

# Rapport Devoir 2

ÉCONOMÉTRIE APPLIQUÉE

[ECN2260](#)

**Auteur(e)s:**

Kelly Varin  
Johny Huynh  
Haris Imam

17 Avril 2024

# Introduction

La réussite scolaire est un sujet sociétale important pour le bon fonctionnement de la communauté. Ses répercussions semblent nombreuses dans la littérature que ce soit dans des domaines comme la réduction d'inégalité, de la pauvreté ou même juste d'une croissance économique plus durable, il semble que l'accès à un système d'éducation compétant est essentiel. Ainsi, le gouvernement a mis en place un programme d'immersion anglaise à des fins de spécialisation dans ce qui pourrait être nécessaire pour que les jeunes d'aujourd'hui puissent parfaire leurs connaissances en anglais. L'impact que peut avoir un tel programme sur les performances académiques ainsi que la participation en classe est le centre de questionnement de ce qui sera l'ultérieure analyse.

Afin de produire une dite analyse complète et rigoureuse des effets de la mise en place d'un tel programme, le portrait de l'étudiant type admis au programme d'immersion sera présenté. Dans un deuxième temps, l'effet que l'immersion anglaise peut avoir sur les notes des étudiants en anglais et en mathématique ainsi que sur la présence en classe sera relaté.

À des fins de clareté, les méthodes employées pour l'analyse des effets de la mise en place du programme seront tout d'abord expliquées et justifiées. Par la suite, les résultats obtenus seront présentés, puis discutés. Finalement, une conclusion à juste titre sera émise sur l'intérêt ou non de la mise en place du programme d'immersion anglaise et des recommandations seront suggérées.

## Méthodes utilisées

### Variables d'intérêt

Afin de bien comprendre les méthodes qui seront utilisées ultérieurement, les bases seront posées pour ce qui est des variables d'intérêt qui seront pertinentes pour l'analyse.

Tout d'abord, notons que le traitement T est désigné par le programme d'immersion anglaise. Le outcome est une réalisation possible sur laquelle le traitement pourrait avoir un effet causal. En effet, il sera exposé si l'immersion au programme d'anglais produit un effet sur les trois outcomes suivants: les notes d'anglais, les notes de mathématiques et la présence en classe.

Pour ce qui est des variables obtenues par les bases de données, il est possible d'y retrouver la liste pour chaque élève avec leur nom et identifiant associé (`eleve_id`), les revenus des parents (`parents_revenu`), le niveau de scolarité des parents (`parents_education`), le quartier de résidence (`Quartier`), une note à un test standardisé avant l'immersion au programme (`Score_0`), la présence en classe une fois le programme implémenté ainsi que l'évaluateur associé pour chaque étudiant (`eval_id`).

## Description des méthodes

### La double différence

En premier lieu, la double différence par régression sera utilisée pour quantifier l'effet causal du programme d'immersion anglaise sur la performance aux examens d'anglais et de mathématique des étudiants faisant partis de l'échantillon d'intérêt. En effet, pour procéder à cette méthode, il sera nécessaire d'avoir un groupe traité et un groupe non traité, le cas échéant la participation ou non au programme d'immersion anglaise. De plus, il est nécessaire que pour ces deux groupes, des données pour deux périodes distinctes soient observées, soit une période pré-traitement et une période

post-traitement. Cela donne un total de quatre groupes, qui seront utilisés pour la double différence tel que: ( $G_1 = PRE_T, G_2 = POST_T, G_3 = PRE_{NT}, G_4 = POST_{NT}$ ). Ainsi, quatre données cruciales seront utilisées: la moyenne au test standardisé avant l'immersion au programme pour les participants et les non-participants et la moyenne au test d'anglais ou de mathématiques une fois le programme complété pour les participants du programme et les non-participants.

Le fonctionnement de la double différence repose sur l'interaction des deux groupes d'intérêts à deux moments distincts. Le fait d'observer les mêmes groupes dans différents moments temporels permet de capturer l'effet temporel sur les groupes. De plus, le fait d'observer les changements pour un seul groupe traité en deux périodes permet d'obtenir la variation au sein du groupe et par le fait même avoir des effets fixes. Puisque le groupe traité et non-traité reste le même tout au long de l'immersion au programme d'anglais, des variables inobservables comme l'intelligence, la motivation, etc permet de capturer et contrôler ces inobservables par effet fixe.

Mathématiquement, la méthode repose sur nos quatre groupes nommés précédemment. Ainsi, la notation attribuée est  $\bar{Y}_{TR}^{PRE}$  la note moyenne au test préliminaire obtenu par les étudiants qui seront traités,  $\bar{Y}_{TR}^{POST}$  la note moyenne obtenue en anglais ou en mathématiques pour les étudiants ayant fait le programme d'immersion,  $\bar{Y}_{NT}^{PRE}$  la note moyenne au test standardisé chez les non-participants au programme et  $\bar{Y}_{NT}^{POST}$  la note moyenne obtenue en anglais ou mathématiques pour les non-participants au programme. Ainsi, pour bien quantifier l'effet qu'aura la participation au programme sur les notes obtenues en anglais et en mathématique, la régression suivante sera estimée :

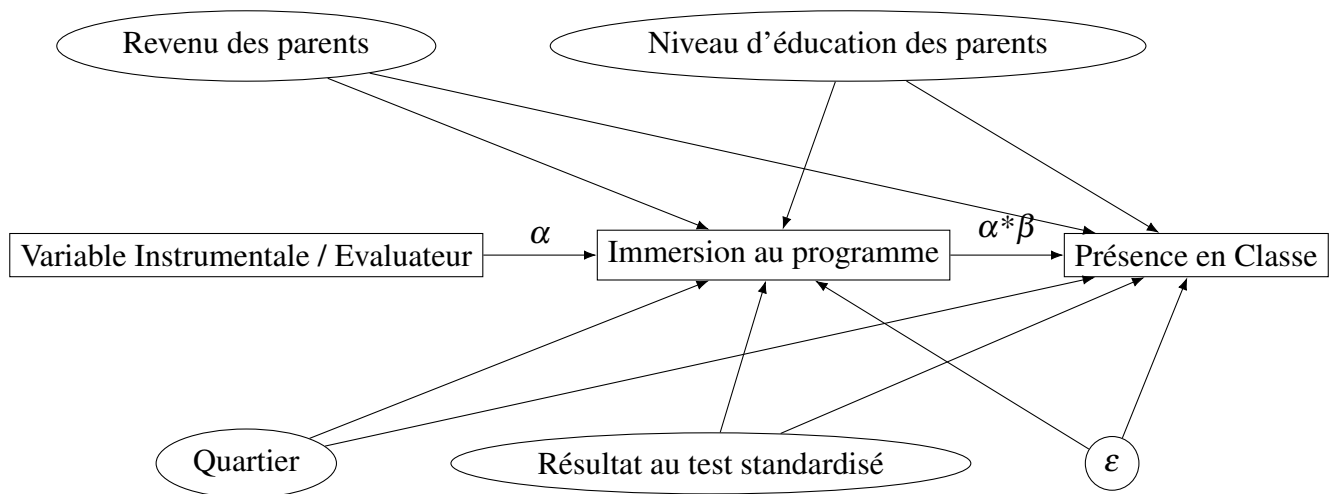
$$\hat{y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 \text{Traite}_i + \hat{\beta}_2 \text{Post}_t + \hat{\beta}_3 \text{Traite}_i * \text{Post}_t.$$

Il est à noter que le coefficient d'intérêt de la régression est  $\hat{\beta}_3$  qui est le coefficient du terme d'interaction entre Traite et Post. En effet, c'est ce coefficient qui indiquera l'effet de traitement qu'aura l'immersion au programme sur les notes obtenues. Il est aussi à noter que les variables traité, post et traite\_x\_post sont des vecteurs binaires de forme  $(\bar{Y}_T^{PRE}, \bar{Y}_T^{POST}, \bar{Y}_{NT}^{PRE}, \bar{Y}_{NT}^{POST})$ , tel que lorsque l'on participe au traitement, Traite= 1, sinon 0. Post=1 lorsque le programme est terminé, 0 sinon et finalement Traite\*Post est un terme d'interaction entre le fait de participer au traitement et quelle période est référée.

Puisqu'une amélioration à la performance scolaire peut être dû à des changements dans les variables inobservables au fil du temps, par exemple un choc de motivation. La méthode de double différence permet de contrôler, pour tout effet temporel constant, entre le nôtre groupe de traitement, c'est-à-dire ceux qui se sont inscrits au programme, et le groupe de contrôle, ceux qui ne sont pas inscrit. À noter que la méthode de double différence donne un effet moyen de traitement sur les traités et non un effet moyen de traitement total.

### Les variables instrumentales

La méthode par variable instrumentale sera utilisée pour ce qui est de l'analyse de l'effet de traitement de l'immersion au programme sur la présence en classe. Notons tout d'abord que le fait de participer ou non au programme semble aléatoire tel que tous les étudiants peuvent s'inscrire sur une base volontaire. De plus, l'évaluateur pourrait sembler avoir un pouvoir incitatif sur les étudiants tel que si celui-ci recommande ou non le fait de s'inscrire au programme, l'étudiant pourrait se voir plus enclin à choisir de participer au programme. Afin de bien visualiser, voici un graphe de causalité montrant les possibles interactions entre les variables d'intérêt.



Le graphe de causalité indique quelles variables ont un effet causal potentiel sur quelles autres variables. À titre d'exemple, le revenu des parents pourrait avoir un impact à la fois sur la participation au programme ainsi que sur les résultats scolaires de leur enfant, le cas échéant de l'étudiant concerné. Il est aussi à noter que la variable  $\varepsilon$  est l'agrégat de toutes les variables inobservables pour ce modèle comme l'intelligence, le niveau de motivation ou tout autres facteurs sociaux qui ne sont pas pris en compte par les données.

Ainsi, la méthode de variable instrumentale peut sembler d'intérêt ici. En effet, afin de bien trouver l'effet de la participation au programme, la variable qui pourrait être assignée comme étant une variable instrumentale serait l'évaluateur associé à chaque étudiant. En effet, il semble qu'un tel instrument respecte les conditions de validité et de pertinence. Tout d'abord, nous avons que l'évaluateur assigné n'affecte que le traitement, soit la participation au traitement. En effet, puisque l'assignation est aléatoire entre l'étudiant et l'évaluateur, en rien ses caractéristiques observables et inobservables semblent être en relation entre elles. Il est à noter que la force de corrélation entre la variable instrumentale et le traitement sera testée ultérieurement pour s'assurer d'un lien causal assez fort entre la variable instrumentale qu'est l'évaluateur et le traitement soit la participation au programme. De plus, il semble bien que la présence en classe sera affecté par le traitement préalablement affecté par notre variable instrumentale. En effet, lorsque toutes variables nécessitant un contrôle seront contrôlées par les régressions pertinentes, il ne restera que l'effet de la variable instrumentale sur la présence en classe induite qu'en passant par le traitement. Ainsi, les conditions de pertinence et de validité de l'instrument seront respectées.

Afin de bien estimer l'effet de la participation au programme sur la présence en classe, il faudra déterminer l'effet entre la variable instrumentale et le traitement, puis l'effet de traitement sur la présence en classe soit le outcome. Pour ce faire, il y aura deux régressions distinctes à faire

$$T_i = \alpha_0 + \alpha_1 Z_i + \varepsilon$$

$$Y_i = \delta_0 + \delta_1 Z_i + \gamma$$

La première régression sera pour déterminer l'effet de la variable instrumentale sur le traitement, soit l'impact de l'assignation d'un certain évaluateur sur le fait de participer ou non au programme d'immersion. La deuxième régression estime l'effet de la variable instrumentale sur l'outcome soit l'effet de l'assignation à un certain évaluateur sur la présence en classe. De plus, il est à remarquer

que dans les deux équations, l'instrument utilisé dénoté par  $Z$  est calculé par

$$Z_{ij} = \frac{\sum T_k}{|N|_j - 1}$$

Où  $N$  représente le nombre d'élève évalué pour un évaluateur  $j$ , et la sommation des  $T_k$  est le nombre de participant total pour un évaluateur lorsque l'on ne compte pas l'étudiant pour lequel l'instrument est calculé.

Avec tous ces outils en main, il sera possible de calculer l'effet de traitement tel que

$$\text{Effet de traitement} = \frac{\delta_1}{\alpha_1}$$

Où le  $\alpha_1$  et le  $\delta_1$  sont les coefficients préalablement trouvés grâce aux deux régressions linéaires précédemment exposées.

## Hypothèses

### Régression

Afin d'obtenir le profil-type des élèves qui participent au programme, il faudra faire une régression logistique sur le traitement en contrôlant toutes les variables observables pour chaque élève. Cela permettra d'associer les variables qui corréleront avec le traitement. Dans ce cas, il y a régression de la variable de traitement en contrôlant les variables pour le quartier, le niveau d'éducation des parents, le revenu des parents et le score au test standardisé avant l'immersion au programme. Cela montre la probabilité que chaque variable observée ajoutent ou soustraient à la probabilité de participer au programme d'immersion.

$$\text{Programme} = \hat{\beta}_0 + \hat{\beta}_1 \text{Quartier} + \hat{\beta}_2 \text{Education} + \hat{\beta}_3 \text{Revenu} + \hat{\beta}_4 \text{Score0.x} + \varepsilon$$

$$\begin{aligned} \text{Programme} = & \beta_0 + \beta_1 \text{MileEnd} + \beta_2 \text{Mont-Royal} + \beta_3 \text{Notre-Dame-De-Grâce} \\ & + \beta_4 \text{Outremont} + \beta_5 \text{Plateau} + \beta_6 \text{Rosemont} \\ & + \beta_7 \text{Saint-Henry} + \beta_8 \text{Vieux-Montreal} + \beta_{10} \text{Westmount} \\ & + \beta_{11} \text{Secondaire} + \beta_{12} \text{Primaire} + \beta_{13} \text{Université} \\ & + \beta_{14} \text{Revenu} + \beta_{15} \text{Score0.x} + \varepsilon \end{aligned}$$

### La double différence

L'hypothèse clé de la double différence est la *parallel trends assumption*. Cette hypothèse stipule qu'en absence de programme, en moyenne la tendance des performances scolaires entre les étudiants inscrits au programme d'immersion anglaise et les étudiants non-inscrits serait la même. Cela fait en sorte que toute variation aux performances est seulement attribuée à l'effet du traitement, le cas échéant le programme d'immersion anglaise.

### Variable Instrumentale

Deux conditions doivent être satisfaites. La première est que l'instrument doit être jugé pertinent. En effet, les évaluateurs dans le modèle affectent la participation au traitement en incitant les élèves qui performaient déjà bien au paravant. Afin de s'assurer de la pertinence de l'instrument, la statistique-t associée au coefficient  $\alpha_1$  associé à la régression du traitement en fonction de l'instrument  $z$  soit

supérieur à  $\sqrt{10}$ . La deuxième condition est que l'instrument doit être valide. Pour la bonne validité de l'instrument, les évaluateurs doivent affecter l'outcome qui est la présence en classe en passant que par la variable de traitement. Cela se constate en regardant le graphe de causalité exposé dans la description de la méthode pour les variables instrumentale plus haut. En effet, les chemins possibles comportant les directions par lesquelles les variables semblent s'influencer sont les suivantes

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \rightarrow \text{Presence en Classe} \quad (1)$$

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \leftarrow \text{Revenue} \rightarrow \text{Presence} \quad (2)$$

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \leftarrow \text{Niveau d'éducation} \rightarrow \text{Presence} \quad (3)$$

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \leftarrow \text{Quartier} \rightarrow \text{Presence} \quad (4)$$

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \leftarrow \text{Score0} \rightarrow \text{Presence} \quad (5)$$

$$\text{Evalueur\_Id} \rightarrow \text{Immersion} \leftarrow \varepsilon \rightarrow \text{Presence} \quad (6)$$

Il est possible d'observer que pour les chemins (2) à (6), la variable traitement qui est le programme d'immersion, devient une variable *collider*, c'est-à-dire qu'elle est causée par deux variables observables pour lesquelles il est possible de contrôler l'effet. La présence d'une variable *collider* ferment les chemins dans lesquelles elle se retrouve de manière automatique. Ainsi, ces chemins qui sont considérés comme n'étant pas de bons chemins ne sont plus source d'inquiétude tel qu'ils sont, comme que mentionné, fermés d'eux-mêmes. La variable instrumentale permet donc la possibilité de calculer un effet entre l'évaluateur et la présence en classe en passant par le traitement de l'immersion sans avoir à s'inquiéter de variables observables et inobservables.

## Résultats

### Profil type des participants

À l'aide de la régression logistique multiple des variables observables sur le traitement, le profil type des élèves qui participent au programme d'immersion anglaise ainsi que plusieurs constantes attribuées à chaque variable de la régression sont obtenues. Cela estime:

$$\begin{aligned} P(\text{Programme}) = & -4.35 \times 10^{-1} + 2.509 \times 10^{-2} \text{MileEnd} + 2.313 \times 10^{-2} \text{Mont-Royal} \\ & + 4.31 \times 10^{-2} \text{Notre-Dame-De-Grâce} - 3.563 \times 10^{-2} \text{Outremont} \\ & + 2.846 \times 10^{-2} \text{Plateau} + 3.023 \times 10^{-2} \text{Rosemont} \\ & + 3.718 \times 10^{-2} \text{Saint-Henry} + 1.034 \times 10^{-1} \text{Vieux-Montreal} \\ & + 1.445 \times 10^{-2} \text{Westmount} + 9.712 \times 10^{-2} \text{Secondaire} \\ & + 1.674 \times 10^{-1} \text{Primaire} - 3.94 \times 10^{-2} \text{Université} \\ & + 3.259 \times 10^{-6} \text{Revenu} + 1.121 \times 10^{-2} \text{Score0.x} + \varepsilon \end{aligned}$$

## Impact sur la performance académique

Tel qu'exposé précédemment, afin de calculer l'effet de l'immersion au programme d'anglais sur les résultats obtenus en anglais et en mathématique, la régression suivante sera estimée

$$\hat{y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 \text{Traite}_i + \hat{\beta}_2 \text{Post}_t + \hat{\beta}_3 \text{Traite}_i * \text{Post}_t + \varepsilon$$

Le coefficient d'intérêt pour la double différence est  $\hat{\beta}_3$  qui est le coefficient du terme d'interaction entre les variables *Traite* et *Post*. Ayant calculé le résultat moyen du test préalable au programme pour les participants et les non-participants distinctement ainsi que le résultat moyen en anglais ou en mathématique pour ces deux mêmes groupes, les données obtenues sont les suivantes.

Score moyen	Traite	Post	Traite*Post
63.29377	1	0	0
75.29767	1	1	1
47.53292	0	0	0
43.36214	0	1	0

Table 1: Données pour les scores moyens d'anglais et du test préliminaire (data\_dd\_ang)

Score moyen	Traite	Post	Traite*Post
63.29377	1	0	0
72.44942	1	1	1
47.53292	0	0	0
43.35185	0	1	0

Table 2: Données pour les score moyen de mathématiques et du test préliminaire (data\_dd\_math)

Avec les données exposées ci-haut, voici les régressions obtenues

Régression avec les notes d'anglais:

$$\hat{y}_{it} = 47.533 + 15.761 \text{Traite}_i - 4.171 \text{Post}_t + 16.175 \text{Traite}_i * \text{Post}_t + \varepsilon$$

Régression avec les notes de mathématique:

$$\hat{y}_{it} = 47.533 + 15.761 \text{Traite}_i - 4.181 \text{Post}_t + 13.337 \text{Traite}_i * \text{Post}_t + \varepsilon$$

Puisque le coefficient d'intérêt est  $\hat{\beta}_3$ , ce qui correspond à l'effet de traitement obtenu par la double différence, il est obtenu qu'en moyenne l'effet de la participation au programme sur les résultats scolaires en anglais et en mathématiques sont respectivement de 16.175 et 13.33 soit la valeur du  $\hat{\beta}_3$  estimée. Cela signifie que l'effet du programme d'immersion anglaise augmente les résultats aux examens d'anglais des élèves de 16.175 points de pourcentage. Similairement, l'effet du programme d'immersion anglaise augmente les résultats aux examens de mathématique des élèves ayant participé de 13.33 points de pourcentages en moyenne. Notons cependant que la méthode de double différence nous rapporte un effet de traitement chez les traités et non un effet de traitement moyen.

## Impact sur la présence en classe

En appliquant la méthode pour les variables instrumentales exposée plus haut, la première régression du traitement en fonction de la variable instrumentale  $z$  est de la forme suivante

$$T_i = \alpha_0 + \alpha_1 Z_i + \varepsilon$$

$$\hat{T}_i = 0.081 + 0.842 \hat{Z}_i$$

De plus, en faisant la régression de l'outcome, soit le taux de présence en classe, en fonction de la variable instrumentale  $z$ , la régression suivante est obtenue

$$Y_i = \delta_0 + \delta_1 Z_i + \gamma$$

$$\hat{Y}_i = 1.99 + 0.109 \hat{Z}_i$$

Il est à noter que le  $\alpha_1$  trouvé détient une statistique-t de 10.257. Il a été mentionné préalablement que pour qu'une variable instrumentale soit valide, il faut que la statistique-t associée au  $\alpha_1$  soit supérieure à  $\sqrt{10}$ , ce qui dans le cas échéant est respecté. Ainsi, il y a bel et bien respect de la condition de validité de l'instrument.

Ainsi, l'effet de traitement calculé est:

$$\text{Effet de traitement} = \frac{\delta_1}{\alpha_1} = \frac{0.109}{0.842}$$

$$\text{Effet de traitement} = 0.129$$

## Analyse des résultats

### Profile type

La régression dévoile que les élèves provenant de la région de Mont-Royal ont en moyenne 2.313% plus de chance de participer au programme, du côté de Mile End, les élèves ont en moyenne 2.509% de chance de participer. De plus, la région de Notre-Dame-De-Grâce ajoute 4.31% de chance à la participation et malheureusement la probabilité de participer diminue en moyenne de 3.563% si l'élève provient du secteur d'Outremont. Pour finir, les quartiers du Plateau, Rosemont, Saint-Henry, du Vieux-Montréal et de Westmount augmentent les probabilités de participer au programme de 2.846%, 3.023%, 3.718%, 10.34% et 1.445% respectivement. On peut constater que les quartiers les plus riches, par exemple Mont-Royal, Westmount et Plateau, ont moins de probabilité de participer au programme car les évaluateurs incitent surtout les étudiants qui étaient de milieu socio-économiques moins favorable. Pour le niveau d'éducation des parents, on voit que s'ils ont fini le secondaire, les chances augmentent de 9.712%, de 16.74% pour les parents qui ont fini le primaire seulement et une réduction de 3.94% pour les parents qui ont complétés leurs études universitaires. Chez le revenu des parents, on constate que pour chaque dollar gagnés par les parents, la probabilité de participer au programme augmente en moyenne de  $3.259 \times 10^{-4}\%$ . Enfin, on trouve que pour chaque pourcentage que l'élève avait reçu à l'examen standardisé avant l'immersion augmente la probabilité de participer au programme de 1.121%.



## **Impact sur la performance scolaire**

Tel que mentionné précédemment, le coefficient  $\hat{\beta}_3$  estimé pour l'impact de l'immersion anglaise sur les notes d'anglais est de 16.175. Cela signifie que l'effet de l'immersion sur les notes d'anglais semble être supérieur de 16.175 point de pourcentage chez les groupes d'élève ayant participé au programme versus ceux qui ne l'auraient pas fait.

Pour ce qui est du  $\hat{\beta}_3$  estimé pour l'impact de l'immersion anglaise sur les notes en mathématiques, l'estimation est de 13.33. Cela implique que les notes pour le groupe d'élève ayant participé au programme d'immersion tend à avoir une note supérieure en mathématiques que le groupe d'élèves n'ayant pas participé, et ce d'une différence de 13.33 points de pourcentage.

Compte tenu de la nature du programme étant un programme d'immersion linguistique anglophone, le fait que l'impact sur les notes soit plus grande pour les notes d'anglais versus celles de mathématiques pourrait être causé par le fait que le programme avait comme objectif principal l'amélioration des compétences visées en anglais. Ainsi, une plus grande augmentation dans les notes d'anglais versus celles de mathématiques semble aller dans le même sens que cet objectif.

## **Impact sur la présence en classe**

Tel que calculé avec la méthode de variables instrumentales précédemment, l'effet de traitement de la participation au programme sur la présence en classe est de 0,129. Cela représente donc un effet de l'implantation du programme sur la présence en classe de 12,9% de plus chez les étudiants ayant participé au programme versus ceux qui ne l'aurait pas fait. Ainsi, il semble clair que le fait de participer au programme implique un effet positif sur la présence en classe des élèves, en voyant le nombre moyen de participation de 12,9% de plus que pour le groupe d'élèves n'ayant pas participé au programme d'immersion.

## **Conclusion**

### **Recommandation**

Pour conclure, tel qu'expliqué dans l'analyse de résultat, le programme d'immersion anglais a un effet causal positif sur la présence et la performance aux examens de mathématiques et d'anglais des étudiants. Dorénavant, quelques améliorations peuvent-êtes apportées à ce programme pour de meilleurs résultats qui seront moins biaisés et pour avoir un programme qui serait plus équitable. En premier lieu, comme analysé précédemment, les évaluateurs influencent beaucoup la prise de décision des élèves mais surtout pour ceux qui performaient déjà bien ou ceux de milieu socio-économique moins favorable. Or pour améliorer ce programme, les évaluateurs pourraient ne pas inciter les élèves à participer de façon proactive en étant plus présent pour aider les élèves qui sont intéressés sans aucune intervention réactive tel que la recommandation ou non d'élèves spécifiques. Or, cela rendrait le programme plus juste pour les élèves intéressés, mais qui habitent dans un milieu socio-économique plus favorable ou ceux qui ne performaient pas necessairement bien initialement, mais qui ont quand même un interet à s'améliorer ou juste une curiosité intellectuelle.