outcomes," NBER Technical Working Paper 277, June 2002.
- Patrinos, Harry A. and Ariasingham, David L. *Decentralization of Education: Demand-Side Financing*. Washington, DC: The World Bank, 1997.

- Psacharopoulos, George; Tan, J., and Jimenez, E. *Financing Education in Developing Countries: An Exploration of Policy Options*. Washington, DC: The World Bank, 1986.
- Psacharopoulos, George, and Velez, Eduardo. "Education Quality and Labor Market Outcomes: Evidence from Bogota, Colombia." *Sociology of Education*, April 1993, 66(3), 130-145.
- Ribero, Rocío and Tenjo, Jaime. University de los Andes, Department of Economics Working Paper, 1997.
- Rouse, Cecilia Elena. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics*, May 1998, 13(2), 553-602.
- US Department of Education, Office of Education Research and Improvement. *The Condition of Education 1998*, NCES 98-013, Washington, DC: USGPO, 1998.
- The World Bank, Research Department, *Staff Appraisal Report: Colombia, Secondary Education Project*, Human Resources Operations Division, Latin America and the Caribbean Region, Report No. 11834-CO, November 19, 1993.
- The World Bank, *World Development Report 1998/99*. New York: Oxford University Press, 1999.

Tableau 1. Caractéristiques de l'échantillon de dossiers appariés de l'ICFES selon la situation d'obtention du chèque-éducation

	Moyenne		Différence selon la situation d'obtention du chèque			
	Échantillon complet	Échantillon avec âge valide	Échantillon complet	Échantillon avec âge valide	N° d'ID et âge valides	N° d'ID et âge valides et a le téléphone
	(1)	(2)	(3)	(4)	(5)	(6)
A gagné un chèque-éducation	0,588	0,585				
N° d'ID valide	0,876	0,967				
Âge au moment de la demande	12,7 (1,8)	12,7 (1,3)	-0,137 (0,064)	-0,086 (0,045)	-0,085 (0,044)	-0,091 (0,047)
Sexe masculin	0,487	0,493	0,004 (0,016)	0,011 (0,017)	0,012 (0,017)	0,008 (0,018)
Téléphone	0,882	0,886	0,013 (0,010)	0,008 (0,011)	0,008 (0,011)	---
N	4 044	3 661	4 044	3 661	3 542	3 139

Nota : Les erreurs-types robustes sont indiquées entre parenthèses. Les estimations par régression des différences selon le statut d'obtention du chèque présentées à la colonne 4 sont celles obtenues pour l'échantillon avec données sur l'âge valides intégrées dans le numéro d'identification national. À la colonne 5, l'échantillon est limité aux cas pour lesquels le chiffre de vérification du numéro d'identification est valide et à la colonne 6, l'échantillon est en outre limité aux cas ayant le téléphone. Pour l'« âge au moment de la demande », la colonne 1 comporte 1 520 observations et la colonne 2 en compte 3 664. Les autres tailles d'échantillon sont présentées dans le tableau.

Tableau 2. Effet de la situation d'obtention du chèque-éducation sur la probabilité d'avoir un numéro d'identification valide

	Échantillon complet	Échantillon avec données sur l'âge valides		Échantillon avec âge valide et téléphone	
	(1)	(2)	(3)	(4)	(5)
A. Tous les demandeurs					
Variable dépendante – Moyenne	0,876	0,968		0,968	
Gagnant d'un chèque	-0,010 (0,010)	0,001 (0,006)	-0,0002 (0,0060)	0,001 (0,006)	0,0001 (0,0063)
Âge			-0,010 (0,002)		-0,008 (0,002)
Téléphone			-0,0003 (0,0092)		
N	4 044	3 661	3 661	3 244	3 244
B. Filles					
Variable dépendante – Moyenne	0,862	0,963		0,971	
Gagnante d'un chèque	-0,023 (0,015)	-0,001 (0,009)	-0,002 (0,009)	-0,002 (0,010)	-0,003 (0,009)
Âge			-0,010 (0,003)		-0,010 (0,003)
Téléphone			0,003 (0,013)		
N	2 076	1 857	1 857	1 631	1 631
C. Garçons					
Variable dépendante – Moyenne	0,891	0,971		0,971	
Gagnant d'un chèque	0,004 (0,014)	0,003 (0,008)	0,002 (0,008)	0,004 (0,008)	0,003 (0,008)
Âge			-0,011 (0,003)		-0,006 (0,003)
Téléphone			0,001 (0,013)		
N	1 968	1 804	1 804	1 613	1 613

Nota : Le tableau donne les estimations de l'équation (1) qui figurent dans le texte. Les erreurs-types robustes sont indiquées entre parenthèses.

Tableau 3. Situation d'obtention du chèque et probabilité d'un appariement avec les dossiers de l'ICFES

	Appariement exact du n° d'ID			Appariement du n° d'ID et de la ville			Appariement du n° d'ID et de 7 lettres du nom			Appariement du n° d'ID, de la ville et de 7 lettres du nom		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Tous les demandeurs												
Variable dépendante - Moyenne	0,354		0,352	0,339	0,337	0,331	0,330	0,318	0,317			
Gagnant d'un chèque	0,072 (0,016)	0,059 (0,015)	0,066 (0,015)	0,069 (0,016)	0,056 (0,014)	0,063 (0,015)	0,072 (0,016)	0,059 (0,014)	0,067 (0,015)	0,068 (0,016)	0,056 (0,014)	0,064 (0,015)
Sexe masculin		-0,052 (0,014)	-0,046 (0,015)		-0,053 (0,014)	-0,045 (0,002)		-0,043 (0,014)	-0,038 (0,015)		-0,045 (0,014)	-0,039 (0,015)
Âge		-0,160 (0,005)	-0,162 (0,005)		-0,156 (0,005)	-0,158 (0,002)		-0,153 (0,005)	-0,154 (0,005)		-0,149 (0,005)	-0,150 (0,005)
N	3 542	3 542	3 139	3 542	3 542	3 139	3 542	3 542	3 139	3 542	3 542	3 139
B. Filles												
Variable dépendante - Moyenne	0,387		0,383	0,372	0,368	0,361	0,357	0,348	0,344			
Gagnante d'un chèque	0,067 (0,023)	0,056 (0,021)	0,060 (0,022)	0,069 (0,023)	0,057 (0,021)	0,060 (0,022)	0,071 (0,023)	0,060 (0,021)	0,070 (0,022)	0,073 (0,023)	0,062 (0,021)	0,070 (0,022)
Âge		-0,168 (0,006)	-0,169 (0,007)		-0,164 (0,006)	-0,165 (0,007)		-0,160 (0,006)	-0,161 (0,007)		-0,156 (0,006)	-0,156 (0,007)
N	1 789	1 789	1 571	1 789	1 789	1 571	1 789	1 789	1 571	1 789	1 789	1 571
C. Garçons												
Variable dépendante - Moyenne	0,320		0,321	0,304	0,306	0,302	0,304	0,288	0,290			
Gagnant d'un chèque	0,079 (0,022)	0,063 (0,020)	0,073 (0,021)	0,071 (0,022)	0,055 (0,020)	0,066 (0,021)	0,074 (0,022)	0,059 (0,020)	0,065 (0,021)	0,065 (0,022)	0,050 (0,020)	0,058 (0,021)
Âge		-0,153 (0,007)	-0,154 (0,007)		-0,148 (0,007)	-0,150 (0,007)		-0,146 (0,007)	-0,147 (0,007)		-0,141 (0,006)	-0,143 (0,007)
N	1 752	1 752	1 568	1 752	1 752	1 568	1 753	1 753	1 568	1 753	1 568	1 568

Nota : Les erreurs-types robustes sont indiquées entre parenthèses. L'échantillon inclut les 95 demandeurs de Bogota pour lesquels le numéro d'identification et les données sur l'âge étaient valides. L'échantillon est le même que pour la colonne 5 du tableau 1.

Tableau 4a. Estimations par régression des effets des chèques-éducation sur les notes en **langue**

	Moindres carrés ordinaires			Tobit		
	MCO avec note>0	MCO censuré à 0	MCO censuré à 1 %	Tobit censuré à 0	Tobit censuré à 1 %	Tobit censuré à 10 %
	(1)	(2)	(3)	(4)	(5)	(6)
A. Tous les demandeurs						
Variable dépendante - Moyenne	47,4 (5,6)	16,4 (22,8)	37,3 (8,0)	16,4 (22,8)	37,3 (8,0)	42,7 (4,7)
Gagnant d'un chèque	0,70 (0,33)	3,07 (0,69)	1,14 (0,24)	8,81 (1,99)	3,29 (0,70)	2,06 (0,46)
Sexe masculin	0,73 (0,32)	-1,90 (0,69)	-0,44 (0,24)	-5,76 (1,95)	-1,67 (0,69)	-0,72 (0,44)
Âge	-0,99 (0,19)	-07,66 (0,22)	-2,57 (0,08)	-25,7 (1,03)	-9,03 (0,37)	-5,50 (0,24)
N	1 223	3 541	3 541	3 541	3 541	3 541
B. Filles						
Variable dépendante - Moyenne	47,0 (5,7)	17,7 (23,0)	37,6 (8,1)	17,7 (23,0)	37,6 (8,1)	42,8 (4,7)
Gagnante d'un chèque	0,74 (0,44)	2,76 (0,98)	1,04 (0,34)	7,57 (2,60)	2,88 (0,91)	1,86 (0,59)
Âge	-1,14 (0,26)	-7,95 (0,31)	-2,65 (0,11)	-24,45 (1,34)	-8,54 (0,47)	-5,30 (0,32)
N	672	1 788	1 788	1 788	1 788	1 788
C. Garçons						
Variable dépendante - Moyenne	47,8 (5,5)	15,0 (22,4)	37,0 (7,9)	15,0 (22,4)	37,0 (7,9)	42,5 (4,6)
Gagnant d'un chèque	0,66 (0,48)	3,41 (0,97)	1,25 (0,34)	10,30 (3,10)	3,77 (1,10)	2,29 (0,71)
Âge	-0,80 (0,29)	-7,36 (0,32)	-2,49 (0,12)	-27,29 (1,61)	-9,63 (0,57)	-5,74 (0,37)
N	551	1 753	1 753	1 753	1 753	1 753

Nota : Les erreurs-types robustes sont indiquées entre parenthèses. Le point de censure est le percentile précisé de la distribution des notes d'examen, conditionnellement au fait de passer l'examen. Les écarts-types sont indiqués pour les moyennes de la variable dépendante. Une observation comportant une note de 1999 est perdue comparativement au tableau 3, parce qu'il n'existait pas de covariables pour l'imputation.

Tableau 4b. Estimations par régression des effets des chèques-éducation sur les notes en **mathématiques**

	Moindres carrés ordinaires			Tobit		
	MCO avec note>0	MCO censuré à 0	MCO censuré à 1 %	Tobit censuré à 0	Tobit censuré à 1 %	Tobit censuré à 10 %
	(1)	(2)	(3)	(4)	(5)	(6)
A. Tous les demandeurs						
Variable dépendante - Moyenne	42,5 (4,9)	14,7 (20,4)	35,7 (5,8)	14,7 (20,4)	35,7 (5,8)	37,6 (4,6)
Gagnant d'un chèque	0,40 (0,29)	2,74 (0,63)	0,79 (0,18)	7,9 (1,8)	2,29 (0,51)	1,98 (0,45)
Sexe masculin	0,43 (0,28)	-1,80 (0,62)	-0,32 (0,18)	-5,3 (1,7)	-1,26 (0,49)	-0,85 (0,43)
Âge	-0,37 (0,18)	-6,82 (0,24)	-1,73 (0,07)	-23,0 (0,93)	-6,30 (0,26)	-5,13 (0,23)
N	1 225	3 541	3 541	3 541	3 541	3 541
B. Filles						
Variable dépendante - Moyenne	42,3 (4,8)	15,9 (20,7)	35,9 (5,8)	15,9 (20,7)	35,9 (5,8)	37,8 (4,6)
Gagnante d'un chèque	0,18 (0,38)	2,37 (0,89)	0,62 (0,25)	6,68 (2,34)	1,84 (0,66)	1,60 (0,58)
Âge	-0,50 (0,23)	-07,09 (0,34)	-1,79 (0,10)	-21,88 (1,20)	-6,06 (0,34)	-4,89 (0,30)
N	673	1 788	1 788	1 788	1 788	1 788
C. Garçons						
Variable dépendante - Moyenne	42,8 (5,0)	13,5 (20,1)	34,8 (6,1)	13,5 (20,1)	34,8 (6,1)	38,9 (3,8)
Gagnant d'un chèque	0,70 (0,44)	3,13 (0,88)	1,02 (0,27)	9,40 (2,78)	3,03 (0,85)	2,21 (0,62)
Âge	-0,18 (0,27)	-6,54 (0,33)	-1,80 (0,10)	-24,34 (1,45)	-7,21 (0,44)	-4,58 (0,32)
N	552	1 753	1 753	1 753	1 753	1 753

Nota : Les erreurs-types robustes sont indiquées entre parenthèses. Le point de censure est le percentile précisé de la distribution des notes d'examen, conditionnellement au fait de passer l'examen. Les écarts-types sont indiqués pour les moyennes de la variable dépendante. Une observation comportant une note de 1999 est perdue comparativement au tableau 3, parce qu'il n'existait pas de covariables pour l'imputation.

Tableau 5a. Bornes des effets du chèque sur les notes en **langue**, pour les demandeurs qui auraient fait l'examen de l'ICFES de toute façon

Note seuil	Percentile	Note moyenne des perdants supérieure au quantile (1)	Sans variables de contrôle		Avec variables de contrôle	
			Borne inférieure (2)	Borne supérieure (3)	Borne inférieure (4)	Borne supérieure (5)
A. Tous les demandeurs						
0	72 ^e percentile	46,92 (5,52)	0,68 (0,33)	2,81 (0,31)	0,70 (0,33)	2,80 (0,31)
41	75 ^e percentile	48,68 (3,94)	0,46 (0,26)	2,47 (0,26)	0,49 (0,26)	2,46 (0,26)
47	85 ^e percentile	51,23 (3,00)	0,49 (0,27)	2,39 (0,28)	0,50 (0,27)	2,37 (0,28)
52	95 ^e percentile	55,60 (1,73)	0,35 (0,31)	1,38 (0,34)	0,36 (0,31)	1,39 (0,34)
B. Filles						
0	67 ^e percentile	46,61 (5,66)	0,65 (0,45)	2,29 (0,42)	0,74 (0,45)	2,33 (0,42)
43	75 ^e percentile	49,18 (3,66)	0,46 (0,35)	1,86 (0,35)	0,52 (0,35)	1,87 (0,35)
47	85 ^e percentile	51,49 (3,08)	0,31 (0,37)	2,32 (0,40)	0,32 (0,37)	2,31 (0,40)
52	95 ^e percentile	55,73 (1,82)	0,11 (0,45)	1,32 (0,52)	0,11 (0,45)	1,34 (0,53)
C. Garçons						
0	74 ^e percentile	47,33 (5,31)	0,67 (0,48)	3,19 (0,45)	0,66 (0,48)	3,18 (0,45)
36	75 ^e percentile	47,91 (4,63)	0,54 (0,44)	2,61 (0,42)	0,54 (0,43)	2,62 (0,41)
47	85 ^e percentile	51,57 (2,92)	0,69 (0,39)	1,36 (0,39)	0,69 (0,38)	1,35 (0,38)
51	95 ^e percentile	54,58 (2,08)	0,20 (0,44)	2,33 (0,46)	0,27 (0,45)	2,35 (0,47)

Nota : Le tableau présente les bornes calculées au moyen des formules de la proposition 1 dans le texte. Les moyennes et les écarts-types sont présentés dans la colonne 1. Les bornes estimées et les erreurs-types sont présentées dans les colonnes 2 à 4.

Tableau 5b. Bornes des effets du chèque sur les notes en **mathématiques**, pour les demandeurs qui auraient fait l'examen de l'ICFES de toute façon

Note seuil	Percentile	Note moyenne des perdants supérieure au quantile	Sans variables de contrôle		Avec variables de contrôle	
			Borne inférieure	Borne supérieure	Borne inférieure	Borne supérieure
		(1)	(2)	(3)	(4)	(5)
A. Tous les demandeurs						
0	70 ^e percentile	42,27 (4,75)	0,40 (0,29)	2,40 (0,27)	0,40 (0,29)	2,41 (0,27)
37	75 ^e percentile	43,74 (3,75)	0,35 (0,25)	1,76 (0,25)	0,34 (0,25)	1,76 (0,25)
42	85 ^e percentile	46,23 (3,18)	0,24 (0,28)	1,44 (0,28)	0,27 (0,28)	1,48 (0,28)
47	95 ^e percentile	50,29 (2,40)	-0,09 (0,39)	1,85 (0,42)	-0,11 (0,39)	1,80 (0,43)
B. Filles						
0	68 ^e percentile	42,25 (4,67)	0,14 (0,38)	1,63 (0,36)	0,18 (0,38)	1,66 (0,36)
39	75 ^e percentile	44,33 (3,34)	0,24 (0,32)	1,08 (0,32)	0,27 (0,32)	1,09 (0,32)
42	85 ^e percentile	46,22 (2,79)	0,28 (0,34)	1,53 (0,34)	0,32 (0,34)	1,55 (0,34)
47	95 ^e percentile	49,88 (1,56)	0,06 (0,40)	2,19 (0,42)	0,06 (0,40)	2,18 (0,42)
C. Garçons						
0	74 ^e percentile	42,31 (4,86)	0,70 (0,44)	3,27 (0,42)	0,70 (0,44)	3,28 (0,42)
35	75 ^e percentile	43,19 (4,26)	0,66 (0,40)	2,39 (0,40)	0,65 (0,40)	2,39 (0,39)
42	85 ^e percentile	46,30 (3,66)	0,27 (0,46)	1,31 (0,47)	0,32 (0,46)	1,37 (0,47)
45	95 ^e percentile	49,75 (3,16)	-0,60 (0,60)	2,45 (0,70)	-0,61 (0,60)	2,44 (0,71)

Nota : Le tableau présente les bornes calculées au moyen des formules de la proposition 1 dans le texte. Les moyennes et les écarts-types sont présentés dans la colonne 1. Les bornes estimées et les erreurs-types sont présentées dans les colonnes 2 à 4.

Tableau 6. Résultats de la régression par quantile

Seuil pour les langues	Seuil pour les mathématiques	Coefficient pour le chèque au	Notes en langue		Notes en mathématiques	
			Sans var. de contrôle	Avec var. de contrôle	Sans var. de contrôle	Avec var. de contrôle
A. Tous les demandeurs						
44	40	75 ^e percentile	3,00 (1,09)	5,50 (0,67)	2,00 (0,74)	4,40 (0,27)
48	43	85 ^e percentile	3,00 (0,41)	3,00 (0,44)	1,00 (0,45)	1,00 (0,72)
54	48	95 ^e percentile	2,00 (0,40)	1,50 (0,65)	1,00 (0,39)	1,00 (0,50)
N			3 541	3 541	3 541	3 541
B. Filles						
45	40	75 ^e percentile	3,00 (1,33)	4,80 (1,25)	1,00 (1,34)	3,00 (0,88)
48	43	85 ^e percentile	3,00 (0,27)	3,00 (0,71)	0,00 (0,77)	1,61 (0,99)
54	48	95 ^e percentile	2,00 (0,80)	1,50 (0,76)	1,00 (0,60)	1,00 (0,86)
N			1 788	1 788	1 788	1 788
C. Garçons						
44	39	75 ^e percentile	8,00 (1,76)	4,00 (0,44)	5,00 (1,32)	1,75 (0,76)
48	43	85 ^e percentile	3,00 (1,02)	2,00 (1,07)	1,00 (0,44)	1,00 (0,72)
54	47	95 ^e percentile	2,00 (0,82)	2,22 (0,93)	1,00 (0,58)	2,00 (0,86)
N			1 753	1 753	1 753	1 753

Nota : Le tableau présente les coefficients de la régression par quantile aux quantiles indiqués. Les notes seuils par quantile sont déterminées d'après les distributions des notes en fixant à zéro la note des élèves qui ne se sont pas inscrits à l'examen.

Figure 1a. Coefficients tobit selon le percentile de censure de la distribution des notes en langue

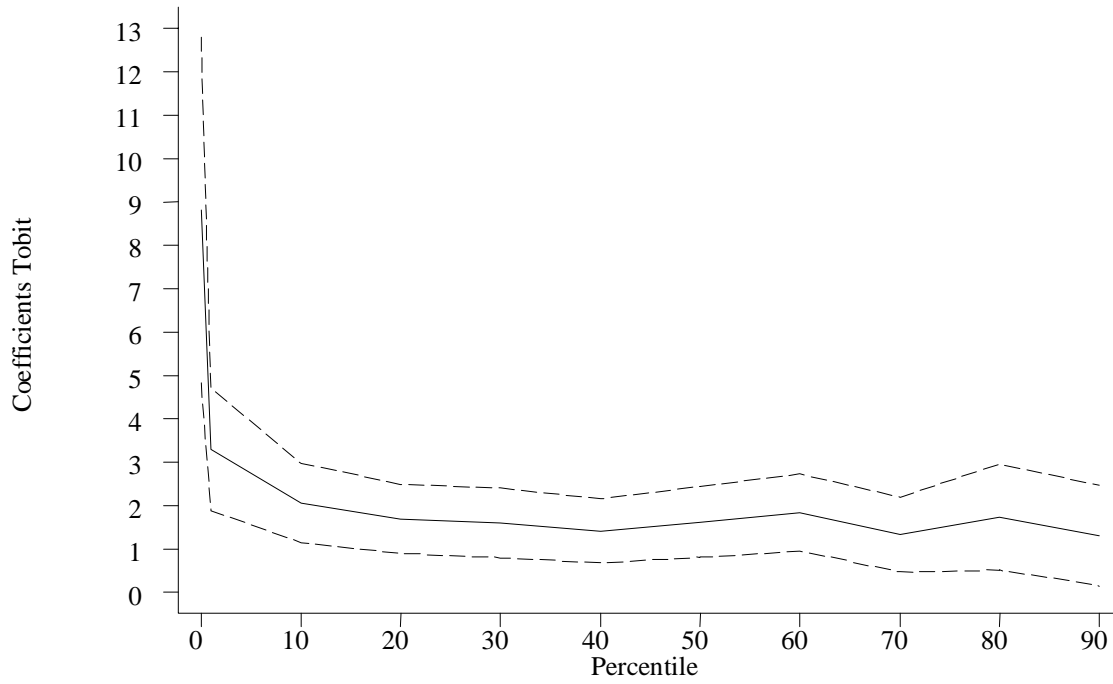


Figure 1b. Coefficients tobit selon le percentile de censure de la distribution des notes en mathématiques

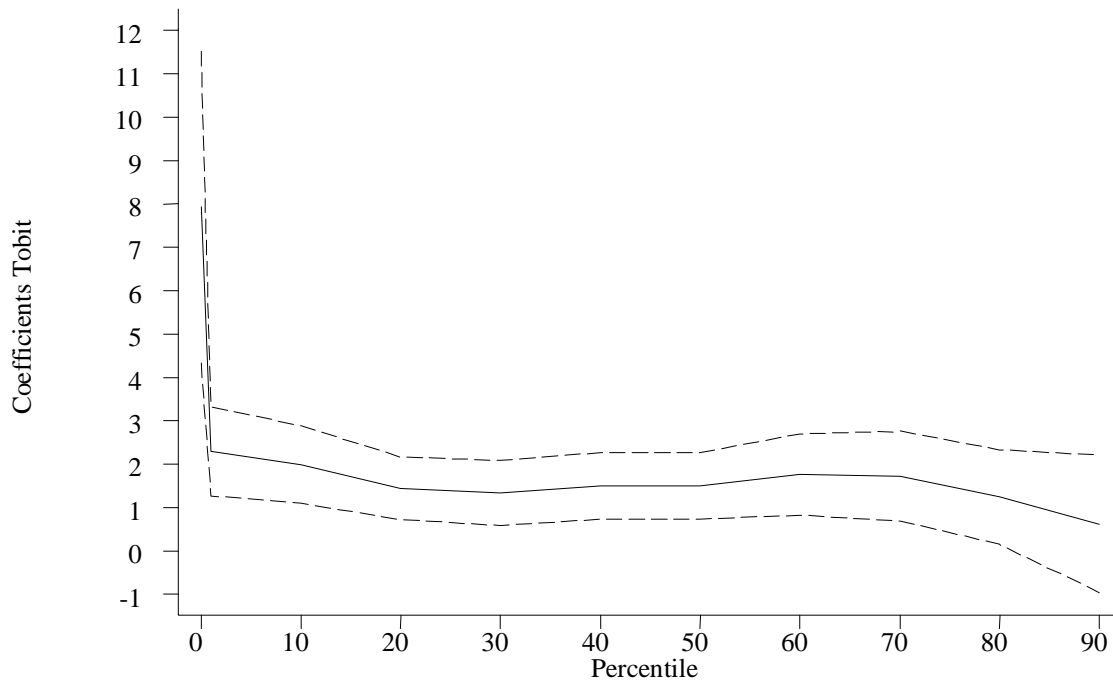


Figure 2a. Notes en langue selon la situation d'obtention du chèque (pas de correction pour le biais de sélection)

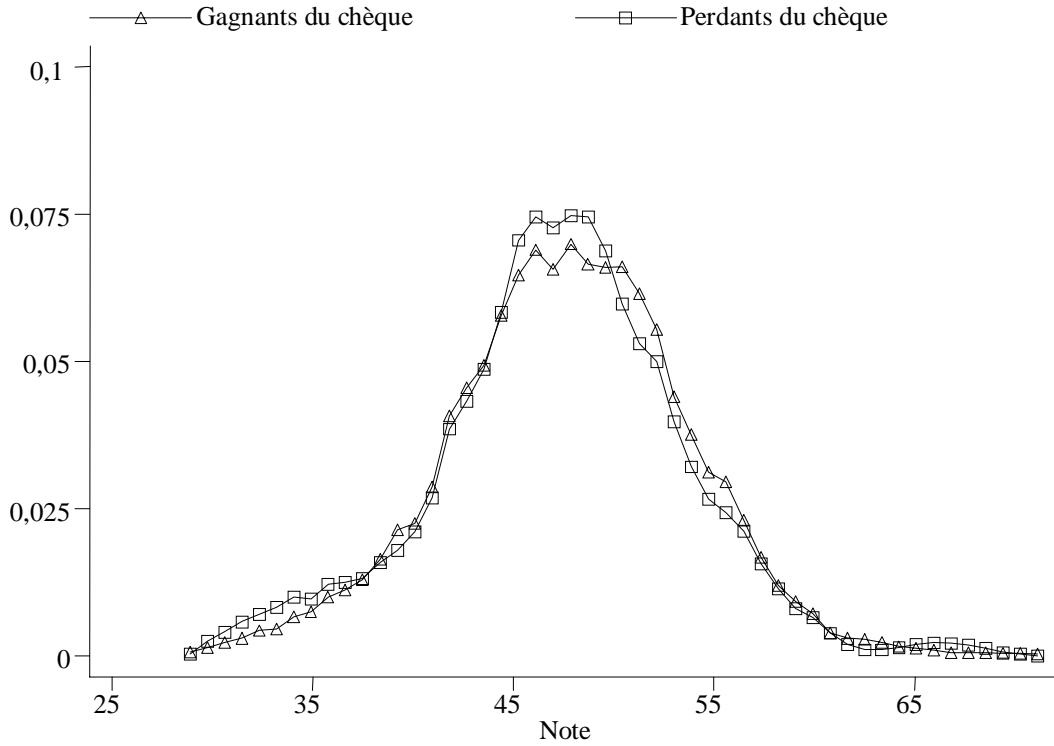


Figure 2b. Notes en mathématiques selon la situation d'obtention du chèque (pas de correction pour le biais de sélection)

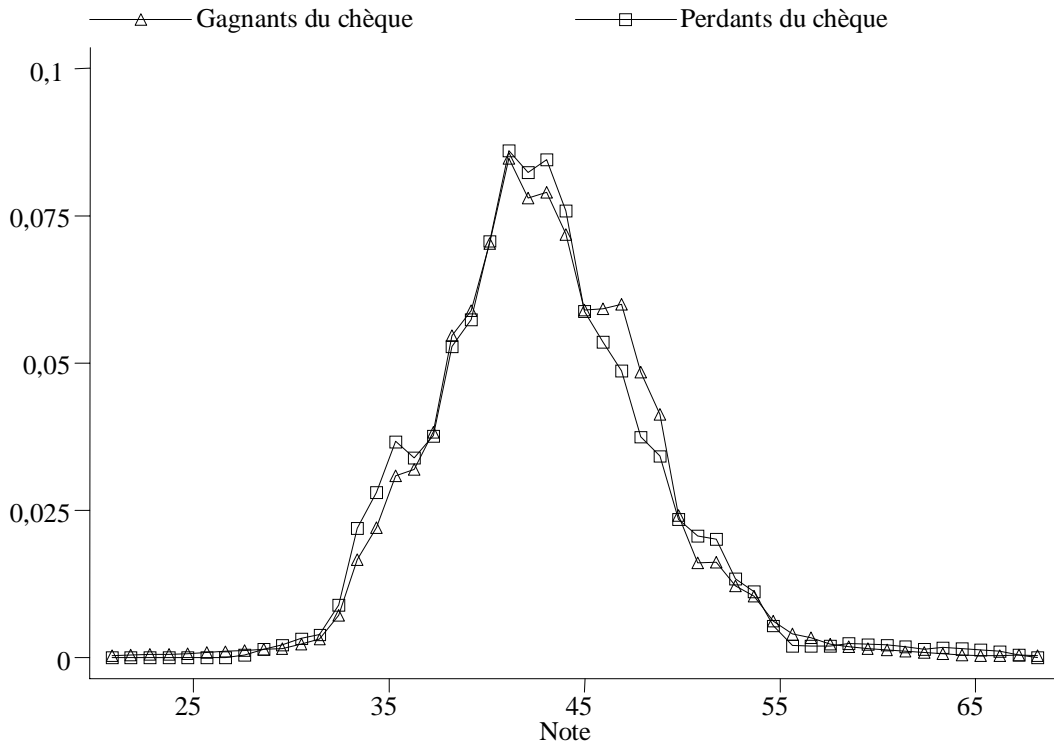


Figure 3a. Distribution des notes en langue selon la situation d'obtention du chèque pour des proportions égales de gagnants et de perdants

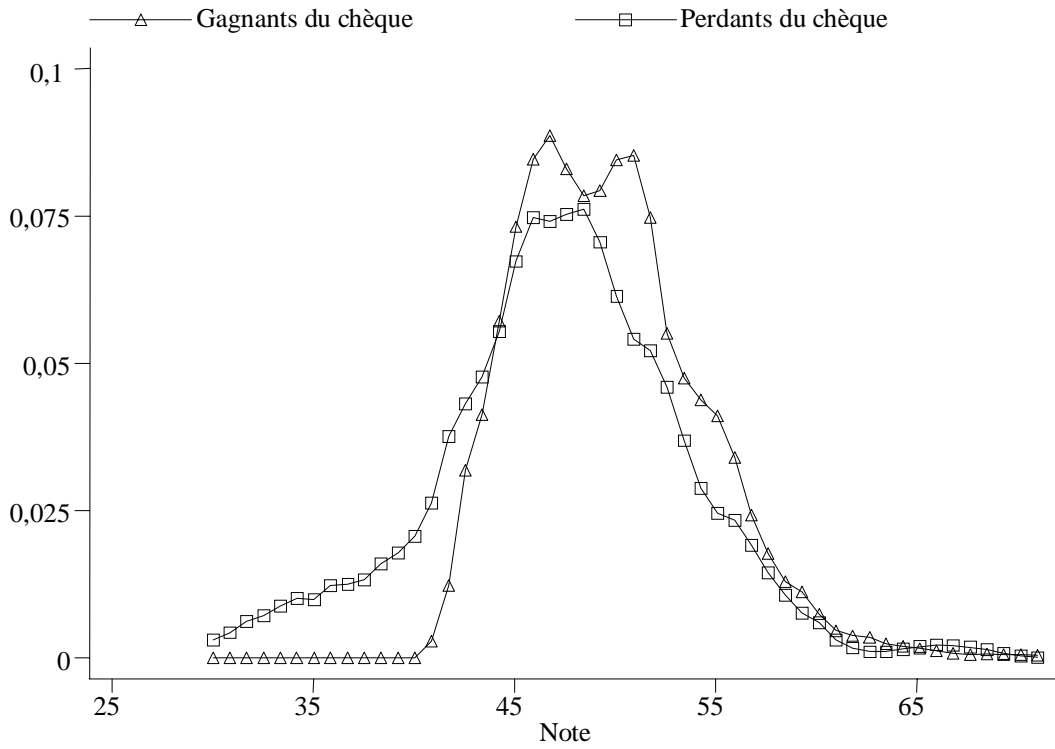


Figure 3b. Distribution des notes en mathématiques selon la situation d'obtention du chèque pour des proportions égales de gagnants et de perdants

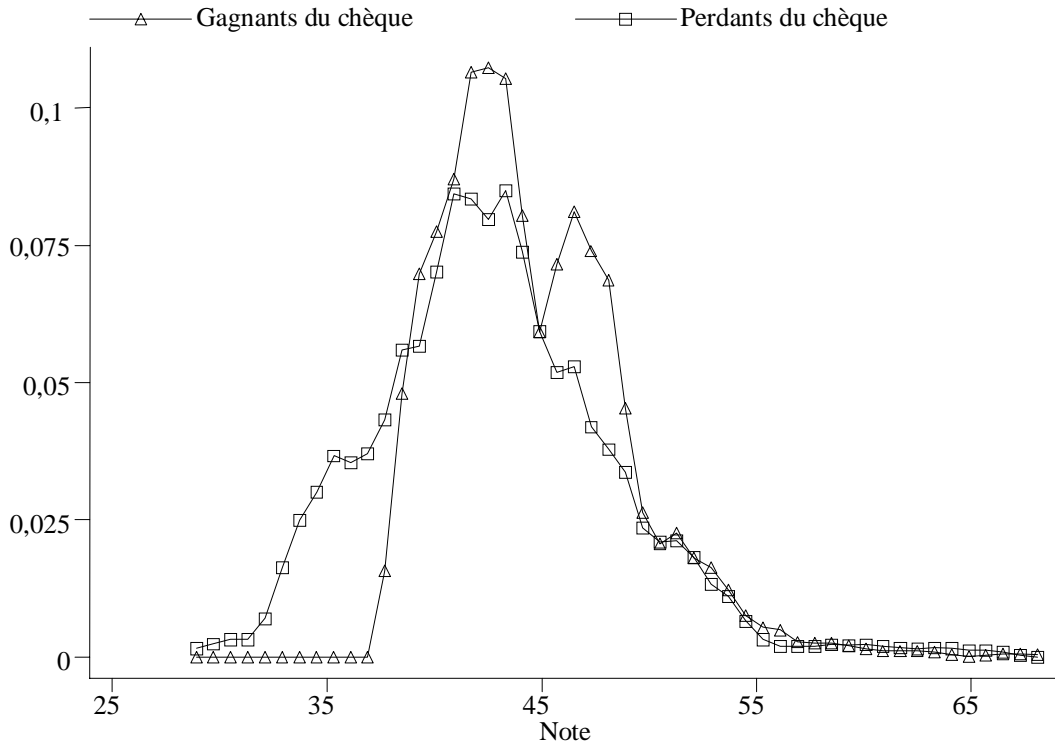


Figure Ia en annexe. Distribution des notes en langue à Bogota comparativement à l'échantillon PACES

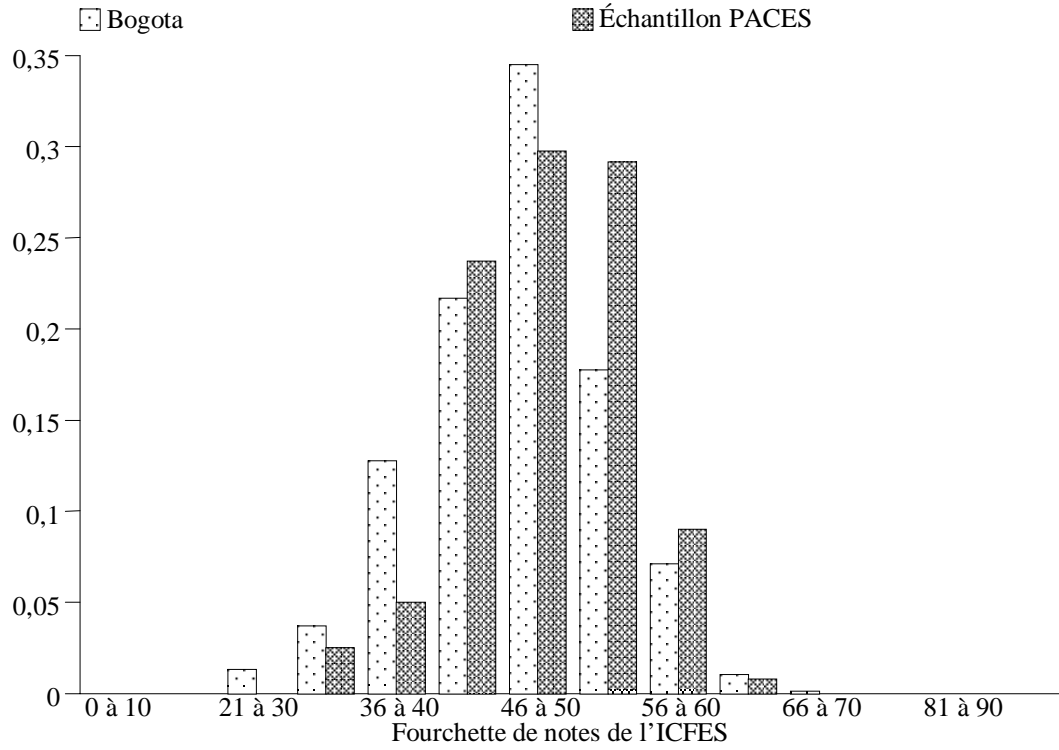


Figure Ib en annexe. Distribution des notes en mathématiques à Bogota comparativement à l'échantillon PACES

