
Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?

Auteur : Laboureur, Lola

Promoteur(s) : Baye, Ariane

Faculté : Faculté de Psychologie, Logopédie et Sciences de l'Éducation

Diplôme : Master en sciences de l'éducation, à finalité spécialisée en enseignement

Année académique : 2022-2023

URI/URL : <http://hdl.handle.net/2268.2/17074>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.



Université de Liège
Faculté de Psychologie, Logopédie & Sciences de l'Éducation
Département des Sciences de l'Éducation

Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?

Promotrice : Ariane Baye

Assistant doctorant : Dylan Dachet

Lectrices : Dominique Lafontaine et Valérie Quittre

Mémoire présenté par Lola Laboureur en vue de l'obtention du grade de Master en Sciences de l'Éducation à finalité spécialisée en Enseignement.

Année académique 2022 – 2023

Remerciements

Je tiens tout d'abord à remercier Monsieur Dylan Dachet, pour sa disponibilité, ses relectures et ses précieux conseils durant les deux années écoulées. Monsieur Dachet, merci pour l'intérêt porté à l'égard de ce travail.

Un tout grand merci à Madame Baye, promotrice de ce mémoire, pour ses conseils avisés, sa confiance en mon travail et l'aide apportée lors de la récolte de données.

Je remercie mes lectrices, Madame Lafontaine et Madame Quittre pour l'attention qu'elles porteront à ce travail.

Enfin, je tiens également à remercier chacun des membres de ma famille et particulièrement mes parents, qui ont fait preuve d'un soutien sans faille durant mon cursus universitaire.

Pour terminer, merci à toi Samir, pour la patience et le réconfort dont tu as fait preuve durant ces années d'étude.

Table des matières

1	INTRODUCTION	1
2	REVUE DE LA LITTÉRATURE.....	3
2.1	La docimologie et l'évaluation	3
2.2	Les élèves issus de l'immigration et l'école.....	6
2.2.1	Leurs performances.....	6
2.2.2	Leurs parcours scolaires	7
2.2.3	La discontinuité culturelle et la discrimination systémique.....	7
2.3	Le poids de l'origine sociale et de l'origine socio-économique des élèves.....	10
2.4	L'impact des stéréotypes.....	11
2.4.1	L'activation des stéréotypes par les enseignants.....	11
2.4.2	Sur les élèves	12
2.5	Les biais de l'évaluation.....	13
2.5.1	Les biais liés à l'évalué.....	14
2.5.2	Des biais liés aux modalités de testing et à l'évaluateur.....	15
2.5.3	Peut-on éviter ces biais ?.....	16
2.6	Les conséquences de ces biais	17
2.6.1	Le comportement des évaluateurs.....	17
2.6.2	L'avenir des élèves.....	20
3	OBJET DE RECHERCHE ET HYPOTHESES	21
3.1	Objet de recherche.....	21
3.2	Hypothèses	21
3.2.1	Hypothèse n°1 (LUCIEN VS JORDAN VS AYOUB)	21
3.2.2	Hypothèse n°2 (LUCIEN VS JORDAN).....	22
3.2.3	Hypothèse n°3 (LUCIEN VS AYOUB)	22
3.2.4	Hypothèse n°4 (JORDAN VS AYOUB).....	22
3.2.5	Hypothèse n°5 (LUCIEN VS JORDAN VS AYOUB)	22
4	METHODOLOGIE.....	23
4.1	Design de la recherche	23
4.2	Instruments de mesure et données récoltées	25
4.2.1	Instruments de mesure	25
4.2.1.1	Les trois rédactions.....	25
4.2.1.2	Les grilles d'évaluation	27
4.2.2	Données récoltées.....	28
4.3	Public cible et échantillon	28

4.4	Procédure de recrutement.....	30
4.5	Aspects éthiques	30
4.6	Les analyses statistiques.....	31
4.6.1	Indices de tendance centrale	31
4.6.2	Indices de dispersion	32
4.6.3	Ampleur de l'effet	32
4.6.4	Analyse de la variance simple (ANOVA)	33
5	RESULTATS ET INTERPRETATIONS	34
5.1	Hypothèse n°1 : LUCIEN VS JORDAN VS AYOUB.....	35
5.2	Hypothèse n°2 : LUCIEN VS JORDAN	38
5.3	Hypothèse n°3 : LUCIEN VS AYOUB.....	40
5.4	Hypothèse n°4 : JORDAN VS AYOUB	41
5.5	Hypothèse n°5 : LUCIEN VS JORDANS VS AYOUB.....	42
6	DISCUSSION	47
6.1	Discussion autour des différentes hypothèses	47
6.1.1	Hypothèse n°1 : LUCIEN VS JORDAN VS AYOUB.....	48
6.1.2	Hypothèse n°2 : LUCIEN VS JORDAN	50
6.1.3	Hypothèse n°3 : LUCIEN VS AYOUB.....	52
6.1.4	Hypothèse n°4 : JORDAN VS AYOUB	53
6.1.5	Hypothèse n°5 : LUCIEN VS JORDAN VS AYOUB.....	54
6.2	Limites de cette recherche et perspectives	58
7	CONCLUSION.....	59
8	BIBLIOGRAPHIE.....	61
9	RÉSUMÉ	67
10	ANNEXES	68

1 INTRODUCTION

Les élèves issus de l’immigration représentent une partie importante de la population belge. Dans notre système scolaire, en Fédération Wallonie-Bruxelles, plus d’un cinquième des élèves sont d’origine immigrée¹ (Lafontaine et al., 2017). De nombreuses classes sont donc fréquentées, en partie, par des élèves issus de l’immigration. Beaucoup de préjugés et de stéréotypes tournent autour de l’immigration en général, notamment centrés sur les différentes religions, cultures, et cela se répercute sur les enfants concernés, et ce dès leur plus jeune âge. Certains enseignants sont influencés, de manière consciente ou non, par ces stéréotypes. Dans un pays comme le nôtre, ayant connu une immigration de grande ampleur, certains discours médiatiques voire politiques considèrent l’immigration comme étant un problème sociétal. Au même titre, un enfant d’immigré en échec scolaire est davantage caractérisé par un manque de volonté d’intégration que par un dysfonctionnement du système scolaire (Ichou, 2018).

L’école fait-elle preuve d’équité face à la diversification de nos publics scolaires ? L’efficacité et l’équité de nos systèmes éducatifs font l’objet d’enquêtes internationales comme PISA (le Programme International pour le Suivi des Acquis des élèves) organisé par l’OCDE. Demeuse et Baye (2005, p.11) définissent l’équité comme étant un des principes fondamentaux des « démocraties contemporaines ». De plus, ces auteurs expliquent que l’objectif de l’école est de mettre en place une équité de traitement qui pourrait permettre à tous les enfants d’apprendre de manière égale. Dans le cadre de ce mémoire, nous nous intéressons aux inégalités de traitement que certains élèves, défavorisés sur les plans culturels et sociaux, pourraient subir dans l’évaluation de leurs compétences.

« L’école est un lieu dans lequel l’évaluation est omniprésente » (Leclercq et al., 2004, p.273). Autin et al. (2019) annoncent que les établissements scolaires doivent porter une grande attention au rôle et à la signification de l’évaluation. En effet, l’utilisation des évaluations, et surtout celles à caractère normatif/sélectif, pourrait conduire les évaluateurs à créer artificiellement un écart de performances entre des élèves issus de classes sociales différentes, même s’il n’existe objectivement aucun écart de performances entre ces élèves. Différentes recherches en docimologie se sont intéressées aux biais impactant tout jugement évaluatif. Ainsi, certains chercheurs ont montré qu’à performances réelles identiques, des élèves porteurs du stéréotype « favorisé » sont surévalués comparativement aux élèves qui portent le stéréotype

¹ Élèves nés à l’étranger ou dont les parents sont nés à l’étranger.

« statut socio-économique faible » qui eux, sont sous-évalués (Autin et al., 2019 ; Rangvid, 2015). Par ailleurs, Bonefeld et Dickhäuser (2018), Rangvid (2015) et Sprietsma (2013) ont mis en lumière des résultats similaires au désavantage des élèves d'origine immigrée. Ces constats interpellent puisque si ce phénomène se reproduit pour un grand nombre d'évaluations, la différence au bout du compte pourrait avoir des conséquences dans le parcours scolaire de ces élèves discriminés. Nos (futurs) enseignants en Fédération Wallonie-Bruxelles (FW-B) sont-ils, eux aussi, influencés dans leur cotation par l'origine ethnique et l'origine sociale des élèves ?

La mise en parallèle de ces deux constats, à savoir, l'existence de biais de l'évaluation au détriment des élèves défavorisés (Autin et al., 2019) et au détriment des élèves d'origine immigrée (Rangvid, 2015) soulève une certaine interrogation. Les effets liés aux stéréotypes sur les élèves d'origine immigrée sont-ils dépendants ou non des effets liés au statut socio-économique ? A notre connaissance, aucune recherche n'a étudié l'influence conjointe de ces deux variables.

Le présent mémoire vise donc à étudier l'existence d'un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale. Dans ce cadre, une expérience docimologique sera conduite auprès d'enseignants et de futurs enseignants. L'objectif de la présente recherche est de questionner les processus qui contribuent à l'écart des performances entre les élèves par le biais de l'évaluation elle-même et de vérifier s'il existe une différence de cotation entre un élève natif, un élève natif issu d'un milieu socio-économique défavorisé et un élève issu de l'immigration.

En d'autres termes, ce mémoire vise à répondre à la question suivante : « *Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?* »

Ce mémoire est articulé autour de huit parties dont la première est la présente **introduction**. Ensuite, une **revue de la littérature** articulée autour des grands thèmes de la question de recherche sera exposée. Par la suite, une partie dédiée à **l'objet de recherche et aux hypothèses** sera présentée. En découle, la **méthodologie** de cette recherche. Ensuite, nous exposerons les **résultats** et dans la partie suivante, nous les **discuterons** à la lumière de la littérature scientifique. Enfin, pour terminer ce travail, la **conclusion générale** ainsi que **les limites et perspectives** seront décrites.

2 REVUE DE LA LITTÉRATURE

La revue de la littérature reprend des points théoriques que nous estimons pertinents dans le cadre de cette recherche. Elle est organisée en six chapitres. Nous aborderons dans un premier temps, la docimologie ainsi que l'évaluation dans le premier chapitre. Ensuite, nous aborderons dans le second chapitre, différents points concernant les élèves issus de l'immigration. Nous discuterons du poids des origines sociales et du statut socio-économique des élèves dans le troisième chapitre. Dans la foulée, nous considérerons dans le chapitre quatre l'impact des stéréotypes. En découlent, dans le chapitre cinq, les biais de l'évaluation. Enfin, dans le sixième et dernier chapitre, nous terminerons par un point concernant des conséquences de ces biais sur le comportement des évaluateurs ainsi que sur l'avenir des élèves.

2.1 La docimologie et l'évaluation

Avant de rentrer dans le vif du sujet de cette section dédiée à la docimologie et à l'évaluation, il semble opportun de définir certains concepts clés. De Landsheere (1992, p.17) définit la docimologie en ces termes « une science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés ». Detroz (2021) complète en indiquant que la docimologie est une discipline scientifique qui s'intéresse à la manière dont les notes sont attribuées par les évaluateurs. Dans le but d'assurer davantage de justice scolaire, la docimologie offre des méthodes de mesure qui permettent plus d'objectivité dans l'évaluation des performances des élèves (De Landsheere, 1992). De Ketele (1993) explique que le paradigme docimologique étudie la fiabilité des évaluations. De nos jours, on parle aussi de psychologie de l'évaluation puisqu'on étudie et analyse le comportement de l'évaluateur et de l'évalué (Noizet & Caverni, 1978, cités par De Landsheere, 1992).

Detroz (2021) s'inspire de Romainville (2017) et compare l'action d'évaluer à une danse à trois temps. Le premier temps est consacré au recueil de données, à la prise d'informations. Ensuite, dans un deuxième temps, l'évaluateur va interpréter ces informations et les analyser. Pour terminer, l'évaluateur fournit un feedback et/ou une décision.

En outre, De Landsheere (1992) distingue la mesure de l'évaluation. La première se traduit par un chiffre, mesurer, c'est « assigner un nombre à un objet ou à un événement selon une règle logiquement acceptable » (Guilford, 1973, cité par De Landsheere, 1992, p.21). La seconde, l'évaluation, est un moyen de communication entre les élèves et le maître « aboutissant à la suite d'un nombre imprévisible d'interactions, à l'atteinte des objectifs pédagogiques » (De

Landsheere, 1992, p.22). De Ketele (1993, p.62) complète et définit l'éducatrice comme étant « une science de la mesure dans le champ des sciences de l'éducation ». De Landsheere (1992) distingue également les deux notions suivantes : le score et la note. La première relève du caractère objectif : on additionne ou soustrait le nombre de points en fonction de règles établies. La seconde désigne l'appréciation d'une performance : elle est relative mais peut être objective ou subjective. Selon De Landsheere (1992), l'évaluation couvre un triple rôle. Premièrement, elle endosse un rôle diagnostique qui permet de rendre compte des matières non maîtrisées. Ensuite, l'évaluation a un rôle de jaugeage qui rend possible l'évaluation formative, l'évaluation normative ou encore, un contrôle des acquisitions. Enfin, l'évaluation a un rôle pronostique qui permet de prédire la réussite dans l'étape suivante : l'évalué a-t-il les connaissances suffisantes pour découvrir la prochaine matière ?

Par ailleurs, selon Detroz (2021), la mesure en éducation est décomposée en trois parties :

- ✓ Le score vrai désigne ce que vaut réellement l'étudiant dans une certaine discipline.
- ✓ L'erreur systématique concerne les erreurs qui impactent tous les étudiants.
- ✓ L'erreur aléatoire varie d'un étudiant à un autre.

Nous retrouvons les biais de l'évaluation dans ces deux types d'erreurs de mesure, que nous développerons dans le chapitre 5 de cette revue de la littérature. Selon Detroz (2021), en diminuant ces deux types d'erreurs, on augmenterait le score vrai. Nous verrons dans le chapitre sur les biais s'il est possible de les éviter et dans le cas échéant, de diminuer les erreurs de mesure.

Illustrons à présent un exemple d'erreur aléatoire. Gjorgjevski (cité par Rot & Butas, 1959) réalise une étude dans laquelle il demande à cinq enseignants de corriger 100 copies. Par la suite, quinze copies ayant reçu la note « bien » sont à nouveau corrigées par quatre autres professeurs. Ces derniers ont attribué des notes différentes pour ces quinze copies allant de médiocre à excellent. Sur les 60 notes attribuées (4 correcteurs x 15 copies), seulement un peu plus de 50% des copies ont gardé la note de base : « bien ». La subjectivité des quatre correcteurs est différente de celle des cinq premiers correcteurs puisqu'ils n'avaient pas le même ensemble de copies à corriger (Rot & Butas, 1959). Les auteurs expliquent ce phénomène : « Tout se passe comme si les correcteurs sériaient les copies en attribuant toujours, quelle que soit leur valeur réelle, les meilleures notes à celles des rangs supérieurs, et les mauvaises notes à celles qui occupent les derniers rangs ». Gjorgjevski (cité par Rot & Butas, 1959) a répété l'expérience à plusieurs reprises et a obtenu les mêmes résultats. En réalité, ces

enseignants cherchent à reproduire une distribution gaussienne et illustrent la loi de Posthumus que nous évoquerons dans le chapitre 5 « Les biais de l'évaluation » de ce travail.

Attardons-nous maintenant sur les différences entre **deux fonctions de l'évaluation : l'évaluation formative (à visée formative) et l'évaluation normative (à visée sélective)**. La première est définie par De Landsheere (1992, p.244) comme suit : « L'évaluation formative consiste fondamentalement à déterminer pour chaque unité ou tâche d'apprentissage dans quelle mesure un élève est maître de la difficulté. Il s'agit donc d'une démarche diagnostique ». L'évaluation formative est une évaluation qui offre un feedback pour l'élève et l'enseignant (Crahay et al., 2019). Allal (2008, p.312 ; cité par Crahay et al., 2019) définit la fonction de l'évaluation formative comme étant « l'adaptation des activités d'enseignement et d'apprentissage afin de favoriser la progression des apprenants vers les objectifs de formation ». Crahay et ses collègues (2019) distinguent trois fonctions de l'évaluation (formative, sommative et pronostique) et préconisent la fonction formative qui a pour but d'aider les élèves dans leurs apprentissages. De Landsheere (1992) poursuit et définit l'évaluation normative comme étant une évaluation dans laquelle on juge la performance d'un élève par rapport à celles des autres élèves. Cette fonction de l'évaluation a un rôle de sélection (Autin et al., 2019). Crahay et al. (2019) ajoutent que la fonction normative, elle, cherche à comparer et à hiérarchiser les élèves entre eux. Ces auteurs expliquent, par ailleurs, que l'évaluation normative laisse à penser que les scores des élèves se distribuent normalement selon la courbe de Gauss. Or, les auteurs poursuivent en expliquant que cette idée est complètement fautive : les classes ne sont pas composées d'un échantillon aléatoire et simple et ne font pas partie de la loi des grands nombres, ce qui ne permet donc pas une distribution gaussienne. Selon Pourtois et ses collègues (1978), la pédagogie normative doit évoluer vers une pédagogie formative afin d'éviter les biais liés à l'origine sociale des élèves.

Par ailleurs, Detroz (2021) définit deux visées de l'évaluation. La première est l'évaluation à visée formative. Elle a pour objectif d'informer l'élève ainsi que son enseignant quant au degré de maîtrise des apprentissages. La seconde concerne l'évaluation à visée sanctionnante. Dans ce contexte, l'évaluation est au service d'une prise de décision. Cette seconde visée de l'évaluation correspond à l'évaluation normative dans la typologie d'Autin et ses collègues (2019).

Finalement, Van Ewijk (2011) annonce que les notes reçues, lors d'évaluation en classe, par les élèves ont une importance capitale dans leur parcours scolaire. Nous mentionnerons les

différentes conséquences de l'évaluation dans la future vie des élèves dans le dernier et sixième chapitre de cette revue de la littérature.

Maintenant que nous avons présenté les différentes fonctions de l'évaluation, intéressons-nous à un public qui semble être, la plupart du temps, désavantagé face à l'évaluation : les élèves issus de l'immigration.

2.2 Les élèves issus de l'immigration et l'école

En Fédération Wallonie-Bruxelles, 22,2% des élèves sont d'origine immigrée², ce qui représente une importante population d'élèves puisqu'en moyenne, les pays de l'OCDE en accueillent 12,6% (Lafontaine et al., 2017). Les pays d'origine de ces élèves sont, entre autres, la France, les pays du Maghreb, les pays d'Afrique (hors Maghreb), les pays d'Europe centrale, d'Europe de l'Est et d'Europe de l'Ouest. Dans ce chapitre, nous nous intéressons aux performances de ces élèves ainsi qu'à leurs parcours scolaires. Nous terminerons par deux hypothèses qui tentent d'expliquer les différences de performances entre les élèves natifs et les élèves issus de l'immigration.

2.2.1 Leurs performances

Lafontaine et al. (2017) constatent, grâce à leurs analyses des enquêtes PISA, que les élèves issus de l'immigration (de première et seconde générations³) présentent des performances scolaires moins élevées que les élèves natifs (au moins un des deux parents est né en Belgique) dans la plupart des pays de l'OCDE. En effet, dans ce même article, les auteurs observent ce résultat dans les trois domaines étudiés par PISA : en mathématiques, en compréhension de l'écrit et en sciences. Selon Ichou (2013), les élèves natifs réussissent mieux à l'école que les élèves immigrés, en termes de résultats scolaires. Sprietsma (2012) explique que, dans un grand nombre de pays industrialisés, les élèves issus de l'immigration ont un retard en termes de performances scolaires dans des enquêtes comme PISA.

En revanche, entre le rapport PISA 2006 et le rapport PISA 2015, l'écart de performances entre ces deux groupes d'élèves (natifs/issus de l'immigration) s'est vu réduit de manière importante dans les trois domaines. Cette diminution des écarts s'explique, notamment, par une amélioration de l'indice du statut économique, social et culturel (SESC) des élèves issus de l'immigration et d'une diminution du SESC des élèves natifs (Lafontaine et al., 2017).

² Élèves nés à l'étranger ou dont les parents sont nés à l'étranger.

³ La première génération concerne les élèves nés à l'étranger et leurs deux parents également. La deuxième génération regroupe les élèves nés en Belgique dont les parents sont tous les deux nés à l'étranger.

2.2.2 Leurs parcours scolaires

Lafontaine et ses collègues (2017) déclarent, sur base de leurs analyses des résultats PISA 2015, que les élèves issus de l'immigration ont davantage connu le redoublement comparativement aux natifs et ce, sous contrôle du niveau socio-économique et de la performance en lecture.

Par ailleurs, certains auteurs se sont intéressés à la perception qu'ont les élèves de leurs compétences et à l'estime de soi. Selon Rangvid (2015), la perception qu'ont les élèves de leurs compétences influence leur probabilité de réussite dans leur parcours scolaire. De plus, le chercheur avance que les performances et les efforts des élèves migrants pourraient en prendre un coup s'ils se voient systématiquement attribuer des scores inférieurs de la part de leur enseignant et cela pourrait nuire à leurs apprentissages. Zafar (2011) annonce que les élèves se basent sur leur moyenne de points antérieure pour leurs attentes quant à leur future moyenne. Rangvid (2015) poursuit en expliquant que les étudiants font des choix éducatifs en fonction de leurs points. L'estime de soi des élèves pourrait être affectée par les faibles notes que les enseignants leur attribuent. Ces faibles résultats scolaires pourraient porter préjudice à la scolarité de ces élèves en modifiant leurs décisions éducatives vers des choix et des filières moins ambitieux. Par conséquent, l'auteur déclare que si les élèves migrants avaient reçu, de la part des enseignants, les mêmes scores que les natifs (élèves ayant des scores similaires aux évaluations externes), les élèves migrants auraient probablement fait des choix plus ambitieux quant aux filières d'enseignement choisies (Rangvid, 2015).

2.2.3 La discontinuité culturelle et la discrimination systémique

Selon Felouzis et son équipe (2015), différentes causes expliqueraient les inégalités entre les élèves natifs et les élèves avec un parcours migratoire. Ils regroupent ces causes en deux grandes catégories d'explication de ces inégalités : la discontinuité culturelle et la discrimination systémique. La première, **la discontinuité culturelle**, concerne les causes ayant des sources familiales, provenant des caractéristiques de la famille. La différence de normes à l'école et à la maison ferait partie d'une des causes de cette catégorie. Les causes inscrites dans cette première catégorie justifient les inégalités envers les élèves issus de l'immigration par l'éloignement culturel de ces élèves. Les inégalités d'acquis entre les élèves natifs et les élèves issus de l'immigration observées dans les enquêtes internationales, comme PISA, seraient la conséquence de l'inégalité culturelle de départ des élèves issus de l'immigration. Ichou (2013) explique les désavantages scolaires par le faible niveau socio-économique ainsi que par le faible niveau d'éducation des parents d'élèves immigrés, ce qui relève du paradigme de la discontinuité culturelle chez Felouzis et al. (2015). Les désavantages scolaires que subissent les

élèves issus de l'immigration comparativement aux natifs peuvent être en partie, expliqués par leur niveau d'éducation, la position sociale des parents et également, la taille de la fratrie (Valet & Caille, 2000, cités par Ichou, 2013). Ichou (2013) rencontre cette même idée qu'il retrouve dans différents travaux (Brinbaum et al., 2010 ; Caille, 1993 ; Simon 2003): la position sociale des parents d'un élève immigré expliquerait les difficultés scolaires de l'enfant.

Ichou (2013) analyse les données du Panel d'écoliers entrés au CP en 1997 et son objectif est d'observer les différences de résultats entre les élèves natifs et les élèves issus de l'immigration (93,1% des enfants d'immigrés de l'échantillon sont nés en France) durant toutes les années de l'école obligatoire (du début de l'école primaire à la fin du collège). Ce « Panel 1997 » a mesuré les résultats scolaires de ces élèves en français et en mathématiques à trois reprises : au CP, en 6^e et en 3^e (Ichou, 2013, p.16). Par la méthode statistique qu'est le matching, Ichou (2013, p.17) cherche à « étudier sociologiquement les scolarités des enfants d'immigrés » en associant chaque enfant d'immigré à un enfant natif. Par cette méthode, il induit des variables contrôles telles que la profession du père, la taille de la fratrie, ... Ces variables ne pourront donc pas expliquer les différences de résultats entre l'élève natif et l'élève d'immigré. Par contre, Mathieu Ichou (2013) annonce tout de même que, en général, les parents des enfants natifs occupent des positions sociales supérieures à celles des enfants d'immigrés. En effet, seulement 4,8% des mères nées en France ont arrêté leurs études au niveau du primaire pour 43,9% des mères immigrées. Au niveau de la profession du père, les pères immigrés sont, pour 81,7% d'entre eux ouvriers et employés et seulement 5,2% sont enseignants ou cadres supérieurs. Pour les pères natifs, on retrouve 21,5% d'enseignants ou cadres supérieurs tandis que 51,1% d'entre eux sont ouvriers et employés. Les enfants d'immigrés ont un score moyen en CP de 60,3% tandis que les enfants natifs ont un score moyen en CP de 70,3%. Dès l'entrée en primaire, le chercheur met en évidence l'existence de différences de résultats entre les natifs et les enfants d'immigrés, à l'avantage de ces premiers. En revanche, l'auteur atteste qu'en prenant en compte les origines sociales et les caractéristiques socio-démographiques, les différences entre enfants natifs et enfants d'immigrés sont réduites. En résumé, Ichou (2013) démontre l'importance de prendre en compte les positions sociales des enfants d'immigrés puisque ce sont celles-ci qui engendrent des désavantages scolaires envers ces enfants. De plus, une fois ces positions sociales contrôlées, l'auteur décrit un amincissement considérable des différences scolaires jusqu'à des différences statistiquement non significatives pour un bon nombre des origines (Maroc, Algérie, Tunisie, Europe du Sud, ...). Comme le souligne Ichou (2013), les résultats

des enfants d'immigrés s'expliquent par la position sociale de leurs parents : ces enfants sont plus nombreux que les natifs à avoir des parents socialement et économiquement plus faibles.

Pourtant, comme l'expliquent Monseur et Baye (2015) à travers différentes analyses de l'enquête PISA 2015, les difficultés scolaires des élèves issus de l'immigration ne peuvent être expliquées uniquement par les paramètres socio-économiques et culturels de ces élèves. D'autres auteurs expliquent ces inégalités par la deuxième catégorie présentée par Felouzis et ses collègues : **la discrimination systémique**, théorie dépassant le paradigme de la discontinuité culturelle. Felouzis et al. (2016) expliquent cette théorie en montrant que l'école n'apporte pas les mêmes opportunités à tous les élèves et ne leur offre pas les mêmes programmes. Ces auteurs exemplifient cette théorie par le fait que les enseignants socioles mieux formés, les conditions les plus propices aux apprentissages sont dédiées aux élèves les plus favorisés. Par ailleurs, en Belgique, les écoles les plus défavorisées socio-économiquement parlant, accueillent les élèves les plus faibles ainsi que les élèves issus de l'immigration (Baye, 2015 ; Lafontaine et al., 2017). De plus, Felouzis et ses collègues (2015) expliquent que les conditions d'apprentissage sont dégradées dans les établissements scolaires accueillant en grand nombre les élèves issus de l'immigration. Dans le cadre de la discrimination systémique, nous ne sommes plus dans une « indifférence aux différences », mais plutôt dans la production d'inégalités par les institutions scolaires (Felouzis et al., 2015, p. 13).

La discrimination systémique est, notamment, présente par la ségrégation scolaire qui mène à des inégalités sociales (Felouzis et al., 2015). La ségrégation ethnique que l'on retrouve dans les collèges tend à diminuer les résultats des élèves issus de l'immigration. On remarque ces ségrégations par le nombre d'élèves issus de l'immigration par classe. L'existence d'école ghettos dans lesquelles les élèves allochtones sont présents en plus grand nombre témoigne d'une ségrégation ethnique (Felouzis, 2003). Pour résumer cette théorie qu'est la discrimination systémique, nous dirons que la qualité de l'éducation offerte à l'école n'est pas la même pour tous les élèves (Felouzis et al., 2016).

La discontinuité culturelle et la discrimination systémique sont deux hypothèses qui s'opposent donc « sur la place et le rôle de l'école dans la formation des inégalités scolaires » (Felouzis et al., 2016, p.16). En comparant les résultats des études PISA 2003 et 2012, Felouzis et ses collègues (2015, p.13) concluent que la théorie de la discontinuité culturelle « n'est pas suffisante pour expliquer l'accroissement des inégalités de compétence liées au parcours migratoires des élèves et/ou de leur famille ». Ce constat suggère donc la présence d'une discrimination systémique dans nos établissements scolaires, dans nos systèmes évaluatifs.

Dans les chapitres 4 et 5, nous verrons comment les enseignants, de manière involontaire et inconsciente pour la plupart d'entre eux, produisent ces discriminations systémiques envers les enfants issus de l'immigration.

Intéressons-nous désormais à un autre public qui semble lui aussi, subir des discriminations systémiques.

2.3 Le poids de l'origine sociale et de l'origine socio-économique des élèves

Les enquêtes internationales telles que PISA témoignent de l'existence d'une relation statistiquement significative entre les ressources culturelles, sociales et économiques de la famille de l'élève et ses performances (Dupriez 2019 ; Monseur & Baye, 2015). Cette relation est très forte dans plusieurs pays, notamment en Belgique (Dupriez, 2019). Selon Yzerbyt et Klein (2019, p.199), l'école est balisée par des valeurs et des règles bien connues des enfants issus d'un contexte favorisé. Or, un enfant défavorisé aura davantage de difficultés et l'arrivée dans le contexte scolaire pourrait être vécu comme « un décalage culturel ».

Goudeau et Croizet (2017) expliquent que les établissements scolaires accroissent l'inégalité sociale en ne la reconnaissant pas. En mettant en avant les différences entre élèves, sans travailler sur leurs origines (un capital social et culturel plus faible pour les élèves défavorisés), l'élève défavorisé et en difficulté se sent incompetent. Dupriez (2019) avance que l'école ainsi que les enseignants ne sont pas impuissants face aux origines sociales des élèves et ont un rôle à jouer quant aux parcours qu'emprunteront ces élèves. L'auteur poursuit en expliquant que certaines pratiques pédagogiques font la différence, en consolidant ces inégalités sociales ou bien en les réduisant.

Pourtois et ses collègues (1978) expliquent que chaque enseignant reçoit des informations concernant la dimension socio-culturelle de ses élèves (par son langage ou bien ses attitudes) qui pourraient l'amener à le classer dans une certaine catégorie. Ces auteurs affirment dans les résultats de leur étude que la variable « appartenance sociale » a très clairement un impact sur la note attribuée par les correcteurs. Dans leur étude, les chercheurs prévoient deux tas de rédactions à corriger. Le premier porte l'étiquette « copies des enfants issus de familles socialement favorisées » et le second « copies des enfants issus de familles socialement défavorisées ». Chacun des quatre correcteurs corrige les deux paquets. Les correcteurs attribuent des notes plus élevées aux enfants issus d'un milieu social favorisé (Pourtois et al.,

1978). Leclercq et son équipe (2004) soutiennent ces propos puisqu'ils expliquent que le jugement d'un enseignant peut dépendre de l'origine sociale de l'élève.

D'après plusieurs auteurs, l'origine sociale de l'élève impacte le jugement évaluatif des correcteurs (Pourtois et al., 1978 ; Sprietsma, 2013). Nous reviendrons sur ce point dans le cinquième chapitre de cette revue de la littérature.

Afin d'éclairer ces différents constats, passons désormais au quatrième chapitre dans lequel nous nous penchons sur l'impact des stéréotypes.

2.4 L'impact des stéréotypes

En préambule, il semble opportun de définir la notion de stéréotype. Dutrévis (2015, p.62) considère que les stéréotypes « présentent des images homogènes, extrêmes, et chargées de valeur. Ces images définissent la place d'un groupe et de ses membres au sein de la société ». Selon Yzerbyt et Klein (2019), en sciences sociales, Walter Lippman (1922) est le premier à avoir introduit la notion de stéréotype. Lippman emprunte cette notion au monde de l'imprimerie qui emploie ce mot pour définir un bloc métallique consacré à reproduire un grand nombre de tirages homologues. Lippman (1922, cité par Yzerbyt & Klein, 2019, p.73), maintient l'idée « d'images dans la tête qui constituent un environnement construit par les gens pour rendre compte de leur réalité sociale ». Yzerbyt et Klein (2019) vont un pas plus loin et avancent que les stéréotypes permettent de placer les individus dans des groupes et ainsi de classer de nombreuses informations efficacement. Les stéréotypes sont donc des croyances qui concernent les caractéristiques d'un individu et qui définissent un groupe (Yzerbyt, 2016).

2.4.1 L'activation des stéréotypes par les enseignants

Abordons désormais les stéréotypes des enseignants à l'égard de certains enfants.

Les concepts d'**endogroupe** et d'**exogroupe** sont avancés par Fresson et Dardenne (2019). Le premier caractérise le groupe d'appartenance d'un individu. Le second, l'exogroupe, c'est l'Autre : c'est le groupe auquel l'individu n'appartient pas et ne se sent pas concerné. La distinction entre ces deux groupes est renforcée par la stigmatisation. Croizet et Leyens (2003, pp.13-14, cités par Fresson et Dardenne, 2019, p.233) spécifient le concept de **stigmaté** qui est « une caractéristique associée à des traits et stéréotypes négatifs qui font en sorte que ses possesseurs subiront une perte de statut et seront discriminés au point de faire partie d'un groupe particulier ; il y aura « eux », qui ont une mauvaise réputation et « nous », les normaux ». Les stigmatés tels que la couleur de peau, l'origine ethnique ainsi que l'origine socio-économique sont présents dans le domaine scolaire. Tajfel (1959, cité par Yzerbyt & Klein, 2019) évoque

la notion de catégorisation. Cette dernière a lieu lorsque l'on accentue les différences entre individus de groupes différents et que l'on minimise les différences entre individus au sein d'un groupe. Yzerbyt et Klein (2019) expliquent que le fait de catégoriser les individus est un besoin afin de représenter et de donner du sens aux différents groupes sociaux qui nous entourent.

Selon Trouilloud et Sarrazin (2003), lorsqu'un individu interagit avec autrui, il développe des attentes envers cette personne. Ces attentes peuvent influencer les comportements et les pensées vis-à-vis de l'interlocuteur (Snyder & Stukas, 1999, cités par Trouilloud & Sarrazin, 2003). En effet, les croyances et les attentes affectent souvent, involontairement, le comportement d'une personne (Hofmann et al., 2005, cités par Sprietsma, 2012).

Selon Sprietsma (2012), les attentes des enseignants allemands envers les élèves d'origine turque sont plus faibles. Les différents comportements des enseignants par rapport à l'origine de leurs élèves ont été étudiés par différents auteurs, notamment par Casteel (1998) et Ferguson (2003) qui ont montré que les élèves afro-américains recevaient moins de félicitations et de compliments de la part de leurs enseignants que leurs camarades d'origines différentes (cités par Sprietsma, 2012). D'après Autin et son équipe (2019), les enseignants auraient des préjugés relatifs aux milieux sociaux qui affecteraient leurs comportements dans une situation d'évaluation. Ces comportements peuvent s'enclencher involontairement. Rangvid (2015) soutient ces auteurs puisqu'il explique que les stéréotypes dont les enseignants sont porteurs peuvent affecter les notes attribuées aux élèves et ainsi, ces stéréotypes ont une influence dans les évaluations des performances des élèves. L'auteur poursuit en précisant que les situations d'examens qui sont corrigées à l'aveugle ne contribuent pas à cette application de stéréotypes car l'évaluateur ne connaît pas l'identité de l'élève. Rangvid (2015) considère que lorsque l'évaluateur est externe, la notation est plus objective. À l'inverse, lorsque l'évaluateur est l'enseignant et corrige la copie de son élève, la note attribuée est alors subjective.

Nous découvrirons dans le cinquième chapitre les biais de l'évaluation qui peuvent apparaître suite à l'existence des stéréotypes.

Nous venons d'aborder les stéréotypes par lesquels les enseignants étaient influencés, mais qu'en est-il de la menace du stéréotype que les enfants peuvent ressentir ?

2.4.2 Sur les élèves

Fresson et Dardenne (2019) expliquent que la menace ressentie par l'individu repose sur ses croyances (réelles ou non) quant au jugement négatif qu'ont les autres individus à son égard et sur son groupe. Ayant peur d'être jugé à travers ce stigmata, l'individu ressent une menace et a

une crainte de confirmer ce stéréotype envers son groupe, ce qui engendre une diminution de ses performances puisqu'il est envahi par ces pensées intrusives (Fresson & Dardenne, 2019). Selon Dutrévis (2015), les effets de **la menace du stéréotype** s'observent tout au long de la scolarité chez les élèves. Ces derniers ont bien souvent conscience des stéréotypes qui tournent autour de leur endogroupe dès l'entrée en primaire, à 6 ans, et en subissent des effets néfastes (Cvencek et al., 2011 ; del Rio et Strasser, 2013 ; McKown et Strambler, 2009 ; Muzzatti et Agnoli, 2007, cités par Fresson & Dardenne, 2019).

L'intériorisation des stéréotypes se manifeste lorsqu'un individu subit des situations de discrimination de manière récurrente, ce qui l'amène à penser que ces images négatives reflètent réellement ses propres compétences (Cooley, 1956, cité par Dutrévis, 2015). Ce processus d'intériorisation a des effets néfastes sur les parcours scolaires des élèves stigmatisés puisqu'ils intègrent ces caractéristiques données à leur groupe d'appartenance (Dutrévis, 2015). Par ailleurs, l'auteur insiste sur l'importance du contexte social, qui est un vecteur supplémentaire de la stigmatisation (Crocker et al., 1998, cités par Dutrévis, 2015).

Merton (1948, p. 195 cité par Trouilloud & Sarrazin, 2003, p.90) parle plutôt de **prophétie autoréalisatrice** et la définit en ces termes : « la prophétie autoréalisatrice est une définition d'abord erronée d'une situation qui suscite un nouveau comportement qui rend exacte cette conception initialement fausse ». Nous aborderons l'effet Pygmalion dans le chapitre suivant.

Dans ce chapitre, nous avons mis en lumière les attentes des enseignants suite à l'activation de leurs stéréotypes ainsi que l'impact de ceux-ci sur les élèves. Intéressons-nous désormais à la section suivante, relative aux biais de l'évaluation que subissent ces élèves porteurs des stéréotypes défavorisés d'un point de vue socio-économique et/ou d'un point de vue ethnique.

2.5 Les biais de l'évaluation

De nombreuses études docimologiques se sont intéressées aux différents biais qui peuvent impacter l'évaluation (Lafontaine & Monseur, 2009). Il existe quatre grandes sources de biais : (1) L'item et les modalités de testing; (2) La copie et le groupe de copies ; (3) L'évaluateur et le groupe d'évaluateurs ; (4) L'évalué et le groupe d'évalués (Detroz, 2021).

Dans le cadre de ce mémoire, nous développerons principalement les biais issus de la dernière source puisque cette thématique est au cœur de notre question de recherche. Ensuite, nous exposerons quelques biais liés aux modalités de testing et à l'évaluateur.

2.5.1 Les biais liés à l'évalué

Comme Bressoux (2012, p.212) l'explique, lorsqu'un enseignant juge la « valeur scolaire » de son élève, il ne se base pas seulement sur ses performances. Le jugement de la « valeur scolaire » est biaisé par certaines caractéristiques appartenant au contexte scolaire mais également, aux caractéristiques des élèves (Bressoux, 2012). Selon De Landsheere (1992), l'évaluateur est influencé par le fait de connaître les résultats antérieurs de l'élève, même s'il ne connaît pas cet élève. L'auteur poursuit en expliquant qu'un premier travail médiocre incite à penser que le prochain le sera également, c'est ce que l'auteur appelle **la stéréotypie** : une contagion des résultats. **L'effet de stéréotypie** consiste à attribuer à un élève les mêmes scores, sans prendre en compte ses progrès, en fonction de ses résultats précédents (Crahay et al., 2019).

Par ailleurs, De Landsheere (1992) évoque **l'effet de halo** : par exemple, le soin et l'écriture de l'élève peuvent avoir un impact dans l'appréciation de ses performances. Ce biais témoigne de l'influence des caractéristiques d'une personne sur la nature et l'orientation d'un jugement évaluatif (De Landsheere, 1992). L'effet de halo a fait l'objet de nombreuses études dans le domaine de la formation d'impression en psychologie sociale. Cet effet peut se caractériser de la manière suivante : « le fait d'avoir une impression favorable à propos d'un trait psychologique sur une dimension nous conduit à attribuer à la personne jugée des traits favorables sur l'autre dimension » (Yzerbyt & Klein, 2019, p.39). D'après Leclercq et ses collègues (2004, p.283), l'effet de halo implique que la note soit « contaminée » par « des caractéristiques de l'étudiant ». Dutrévis (2015) donne un autre exemple de cet effet : les stéréotypes attribués à certains élèves peuvent influencer la note attribuée par l'évaluateur.

Weiss (1969, cité par De Landsheere, 1992) a provoqué l'effet de halo et l'effet de stéréotypie dans ses expériences. Les deux rédactions écrites par des enfants de 4^e primaire ont été dactylographiées sans aucune modification. Avant que les correcteurs ne commencent la correction, il a introduit des préjugés favorables ou défavorables sur les auteurs des copies à corriger :

« Voici deux compositions écrites par des élèves de 4^e primaire. Le travail n°1 est l'œuvre d'un élève moyen qui aime lire les bandes dessinées ; son père et sa mère sont employés. Le travail n°2 a été fait par un enfant doué ; son père est rédacteur d'un quotidien connu. » (De Landsheere, 1992, p.50).

Ainsi, deux groupes de correcteurs ont été formés : le premier groupe a reçu ce mot et le deuxième a reçu un autre mot expliquant la situation inverse (l'élève du travail n°1 est un enfant

doué). Pour les copies ayant reçu un préjugé favorable (n°2 dans le premier groupe de correcteurs, n°1 dans le deuxième groupe de correcteurs), les notes allouées sont significativement supérieures aux autres. Weiss (1969, cité par De Landsheere, 1992) déclare que ces effets seraient d'autant plus grands si l'enseignant connaissait cet élève, son apparence, le statut socio-économique de ses parents, ...

Crahay (2013) soulève la présence d'un « **effet Matthieu** » qui consiste à donner davantage aux élèves les plus favorisés, aux meilleurs élèves, ce qui défavoriserait encore plus les élèves défavorisés. Rosenthal et Jacobson l'ont montré dans leur célèbre étude en 1968, les jugements de la part des enseignants ont d'importantes conséquences sur les performances des élèves (Bressoux, 2012). Ces auteurs ont sélectionné au hasard des élèves et leur ont attribué l'étiquette « prometteurs » qui signifie que ces élèves ont de grandes chances de faire de grands progrès durant l'année. Or, les élèves sélectionnés n'étaient en aucun cas plus prometteurs de progrès que les autres élèves de la classe. Cependant, en fin d'année, les chercheurs constatent que les élèves dits « prometteurs » ont un quotient intellectuel plus élevé que celui des autres élèves de la classe. La prophétie réalisatrice s'est appliquée et les deux auteurs l'ont renommée « **effet Pygmalion** » (Trouilloud & Sarrazin, 2003). Nous savons donc que l'espérance placée en nos élèves influence le développement de leurs capacités (Bressoux, 2012 ; Trouilloud & Sarrazin, 2003 ; Sprietsma, 2012).

2.5.2 Des biais liés aux modalités de testing et à l'évaluateur

Autin et ses collègues (2019) expliquent que l'évaluation à visée sélective est, en réalité, une évaluation normative. Selon ces auteurs, **la fonction donnée à l'évaluation** peut déclencher une évaluation discriminatoire. Ce n'est pas l'évaluation en elle-même qui déclenche l'évaluation biaisée sur les élèves issus des classes sociales les plus faibles. L'évaluation sélective prend appui sur la méritocratie et l'évaluation formative, elle, va de pair avec l'égalité. Quand l'évaluation adopte un rôle de sélection, c'est-à-dire ayant pour but de trier les élèves, ces auteurs remarquent un écart de points entre les élèves issus de différentes classes sociales (Autin et al., 2019). En s'appuyant sur leurs expériences, ces chercheurs affirment que les deux fonctions de l'évaluation (sélection ou formation) peuvent avoir des conséquences en termes d'inégalités sociales. La docimologie ainsi que la sociologie de l'éducation ont démontré un grand nombre de fois l'influence des modalités de l'évaluation sur les résultats obtenus (Linn, 2001 ; Merle, 2007, cités par Ichou, 2013). En effet, lorsque l'élève est en situation d'évaluation normative, sa performance est alors évaluée en référence à celles de ses camarades (Crahay et al., 2019). Gjorgjevski (cité par Rot & Butas, 1959) a illustré ce phénomène, **la loi de**

Posthumus, dans son expérience décrite dans le premier chapitre de cette revue de la littérature. De Landsheere (1980, p. 224, cité par Crahay et al., 2019) définit cette loi comme suit : « un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver d'année en année, approximativement la même distribution (gaussienne) de notes ». Par conséquent, un élève moyen dans une classe faible paraîtra meilleur que ce même élève moyen dans une classe où le niveau est plus fort. Posthumus, en 1947, a dénoncé cet effet qui repose sur une croyance erronée de la part de l'enseignant quant à l'évaluation des performances de ses élèves (Crahay et al., 2019).

Un autre biais semble également important à pointer : **le biais de confirmation**. Yzerbyt et Klein (2019) exposent ce biais qui s'active lorsqu'une personne cherche à confirmer ses hypothèses de départ plutôt que de chercher à les infirmer. Dans ce contexte, les individus sont plus attentifs aux informations qui vont dans le sens de leurs hypothèses et sont davantage disposés à expliquer des éléments ambigus dans ce même sens. L'interprétation de traits observés ultérieurement confirment les hypothèses initiales. Ces dernières viennent colorer l'interprétation d'un comportement observé. En effet, comme nous l'avons développé dans le chapitre précédent, l'enseignant développe des préjugés et des attentes concernant ses élèves, ce qui affecterait les notes attribuées (Autin et al. , 2019; Rangvid, 2015). Ce biais influence donc la manière d'interagir avec la personne victime de ces hypothèses initiales (Yzerbyt & Klein, 2019).

2.5.3 Peut-on éviter ces biais ?

Maintenant que nous avons pris connaissance, de façon non-exhaustive, de différents biais, intéressons-nous à la manière dont nous pouvons ou non les contrer. Frayssinhes (2022) déclare qu'il est possible de limiter les impacts des biais mais qu'il est irréal de penser les supprimer. Certains auteurs affirment qu'il est bon d'être conscient des biais implicites car en rendant compte ces biais, nous pouvons réduire leurs effets (Rudman et al., 2001 ; Frayssinhes, 2022).

Par ailleurs, il serait intéressant de sensibiliser les enseignants à leurs stéréotypes et à ce qu'ils peuvent engendrer (Van Ewijk, 2011). Rangvik (2015) propose de faire passer des tests informatisés qui sont corrigés automatiquement et d'oublier les tests traditionnels manuscrits que l'enseignant corrige. C'est d'ailleurs le cas au Danemark pour des branches comme la géographie et la biologie. Ces tests permettent une évaluation objective complètement à l'aveugle : de fait, l'ordinateur ne juge pas le prénom de l'élève. Selon Rangvik (2015), les conditions pour une correction objective sont une notation externe et à l'aveugle. Des études

plus approfondies sont nécessaires pour identifier quel type de formation pourrait diminuer le biais de notation, notamment, le biais du prénom (Sprietsma, 2012).

Toutefois, même s'il est possible, dans certains cas, de réduire l'effet des biais (Rudman et al., 2001 ; Frayssinhes, 2022), il apparaît impossible de les supprimer (Frayssinhes, 2022). Ce constat nous amène au chapitre suivant, relatif aux conséquences des biais.

2.6 Les conséquences de ces biais

Dans ce sixième et dernier chapitre, nous aborderons dans un premier temps les conséquences des biais sur le comportement des évaluateurs et dans un second temps, les conséquences de ces biais sur l'avenir des élèves.

2.6.1 Le comportement des évaluateurs

Attardons-nous désormais aux comportements des personnes en position d'évaluateur. Selon De Landsheere (1992), un correcteur n'est pas égal à lui-même : une infidélité intra-correcteur est présente et dépend de l'humeur du correcteur, de sa santé mentale et physique et d'un grand nombre de facteurs. Piéron (1962, cité par De Lansheere, 1992) explique que la correction d'une évaluation en français nécessite 78 corrections de correcteurs différents et l'évaluation d'une dissertation philosophique requiert 127 correcteurs différents pour obtenir une note stable. De plus, Sprietsma complète en expliquant que les attentes des enseignants ont un impact plus fort dans la cotation lorsqu'il s'agit d'une évaluation laissant place à la subjectivité, comme par exemple, l'évaluation du comportement en classe (Dee, 2005).

Le prénom peut-il influencer la correction de l'évaluateur ? Sprietsma (2012) a démontré lors de son étude que les notes attribuées par les professeurs d'allemand aux productions d'écrits des élèves ayant un prénom d'origine truque (élève d'origine turque) étaient inférieures aux notes attribuées aux élèves ayant un prénom allemand (élève allemand natif). En effet, cet auteur conclut, suite aux résultats de son étude, que les élèves ayant un prénom d'origine turque reçoivent des notes plus faibles que les natifs. Or, les copies étaient pourtant bien identiques. Rangvid (2015) explique dans son article que la dimension de l'immigration se repère facilement grâce au prénom de l'élève, ce qui peut induire des différences de notation systémiques entre les natifs et les élèves issus de l'immigration, même si l'on ne connaît pas l'élève. A contrario, un évaluateur externe ne pourra définir précisément le niveau socio-économique de l'enfant sur base du prénom et sera donc biaisé dans de moindres mesures par rapport au SES (Rangvid, 2015).

Cependant, une étude réalisée aux Pays-Bas par Van Ewijk (2011) n'a conclu à aucun effet des prénoms (origine immigrée ou natif) sur la note attribuée par les enseignants. Sprietsma (2012) évoque cette étude dans son article et apporte, entre autres, deux explications : soit, ces enseignants ne sont pas biaisés dans leur notation ou bien, ils sont conscients d'une tendance à être influencés par l'ethnicité de l'élève et ils s'y adaptent intentionnellement en y étant vigilants.

Pour son expérience, Sprietsma (2012) s'est basé sur le même schéma expérimental que le chercheur Van Ewijk. Sprietsma (2012) explique que le fait de répliquer des schémas expérimentaux dans différents pays permet d'analyser les différences de résultats entre pays et cette analyse peut révéler des problèmes spécifiques à un pays. Dans ce cadre, Hanna et Linden (2009, cités par Sprietsma, 2012) ont réalisé une étude similaire en Inde et constatent que l'origine des élèves affecte la note attribuée par les correcteurs.

Malgré le fait que, dans son étude, les notations ne soient pas influencées par l'ethnicité de l'élève, Van Ewijk (2011) annonce que les enseignants ont moins d'attentes à l'égard des élèves issus de la diversité. Ce chercheur relève tout de même dans son étude, des comportements chez les enseignants, tels que de faibles attentes envers les minorités ethniques, qui pourraient impacter négativement les performances scolaires de ces élèves. De plus, l'auteur complète en indiquant que ces enseignants ont des attitudes assez négatives envers ces élèves. Sprietsma (2012) annonce que la manière de coter un élève est affectée par les attentes de l'enseignant envers cet élève. Darley et Gross (1983) avancent que les recherches psychologiques ont prouvé qu'une personne cherchera davantage d'indices permettant de confirmer ses attentes que le cas inverse. Comme expliqué supra, Fresson et Dardenne (2019) avancent les concepts d'endogroupe et d'exogroupe. Un enseignant ayant des attentes faibles envers son exogroupe, c'est-à-dire les élèves turcs (Sprietsma, 2012), pourrait avoir tendance à chercher des indices confirmant ses attentes (Darley & Gross, 1983).

L'évaluation subjective des compétences d'un élève faite par un enseignant manifeste des disparités dans les scores et sont à l'avantage des élèves favorisés, c'est-à-dire les élèves natifs ou encore les élèves dont les parents sont instruits (Rangvid, 2015). Sprietsma (2012) annonce que, malgré certaines grilles de notation utilisées par les enseignants pour évaluer leurs élèves, les attentes subjectives des enseignants peuvent impacter l'estimation des performances de ces élèves. De Landsheere (1992) ajoute que l'avis personnel du correcteur ne doit pas intervenir lors de la correction pour obtenir une notation objective. Dans son étude, Rangvik (2015) a pour objectif d'examiner les écarts de notations entre un correcteur subjectif (l'enseignant de l'élève)

et un correcteur objectif (correcteur externe) et d'observer le rôle que ces notations ont dans les décisions éducatives des élèves, comme le choix de filière. Le chercheur va créer des pairs : un autochtone et un enfant migrant ayant les mêmes scores aux évaluations externes. Il compare ensuite les résultats obtenus par l'élève migrant à l'évaluation corrigée par l'enseignant aux résultats homologues de l'élève autochtone (Rangvik, 2015). Il constate que les migrants sont évalués plus négativement de manière systématique par les enseignants que les natifs avec des performances similaires. Crahay et ses collègues (2019) dénoncent, en FWB, un grand arbitraire dans nos systèmes de cotations chez les enseignants puisque leurs exigences envers leurs élèves sont singulières et différentes d'un enseignant à un autre.

Les résultats d'une étude menée par Autin et son équipe (2019) indiquent que lorsque l'évaluateur est en situation d'évaluation pour la sélection, cette personne trouve davantage d'erreurs aux élèves ayant un SES faible qu'aux élèves ayant un SES élevé. Ces chercheurs concluent de leurs expériences que l'évaluation dotée de la fonction de sélection, dans l'absence de différences objectives de performances, conduit les évaluateurs à créer de plus grands écarts de performances entre les élèves SES faibles et SES élevés, ce qui contribue à la reproduction des inégalités sociales. Darley et Gross (1983) ont examiné le processus menant à la confirmation des attentes en attribuant de fausses étiquettes sociales (SES faible et SES élevé) à un enfant. Deux groupes de personnes ont dû évaluer l'enfant : l'un pensant l'enfant doté d'un SES élevé et l'autre pensant que l'enfant a un SES faible. Le premier groupe a évalué les capacités de l'enfant au-dessus de son niveau scolaire et le deuxième groupe, sous son niveau scolaire. Les auteurs ont conclu que certaines informations stéréotypées, comme le SES, créaient de fausses hypothèses sur les individus portant ces informations stéréotypées.

En outre, les résultats obtenus par ces différents auteurs témoignent d'une discrimination envers les élèves d'origine immigrée puisque ceux-ci reçoivent des notes inférieures aux élèves autochtones, même si les performances réelles sont identiques. Autin et al. (2019) soulèvent donc la présence d'une discrimination dans les notes attribuées à l'égard des élèves issus de l'immigration. À performances objectivement identiques, l'enseignant évalue différemment l'élève en fonction de son niveau socio-économique, de son physique, etc (Bressoux et Pansu, 2003, cités par Bressoux, 2012). Or, l'article 6 du Décret Missions prévoit les quatre missions prioritaires, dont celle d'« assurer à tous les élèves des chances égales d'émancipation sociale » (Communauté française, 1997, p.7). En effet, comme le soulignent Lafontaine et Monseur (2009, p. 73), « L'égalité est une valeur centrale en éducation. Tous les élèves [...] doivent être traités de manière égale à l'école. ».

Nous clôturons cette section en rappelant que la plupart des différents travaux présentés témoignent d'une certaine contradiction entre les missions des systèmes scolaires, notamment les articles du Décret Missions qui prévoient une égalité des chances d'émancipation, et les discriminations systémiques produites à l'égard des élèves issus de l'immigration (Felouzis et al., 2015 ; Monseur & Baye, 2015). Toutes ces discriminations envers les enfants issus de l'immigration ont-elles un impact dans l'avenir de ces élèves ?

2.6.2 L'avenir des élèves

Dans leurs études, Sprietsma (2012) et Van Ewijk (2011) demandent aux enseignants de proposer une filière en secondaire pour les élèves évalués : les filières les plus prestigieuses sont plus souvent attribuées aux élèves portant un nom allemand/néerlandais. Sprietsma (2012) en déduit donc que l'estimation, faite par l'enseignant, de la capacité d'un élève à fréquenter une filière plus ou moins prestigieuse est significativement affectée par l'origine du prénom. Dans le même ordre d'idée, les résultats de Rangvid (2015) tendent à indiquer que si les groupes défavorisés (migrant, enfant dont les parents sont peu instruits) avaient des scores égaux aux élèves favorisés (natifs, enfant dont les parents sont très instruits) qui ont des performances similaires, ces élèves défavorisés auraient une plus grande probabilité de s'inscrire à des études plus ambitieuses. Rangvid (2015) complète avec un point déterminant dans l'avenir des élèves : les notes attribuées par les enseignants impactent le choix d'orientation de l'élève qui est basé sur ses résultats scolaires (les scores élevés orientent vers des filières plus exigeantes). En effet, au Danemark, pour accéder à l'enseignement secondaire supérieur, l'élève doit postuler pour accéder à la filière choisie. C'est en fonction des notes attribuées par les enseignants et au certificat attestant que l'élève est capable d'étudier dans une telle filière que cet élève aura accès ou non aux filières. Or, au moment de cette décision, les élèves n'ont passé aucune évaluation externe, ce qui veut donc dire, que c'est le regard subjectif de son enseignant qui détermine, en partie, son avenir. De plus, l'auteur précise que cette décision affecte le niveau final de l'élève, sa réussite ainsi que les taux d'abandon. L'auteur poursuit en indiquant qu'il serait intéressant que ces élèves reçoivent une évaluation externe qui serait corrigée par un évaluateur externe et donc forcément plus objective, avant de devoir faire un choix de filières (Rangvid, 2015).

De plus, les inégalités produites à l'école à l'égard des enfants issus de l'immigration sont grandissantes sur le marché du travail (Silberman & Fournier, 1999, 2006 ; Frickey et al., 2002 ; Meurs et al., 2006, cités par Ichou, 2013). Ces notes attribuées aux élèves pourraient avoir des effets à long terme jusqu'à impacter leurs futurs emplois (Papay et al., 2011, cités par Sprietsma, 2012).

3 OBJET DE RECHERCHE ET HYPOTHESES

3.1 Objet de recherche

Compte tenu des éclairages théoriques exposés supra et des constats posés quant aux biais de l'évaluation liés à l'origine ethnique et à l'origine sociale des élèves, il nous semble opportun d'étudier l'existence d'un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale.

L'objectif de la présente recherche est de mesurer l'effet de l'origine ethnique indépendamment de l'effet du statut socio-économique des élèves sur les notes qui leur sont attribuées par les (futurs) enseignants. Sur base d'un design expérimental, nous testons des enseignants et des futurs enseignants en position d'évaluateurs et nous observons comment ceux-ci se comportent face à la copie d'un élève issu de l'immigration (appelé Ayoub), à la copie d'un élève natif issu d'un milieu socio-économique défavorisé (appelé Jordan) et à la copie d'un élève autochtone issu d'un milieu socio-économique favorisé (appelé Lucien).

Nous analyserons donc les comportements de nos (futurs) enseignants en Fédération Wallonie-Bruxelles (FWB) en situation de correction et observerons s'ils (1) sont influencés dans leurs cotations par l'origine ethnique et/ou l'origine sociale des élèves ; (2) font apparaître, dans leurs cotations, un biais lié à l'origine ethnique indépendant du biais lié à l'origine sociale ; (3) se comportent différemment en fonction de la visée de l'évaluation. Autrement dit, nous tenterons de répondre à la question de recherche suivante :

Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?

3.2 Hypothèses

Suite aux différentes lectures sur le sujet, nos hypothèses de recherches sont dès lors les suivantes :

3.2.1 Hypothèse n°1 (LUCIEN VS JORDAN VS AYOUB)

Tous les élèves ne sont pas égaux face à l'évaluation. A copie identique, le score moyen obtenu par Ayoub, Jordan et Lucien diffère. (Autin et al., 2019 ; Sprietsma, 2012 ; Rangvid, 2015)

3.2.2 Hypothèse n°2 (LUCIEN VS JORDAN)

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »). (Autin et al., 2019 ; Rangvid, 2015)

3.2.3 Hypothèse n°3 (LUCIEN VS AYOUB)

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique »). (Sprietsma, 2012 ; Rangvid, 2015)

3.2.4 Hypothèse n°4 (JORDAN VS AYOUB)

Il existe un biais de l'évaluation lié à l'origine ethnique des élèves indépendant de leur origine sociale et de leur niveau de performance. A copie identique, le score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique ») est inférieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »).

Différents auteurs ont mis en évidence d'une part, un biais lié à l'origine ethnique de l'élève et d'autre part, un biais lié à son origine sociale. A notre connaissance, aucune recherche n'a étudié l'influence conjointe de ces deux biais. Dans ce cas d'indépendance, les évaluations d'Ayoub et de Jordan devraient différer dans le sens décrit ci-dessus.

3.2.5 Hypothèse n°5 (LUCIEN VS JORDAN VS AYOUB)

A copie identique, l'évaluation en vue de la sélection (normative) peut induire une plus grande reproduction des inégalités (entre Lucien, Jordan et Ayoub) que l'évaluation en vue de l'apprentissage (formative). (Autin et al., 2019)

4 METHODOLOGIE

Dans cette partie du travail relative à la méthodologie, nous présenterons tout d'abord le design de la recherche. Ensuite, nous exposerons les instruments de mesure ainsi que les données récoltées. Nous développerons un point sur l'échantillon et le public cible, suivi de la procédure de recrutement. Nous terminerons par les aspects éthiques et enfin, exposerons les analyses statistiques mobilisées dans la section suivante qui concerne les résultats.

4.1 Design de la recherche

L'objectif de la présente recherche est de questionner les processus qui contribuent à l'écart des performances entre les élèves à travers l'évaluation elle-même. À cette fin, le présent mémoire suit un design expérimental visant à mesurer la différence de cotation entre des copies identiques attribuées à un élève natif issu d'un milieu socio-économique favorisé (Lucien), attribuées à un élève natif issu d'un milieu socio-économique défavorisé (Jordan) ou encore attribuées à un élève immigré issu d'un milieu socio-économique défavorisé (Ayoub). Pour ce faire, les participants ont tout d'abord reçu une lettre de consignes générales dans laquelle toutes les démarches à réaliser sont expliquées ainsi que la tâche à effectuer. Quatre lettres différentes ont été écrites et ont été distribuées en fonction du profil du sujet (étudiant ou enseignant) et de la visée de l'évaluation (visée formative ou visée sélective). Ces lettres se retrouvent aux annexes 7, 8, 9 et 10. Pour chacun des sujets, la tâche à effectuer comprend trois productions écrites de trois élèves différents à corriger à l'aide d'une grille d'évaluation : une copie faible, une copie médiane et une copie forte. Nous expliquons dans le point « 4.2.1.1. Les trois rédactions » la procédure de sélection de ces trois copies.

La seule information dont les correcteurs disposent est le prénom de l'élève sur chaque copie (Lucien, Jordan et Ayoub). Il semblerait que le prénom soit un déclencheur puissant d'attentes, encore plus puissant que des informations explicites telles que l'origine de l'élève ou encore son niveau socio-économique (Anderson-Clark et al., 2008 ; Figlio, 2005, cités par Sprietsma, 2012). Le sociologue français Georges Felouzis (2003, p.420) avance que le prénom « apparait comme un marqueur lié à la religion et au pays d'origine ». Dans le cadre expérimental de notre étude, les enseignants ne pourraient être influencés que par le prénom de l'élève puisqu'ils ne le connaissent pas et qu'ils ne possèdent aucune autre information personnelle à son égard.

Chacune des trois copies a été assignée tantôt à un élève porteur d'un stéréotype défavorisé d'un point de vue économique et autochtone (Jordan), à un élève porteur d'un stéréotype défavorisé tant du point de vue ethnique qu'économique (Ayoub) et à un élève porteur d'un

stéréotype favorisé et autochtone (Lucien). Ainsi, 6 groupes ont été organisés : (1) Ayoub fort, Lucien médian et Jordan faible ; (2) Ayoub fort, Jordan médian et Lucien faible ; (3) Lucien faible, Ayoub médian et Jordan fort ; (4) Lucien fort, Jordan faible et Ayoub médian ; (5) Jordan fort, Lucien médian et Ayoub faible ; (6) Jordan médian, Ayoub faible et Lucien fort.

Par ailleurs, deux modalités d'évaluation seront induites. D'une part, certains évaluateurs seront invités à évaluer la copie dans un **contexte sélectif** (prise de décision concernant le redoublement d'un élève). D'autre part, certains évaluateurs seront invités à évaluer la copie dans un **contexte formatif** (identifier les difficultés d'un élève afin qu'il puisse s'améliorer). Par conséquent, douze groupes expérimentaux sont donc constitués : les six groupes décrits ci-dessus sont tantôt évalués dans un contexte formatif (groupes n°1 à 6) et tantôt dans un contexte normatif (groupes n°7 à 12).

Tableau 1. *Organisation des douze groupes expérimentaux*

Groupe expérimental n°	Formative (F) ou Normative (N)	Élève n°1	Élève n°2	Élève n°3
1	F	Ayoub FO 31	Lucien M 29	Jordan FAI 15
2	F	Ayoub FO 31	Jordan M 25	Lucien FAI 19
3	F	Lucien FAI 19	Ayoub M 21	Jordan FO 35
4	F	Lucien FO 39	Jordan FAI 15	Ayoub M 21
5	F	Jordan FO 35	Lucien M 29	Ayoub FAI 11
6	F	Jordan M 25	Ayoub FAI 11	Lucien FO 39
7	N	Ayoub FO 31	Lucien M 29	Jordan FAI 15
8	N	Ayoub FO 31	Jordan M 25	Lucien FAI 19
9	N	Lucien FAI 19	Ayoub M 21	Jordan FO 35
10	N	Lucien FO 39	Jordan FAI 15	Ayoub M 21
11	N	Jordan FO 35	Lucien M 29	Ayoub FAI 11
12	N	Jordan M 25	Ayoub FAI 11	Lucien FO 39
FO = Fort		M = Médian		FAI = Faible

Comme nous pouvons l'observer en rose dans ce tableau pour la copie « Ayoub faible », chacune des neuf copies (3 prénoms x 3 niveaux) apparaît au total quatre fois dans l'ensemble

des groupes expérimentaux : à deux reprises dans un contexte sélectif (évaluation normative) et à deux reprises dans un contexte formatif (évaluation formative). Par ailleurs, nous ne cherchons pas à mesurer l'effet de fatigue ou encore l'ordre des copies. C'est pourquoi, tous les ordres possibles ne sont pas envisagés dans l'organisation des douze groupes expérimentaux.

Intéressons-nous désormais à ces instruments de mesure et à la récolte de données.

4.2 Instruments de mesure et données récoltées

4.2.1 Instruments de mesure

Les instruments de collecte consistent en plusieurs rédactions d'élèves à corriger (annexes 2,3 et 4) ainsi que des grilles de correction à compléter (annexes 5 et 6). La consigne de rédaction destinée à l'élève ainsi que la grille d'évaluation de la tâche sont issues de l'évaluation externe non certificative en Fédération Wallonie-Bruxelles de 2016 destinée aux élèves de 5^{ème} primaire. Cette tâche semble être une base d'expérimentation pertinente puisqu'elle a été pensée et conçue par un groupe de travail composé de conseillers pédagogiques, d'inspecteurs, d'enseignants et d'une chercheuse de l'ULiège.

4.2.1.1 Les trois rédactions

Nous nous sommes rendue dans une classe de 5^e primaire et avons proposé aux vingt élèves présents de réaliser une tâche d'écriture (annexe 0). Nous avons suivi les consignes de passation de test indiquées dans le « dossier de l'enseignant » de l'évaluation externe. Les élèves ont eu 50 minutes pour réaliser cette tâche.

Sur base des 20 rédactions récoltées dans cette classe (annexe 1), nous en avons sélectionnées 6 en concertation avec Madame Baye, Monsieur Pressia et Monsieur Dachet. Nous avons également défini les trois prénoms étudiés dans cette recherche. Ensuite, nous avons demandé à deux enseignants expérimentés de 5^e primaire de sélectionner trois copies sur les six proposées : une copie forte, une copie médiane et une copie faible. Précisons qu'aucun de ces deux évaluateurs n'était titulaire de la classe dans laquelle la tâche d'écriture a été soumise. Autrement dit, les deux instituteurs ne connaissaient pas les enfants qui ont rédigé ces lettres. Ces deux enseignants ont sélectionné ensemble les trois copies en collaborant et en ayant tous les documents nécessaires à leur disposition (consigne, grille de correction). Par conséquent, les trois copies sélectionnées comme instruments de mesure sont des copies jugées comme étant faible, médiane et forte par ces deux enseignants.

Une fois les trois copies sélectionnées, nous avons photocopié trois fois chaque copie afin que les enfants rédacteurs puissent inscrire les trois prénoms étudiés dans cette étude. Chacune de ces copies est tantôt attribuée à Jordan, tantôt à Ayoub et tantôt à Lucien. Finalement, nous avons donc **neuf rédactions à évaluer** (annexes 2A, 2B, 2C, 3A, 3B, 3C, 4A, 4B et 4C).

L'objectif étant de mesurer l'impact du stéréotype défavorisé d'un point de vue socio-économique et du stéréotype défavorisé tant du point de vue ethnique que socio-économique, nous tentons d'activer ce stéréotype implicitement et de manière subtile par le prénom de l'élève rédacteur.

Tableau 2. *Choix des trois prénoms rédacteurs et les stéréotypes susceptibles d'être activés*

Prénoms des rédacteurs :	Stéréotypes susceptibles d'être activés :
Jordan	Défavorisé d'un point de vue socio-économique Élève autochtone
Ayoub (prénom à consonance maghrébine)	Défavorisé tant du point de vue ethnique que socio-économique Élève issu de l'immigration
Lucien	Favorisé d'un point de vue socio-économique Élève autochtone

En outre, les enfants rédacteurs ont été invités à modifier le prénom du destinataire en fonction de l'expéditeur. Ainsi, Ayoub adresse sa lettre à Shahed, Jordan à Kévin et Lucien à Yvan. Nous pensons maximiser l'activation du stéréotype en adaptant le prénom du destinataire dans le même registre de prénoms que l'expéditeur. Tous ces prénoms ont été choisis en concertation avec la promotrice ainsi que le doctorant encadrant ce mémoire.

Par ailleurs, nous avons attribué un numéro à chacune des neuf copies évaluées (annexes 2,3 et 4). Par exemple, le numéro 11 est attribué à la copie « Ayoub faible ». Nous retrouvons ces numéros dans le tableau 1 « *Organisation des douze groupes expérimentaux* » présenté ci-dessus.

Afin de s'assurer de la prise de connaissance des différents prénoms par les correcteurs, nous leur demandons de recopier le prénom de l'élève ainsi que son numéro sur la feuille reprenant la grille de correction.

Enfin, les prénoms attribués aux copies sont tous de genre masculin. Ainsi, nous contrôlons la variable « genre » et les biais liés à celle-ci ne pourront être activés dans cette expérience.

4.2.1.2 Les grilles d'évaluation

Comme expliqué supra, Butera et al. (2019) distinguent l'évaluation en vue de la sélection (normative) et l'évaluation en vue de l'apprentissage (formative). Selon ces auteurs, en mettant en place la première, nous induisons une plus grande reproduction des inégalités. Dès lors, il nous semble important de tester les deux types d'évaluations dans le cadre de la pratique de ce mémoire.

Les deux types d'évaluations sont mobilisés sur base de mêmes rédactions à évaluer. Une grille d'évaluation avec des points à attribuer est construite pour l'évaluation à visée sélective (normative). Nous demandons aux sujets d'entourer un et un seul score par critère pour ensuite, additionner ces différents scores entourés afin d'obtenir un résultat sur 15 points. Pour le deuxième type d'évaluation, l'évaluation à visée formative, les correcteurs reçoivent cette même grille et doivent en plus, ajouter un commentaire pour chaque critère de la grille.

La rédaction de l'élève est évaluée selon quatre grandes compétences : (1) orienter son écrit en fonction de la situation de communication ; (2) élaborer des contenus ; (3) assurer l'organisation et la cohérence d'un texte ; (4) utiliser les unités grammaticales. Ces quatre compétences sont évaluées sous différents critères définis par des indicateurs. Nous retrouvons les grilles d'évaluation aux annexes 5 et 6. Ces grilles ont été remises en page au niveau de la forme mais restent intactes à la source sur le fond, à l'exception de la colonne « commentaires » dans la grille d'évaluation formative qui a été ajoutée.

Par ailleurs, la consigne de correction diffère pour les deux visées de l'évaluation :

Tableau 3. *Les deux visées de correction d'évaluation*

VISÉES :	CONSIGNES :
VISÉE FORMATIVE	L'ENSEIGNANT EST INVITÉ À CORRIGER UNE ÉVALUATION DANS LE BUT D'AIDER L'ÉLÈVE À S'AMÉLIORER.
VISÉE SÉLECTIVE	L'ENSEIGNANT EST INVITÉ À CORRIGER UNE ÉVALUATION DANS LE BUT DE DÉTERMINER SI L'ÉLÈVE EST EN MESURE DE RÉUSSIR SON ANNÉE SCOLAIRE OU BIEN S'IL DOIT RECOMMENCER SON ANNÉE.

Les différentes consignes de correction sont reprises dans les consignes générales (annexes 7,8,9 et 10). Afin de s'assurer de la prise de connaissance des pré-noms sur les trois copies à corriger, nous demandons à chaque sujet de recopier le pré-nom de l'élève ainsi que son numéro attribué sur la grille de correction.

4.2.2 Données récoltées

Les données récoltées sont de nature quantitative et demandent un traitement statistique. Dans le cadre de ce travail, nous analyserons la note totale de chaque production écrite attribuée par chacun des évaluateurs. Nous discuterons de ces éléments dans le point « 4.6. Les analyses statistiques ».

Une seule prise de données est prévue par sujet, d'une durée d'environ 20 à 45 minutes. La récolte de données s'est déroulée en deux phases. La première a débuté en décembre 2022 dans le cadre de deux cours de la Professeure Baye. Les étudiants ont été invités à réaliser la correction de trois copies dans le cadre d'un exercice proposé au cours. Par la suite, pour ceux qui le souhaitaient, nous leur avons proposé de faire partie de notre échantillon en complétant le document de consentement. Nous avons donc comptabilisé ces étudiants volontaires à notre échantillon.

La seconde phase de notre récolte de données a eu lieu durant les mois de février et de mars 2023 et concernait les enseignants du fondamental ainsi que les professeurs de français du secondaire. Chaque participant a reçu un dossier contenant les trois copies à corriger ainsi que les trois grilles d'évaluation et a pu le compléter quand il le souhaitait tout en prenant le temps nécessaire. Dans chaque école, nous avons convenu d'une date à laquelle nous avons récupéré les tests complétés.

Par ailleurs, l'objectif spécifique de l'expérience n'est point révélé aux participants afin que ceux-ci ne soient pas influencés dans leur évaluation. Si l'objectif réel de l'expérience est caché (étudier un éventuel biais de l'évaluation) afin d'éviter d'influencer les évaluateurs, nous ne déformons pour autant pas la visée de la tâche évaluative qui leur est attribuée. En effet, nous mentionnons aux participants que ce projet vise à mesurer les différents facteurs qui interviennent dans l'évaluation d'une rédaction. Dès lors, seuls les facteurs précisément étudiés (l'origine sociale/ethnique de l'élève ainsi que le contexte évaluatif) leur sont dissimulés.

4.3 Public cible et échantillon

Cette recherche sur l'effet du nom des élèves se réalise sur base d'un échantillon de 275 participants afin de garantir des résultats statistiquement valides. Il est nécessaire d'obtenir un nombre d'observations suffisant afin de tenter de répondre aux hypothèses posées. Les participants sont répartis en douze groupes distincts et définis supra, dans le tableau 1. La répartition des 275 sujets dans les douze groupes s'établit comme suit :

Tableau 4. Répartition des 275 sujets dans les douze groupes expérimentaux

Groupe n°	1	2	3	4	5	6	7	8	9	10	11	12
Nombre de sujets par groupe	27	24	25	20	26	26	24	22	20	20	20	21

L'échantillon est scindé en trois grandes catégories. La première concerne les futurs enseignants poursuivant les cours afin d'obtenir le titre d'agrégé de l'enseignement secondaire supérieur (AESS). La deuxième regroupe les étudiants en Master en Sciences de l'éducation, assistants au cours « Introduction aux sciences de l'éducation et de la formation » (ISEF) de la Professeure Baye. Ces étudiants détiennent un diplôme d'enseignant.

Ces deux premières catégories ont passé le test en amphithéâtre lors d'un exercice dans le cadre d'un cours dispensé par la professeure Baye, ils font donc partie de la première phase de récolte de données (décembre 2022).

Enfin, la troisième catégorie concentre les enseignants du fondamental et du secondaire (français) et font partie de la seconde phase de la récolte de données. Nous expliquerons dans le point suivant la procédure de recrutement qui concerne cette catégorie.

De facto, notre échantillon compte 275 sujets et s'organise comme suit :

Tableau 5. Répartition des sujets dans les différentes catégories de l'échantillon

AESS	ISEF	ENSEIGNANTS	TOTAL
N = 106	N = 49	N = 120	275

Des enseignants du fondamental et du secondaire (français), des étudiants en formation initiale (futurs enseignants) et des étudiants en Master en Sciences de l'Education ont été amenés à évaluer trois copies d'élèves (275 x 3 copies évaluées in fine) de 5^e année de l'enseignement primaire. Au total, 825 copies ont été corrigées.

La répartition des sujets dans les différents groupes est strictement aléatoire, ce qui permet d'assurer, avant l'assignation des tests, l'équivalence des groupes (Slavin, 2007, cité par Dachet & Baye, 2021). Dachet et Baye (2021, pp.177-178) nomment ce type d'étude expérimentale comparative « essai contrôlé randomisé ».

Notons que pour la catégorie « enseignants », plus de 200 sujets nous ont contacté afin d'obtenir les documents nécessaires à la récolte de données. Seulement 120 sujets ont rendu les documents complétés.

Préoccupons-nous à présent du recrutement de ces sujets.

4.4 Procédure de recrutement

Quant au recrutement des sujets « enseignants », des lettres d'information ont été envoyées aux directions des écoles fondamentales et secondaires de la province de Namur afin de présenter notre mémoire et en stipulant une recherche de participants. Sur base volontaire, ces enseignants ont pu nous contacter et nous avons ainsi organisé la distribution des tests dans les écoles présentant des enseignants volontaires. Ainsi, nous avons contacté des directions d'écoles jusqu'à l'obtention de notre nombre minimal de participants à notre étude.

Concernant les futurs enseignants en formation initiale ainsi que les étudiants du cours ISEF, nous nous sommes rendus à deux des cours de la professeure Baye afin de leur proposer de participer à notre étude. Sur base volontaire, certains étudiants ont complété un formulaire de consentement et ont rendu un dossier contenant les copies à corriger ainsi que les grilles d'évaluation à utiliser.

Par ailleurs, un message mentionnant l'objet de la recherche ainsi que sa durée a été posté sur les réseaux sociaux faisant appel à des sujets qui seraient volontaires pour participer au test (annexe 12). Nous invitons ces volontaires à nous contacter par message privé afin de leur procurer les documents nécessaires à la passation du test. Tous les participants ont reçu le test en version papier.

4.5 Aspects éthiques

Avant d'entamer toute démarche concernant le prélèvement des données, nous avons dû remplir un dossier et recevoir un avis favorable de la commission de vigilance éthique du département des Sciences de l'éducation de l'Université de Liège. Dans le but de garantir l'éthique dans ce mémoire, l'anonymat et la confidentialité des données sont respectés. Aucun sujet de l'échantillon ne sera reconnaissable à la lecture de ce mémoire : chaque sujet sera numéroté et son identité n'apparaîtra nulle part, ce qui permettra d'assurer l'anonymisation des données. Par ailleurs, la question de l'aveuglement des sujets concernant l'objet de notre recherche a été évoquée dans le dossier.

La commission de vigilance éthique du département des Sciences de l'éducation a apporté un avis favorable en date du 11/11/2022 et autorise la récolte de données dans le cadre de cette recherche (Annexe 13).

Les participants de cette recherche ont complété un formulaire de consentement garantissant l'anonymisation des données récoltées (annexe 14). En ce qui concerne l'ensemble des

documents complétés par les participants en version papier (formulaires de consentement et les grilles de corrections complétées), ils sont conservés dans un lieu sûr pour une durée de deux ans. Une fois cette période écoulée, ces données seront détruites. La promotrice, le doctorant et l'étudiante auront accès à ces données anonymes afin de réaliser les analyses statistiques.

4.6 Les analyses statistiques

Nous récoltons des données quantitatives puisque chaque enseignant attribuera un score à la copie corrigée (De Landsheere, 1992). Nos analyses portent sur le score total attribué à chaque copie de notre échantillon. Au total, 825 copies ont été évaluées (3 copies x 275 sujets).

Dans notre dispositif expérimental, quatre variables sont présentes : trois variables indépendantes, à savoir, le prénom de l'élève (X1), le niveau de la copie (X2) et la visée de l'évaluation (X3) ainsi qu'une variable dépendante, le score de l'élève (Y). Précisons que le seuil de signification est fixé à 0.05, ce qui signifie que les effets observés ont moins de 5 chances sur 100 d'être le fruit du hasard de l'échantillonnage. Grâce à ce seuil, nous déterminerons si les effets observés sont de l'ordre du hasard de l'échantillonnage (>0.05) ou au contraire, si cet effet traduit d'un phénomène réel dans la population de référence (<0.05) (Quertemont, 2023).

En effet, Monseur (2020) annonce que le choix des analyses statistiques dépend de la question de recherche. Dans notre cas, certains tests statistiques sont intéressants afin de tester les hypothèses posées. Ces différents tests sont présentés ci-dessous, pour chacune des hypothèses posées, dans la figure 1.

Figure 1. *Présentation des analyses statistiques pour chacune des cinq hypothèses*

Hypothèse n°1	Hypothèses n°2,3,4 et 5
<ul style="list-style-type: none"> ✓ Indices de tendance centrale ✓ Indices de dispersion 	<ul style="list-style-type: none"> ✓ Indices de tendance centrale ✓ Indices de dispersion ✓ Ampleur de l'effet ✓ Analyse de la variance simple (ANOVA)

Précisons que nos analyses statistiques ont été réalisées à l'aide du logiciel SAS.

4.6.1 Indices de tendance centrale

Tout d'abord, nous nous intéressons à des indices de tendance centrale tels que la **moyenne** et la **médiane**. La moyenne est symbolisée par « *m* » lorsqu'elle est calculée sur les valeurs de

l'échantillon observé (Monseur, 2020). Quant à la médiane, elle permet de séparer l'échantillon en deux parties « de sorte que la moitié des observations lui sont inférieures (ou égales) et la moitié supérieures (ou égales) » (Monseur, 2020, p.4).

Dans un premier temps, en vue de tester les quatre premières hypothèses, nous avons calculé la moyenne des notes attribuées pour chacune des 9 (3 prénoms x 3 niveaux). Nous obtenons donc 9 moyennes. De plus, nous calculons pour chaque copie la médiane, le minimum, le maximum et l'amplitude.

Dans un second temps, en vue de tester l'hypothèse 5, que ce soit pour chacune des neuf copies, nous avons comparé les moyennes des deux évaluations : la moyenne de l'évaluation normative et la moyenne de l'évaluation formative. Nous obtenons donc 18 moyennes (3 prénoms x 3 niveaux x 2 visées de l'évaluation). Pour chacune des neuf copies, nous observons la moyenne attribuée dans le cadre de l'évaluation formative et la moyenne attribuée dans le cadre de l'évaluation sélective.

4.6.2 Indices de dispersion

Ensuite, nous étudions des indices de dispersion tels que **l'écart type** et **l'amplitude**. Le premier indice de dispersion, l'écart-type représente la racine carrée de la variance :

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m_x)^2}$$

Il « est la mesure de dispersion la plus couramment utilisée » (Lafontaine et al., 2019, p.115). Cet indice peut varier de 0 à $+\infty$ mais sa valeur maximale ne dépassera jamais la moitié de l'amplitude des scores observés (Monseur, 2020).

Le second indice de dispersion désigne la distance entre la valeur maximale et la valeur minimale de l'échantillon. L'amplitude ne dépend donc que de deux valeurs et est à interpréter avec précaution.

4.6.3 Ampleur de l'effet

Ensuite, nous avons procédé à une analyse des ampleurs de l'effet. Selon Goudeseune (2018) le g de Hedges et le d de Cohen sont deux mesures de l'ampleur de l'effet très semblables. L'auteur poursuit en expliquant que plus l'ampleur de l'effet est grande, plus il y a de distance entre les courbes de distribution. L'ampleur de l'effet indique la force d'un effet observé, d'une variable sur une autre (Goudeseune, 2018).

Dans le cadre de ce mémoire, nous utiliserons l'analyse du g de Hedges. Il convient de considérer que les valeurs (absolues) allant de 0 à 0.15 comme effet faible, de 0.16 à 0.30 comme effet moyen et au-delà de 0.30, comme effet fort.

Nous calculons des ampleurs de l'effet entre les copies, prises deux par deux, en appliquant la formule suivante :

$$Effet = \frac{Moyenne\ du\ GE - Moyenne\ du\ GC}{Ecart\ -\ type\ moyen\ des\ deux\ groupes}$$

4.6.4 Analyse de la variance simple (ANOVA)

Pour finir, nous avons recourt aux régressions linéaires. Nous avons réalisé une analyse de la variance simple, une ANOVA (ANalysis Of VAriance), à l'aide du logiciel statistique SAS. En effet, comme Monseur (2020) le souligne, nous utilisons une régression lorsque l'on fait l'hypothèse de l'influence de la variable X sur la variable Y ou encore, la prédiction d'un score Y en fonction de la variable X. Dans le cadre de cette analyse, la variable quantitative, expliquée et dépendante correspond à la note totale obtenue par l'élève. Quant aux variables qualitatives, explicatives et indépendantes, elles sont au nombre de trois : le prénom de l'élève (Ayoub vs Jordan vs Lucien), le niveau de la copie (faible vs médian vs fort) et la visée de l'évaluation (formative vs normative). Nous observons donc l'effet respectif de ces variables indépendantes sur le score de l'élève, la variable dépendante.

Y = le score de l'élève

X₁ = le prénom de l'élève

X₂ = le niveau de la copie

X₃ = la visée de l'évaluation

Dans notre cas, nous nous intéressons à l'influence du prénom de l'élève (X₁) sur le score de l'élève (Y), à l'influence du niveau de la copie (X₂) sur le score de l'élève (Y) et à l'influence de la visée d'évaluation (X₃) sur le score de l'élève (Y).

L'équation de la droite de régression :

$$Y = \alpha + \beta_1.(X_1) + \beta_2.X_2 + \beta_3.X_3 + \epsilon$$

$$Y = \alpha + \beta_1.(prénom) + \beta_2.(niveau\ de\ la\ copie) + \beta_3.(visée\ de\ l'éval)$$

Le coefficient α est appelé intercept : c'est l'ordonnée à l'origine, c'est-à-dire la valeur de Y quand X=0. Le coefficient de régression β représente la pente de la droite de régression : «elle correspond à la variation de \hat{Y} associée à l'accroissement d'une unité de X » (Monseur, 2020, p.4).

5 RESULTATS ET INTERPRETATIONS

Passons dès à présent aux résultats et aux interprétations de ce mémoire. Dans cette partie, nous mettons à l'épreuve chacune des cinq hypothèses formulées les unes après les autres. Au travers de ces analyses, nous cherchons tout d'abord à répondre à notre question de recherche par le biais de nos différentes hypothèses posées que nous allons chercher à affirmer ou infirmer. Précisons que l'analyse et le traitement des données récoltées ont été effectués via le logiciel SAS. Pour rappel, nous basons nos analyses sur un échantillon s'élevant à 275 sujets qui ont été amenés à corriger trois copies (825 copies évaluées in fine) de 5^e année de l'enseignement primaire.

Par le biais des résultats des quatre premières hypothèses, nous vérifierons si les stéréotypes subtilement véhiculés à travers le prénom de l'élève rédacteur ont été activés ou non et si ceux-ci ont un impact sur le jugement évaluatif des correcteurs. Au moyen des résultats de l'hypothèse n°5, nous cherchons également à déterminer si la visée de l'évaluation a une incidence sur le score attribué à l'élève. Différents types d'analyses seront présentés à travers les cinq hypothèses : des analyses de moyennes et des écarts-types (indices de tendance centrale et de dispersion), des analyses d'ampleurs de l'effet ainsi qu'une analyse de la variance simple (ANOVA).

Avant de mettre à l'épreuve chacune de nos hypothèses, nous vérifions la **significativité du modèle de l'analyse des variances**. L'ANOVA réalisée en vue de tester les hypothèses 2,3,4 et 5 est basée sur la moyenne des scores totaux. Nous obtenons la valeur F grâce à la formule suivante : $F_{obs} = \frac{CM\ trait}{CM\ erreur}$ qui dans notre cas, correspond au calcul suivant : $\frac{1\ 613.49}{3.14} = 513.76$ Elle obtient donc une valeur de F s'élevant à 513.76 avec une probabilité de dépassement de <0.0001 . Cette probabilité de dépassement étant inférieure au seuil alpha de 5% nous amène à rejeter l'hypothèse nulle selon laquelle les scores attribués ne sont pas influencés par les variables intégrées dans ce modèle et à considérer l'hypothèse alternative qui exprime l'influence de certaines variables de l'analyse de la variance sur la note totale obtenue par l'élève.

Par ailleurs, nous obtenons les carrés moyens du modèle et de l'erreur grâce aux formules

$$suivantes : CM\ trait = \frac{SC\ trait}{dl\ trait} \quad \text{et} \quad CM\ erreur = \frac{SC\ erreur}{dl\ erreur}$$

Pour finir, nous obtenons les degrés de liberté selon les calculs suivants :

$$dl_{total} = N - 1$$

$$dl_{trait} = k - 1$$

$$dl_{erreur} = N - k$$

Tableau 6. *Analyse de la variance simple pour la note totale*

Source	DL	Somme des carrés	Carré moyen	Valeur de F	Pr > F
Modèle (trait)	3	4 840.46	1 613.49	513.76	< 0.0001
Erreur	821	2 578.39	3.14		
Total corrigé	824	7 418.85			

R-carré	Coefficient de variation	Ecart quadratique moyen = Root MSE	Note moyenne
0.65	18.36	1.77	9.65

En outre, le R-carré qui ne peut varier qu'entre 0 et 1, indique une part de la variance de la variable Y qui est expliquée par le modèle (Monseur, 2020). Dans ce cadre, le coefficient de détermination (R^2) de notre modèle s'élève à 0.65, ce qui signifie que les variables indépendantes de ce modèle expliquent 65% de la variance de la variable dépendante, c'est-à-dire le score total de l'élève. En d'autres termes, nous pouvons affirmer que 65% de la variance des scores obtenus par les élèves est expliquée par les variables indépendantes de ce modèle, à savoir le prénom de l'élève, le niveau de la copie et la visée de l'évaluation. 65% des écarts de score sont imputables aux variables présentes dans ce modèle. Nous obtenons une valeur assez élevée du R^2 lorsque la variabilité d'erreur (résiduelle) est assez faible comparativement à la variabilité totale (Monseur, 2020). Le coefficient de R^2 est « le reflet de l'ajustement de la droite de régression aux données » (Monseur, 2020, p.21). Nous pouvons donc en déduire que 35% de la variabilité du score total obtenu par l'élève est expliquée par d'autres variables non incluses dans ce modèle, que nous ne prenons pas en compte dans ce modèle.

5.1 Hypothèse n°1 : LUCIEN VS JORDAN VS AYOUB

Tous les élèves ne sont pas égaux face à l'évaluation. A copie identique, le score moyen obtenu par Ayoub, Jordan et Lucien diffère. (Autin et al., 2019 ; Sprietsma, 2012 ; Rangvid, 2015)

Pour commencer, nous analysons un tableau général reprenant des indices de tendance centrale et des indices de dispersion. Nous pouvons constater dans le tableau 7 présenté à la page suivante que pour les trois niveaux de copies (faible-moyen-fort), le score moyen obtenu par Ayoub, Jordan et Lucien diffère. Pour une même copie, nous n'obtenons pas les mêmes moyennes. En effet, alors qu'Ayoub Fort obtient en moyenne 12.04 points sur 15, Jordan Fort en obtient 12.44 et Lucien Fort 12.63. Pour les copies médianes, nous observons des scores moyens de, respectivement, 11.07, 9.86 et 9.68. Enfin, en ce qui concerne les copies faibles, les scores moyens s'élèvent à, respectivement, 6.16, 6.31 et 6.81.

Tableau 7. *Tableau récapitulatif reprenant le nombre de copies, la moyenne, l'écart-type, la médiane, le minimum, le maximum et l'amplitude pour chacune des 9 copies*

9 copies	N=	Moyenne (écart-type)	Médiane	Minimum	Maximum	Amplitude/ Etendue
Ayoub Faible	93	6.16 (1.96)	6	1	12	11
Ayoub Moyen	85	11.07 (1.75)	11	5	14	9
Ayoub Fort	97	12.04 (1.46)	12	9	15	6
Jordan Faible	91	6.31 (1.57)	6	2	11	9
Jordan Moyen	93	9.86 (1.57)	10	6	14	8
Jordan Fort	91	12.44 (1.56)	13	9	15	6
Lucien Faible	91	6.81 (1.81)	7	2	10	8
Lucien Moyen	97	9.68 (1.94)	10	2	14	12
Lucien Fort	87	12.63 (1.49)	13	8	15	7

Nous analyserons dans les hypothèses 2,3 et 4 le sens de ces différences de moyennes. A ce stade, nous pouvons affirmer que dans notre échantillon, pour une même copie, Ayoub, Jordan et Lucien obtiennent des scores moyens différents.

Comme l'indique le tableau 7, les médianes des différentes copies varient également : de 6 à 7 pour les copies faibles, de 10 à 11 pour les copies moyennes et de 12 à 13 pour les copies fortes. A titre indicatif, les amplitudes, minimums et maximums sont présentés dans ce tableau. Nous ne les étudierons pas étant donné que leurs valeurs dans la distribution dépendent seulement de deux observations. Notons que, quel que soit le prénom, les copies considérées comme faibles obtiennent des scores moyens inférieurs aux copies considérées comme médianes qui elles, obtiennent des scores moyens inférieurs aux copies considérées comme étant fortes. Nous nous attendions à ce résultat puisque comme expliqué dans la méthodologie de ce travail, nous avons volontairement sélectionné trois copies de trois niveaux différents.

Intéressons-nous désormais à la répartition des scores attribués par niveau de copie. Nous analysons tout d'abord les copies faibles, ensuite les copies médianes et pour terminer, les copies fortes. Les graphiques présentés à la page suivante permettent d'observer trois niveaux de la distribution : les percentiles 25, 50 et 75.

En ce qui concerne les **copies faibles** ($N \approx 92$ par prénom), nous observons sur la figure présentée la répartition des scores attribués. Le percentile 50 étant représenté en rouge, nous retrouvons donc la moitié des observations sous cette barre et l'autre moitié, au-dessus de celle-ci. Nous constatons que, pour une même copie, la position du percentile 50 sur ce graphique diffère selon le prénom dans notre échantillon. Il en est de même pour les percentiles 25 et 75 qui ont, eux aussi, des positions différentes selon le prénom.

Figure 2. Répartition des scores attribués aux copies faibles

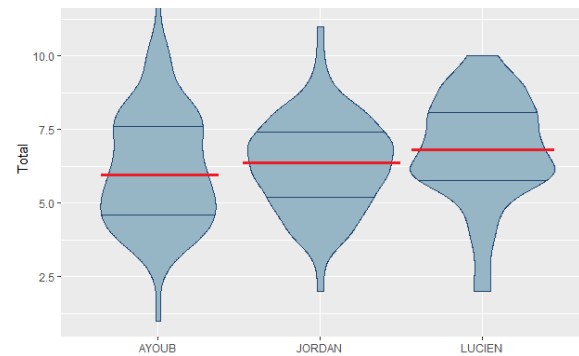
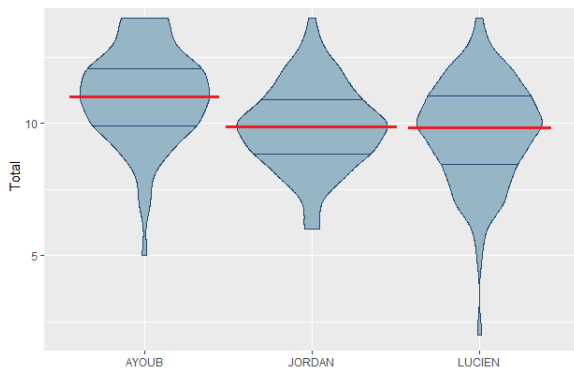


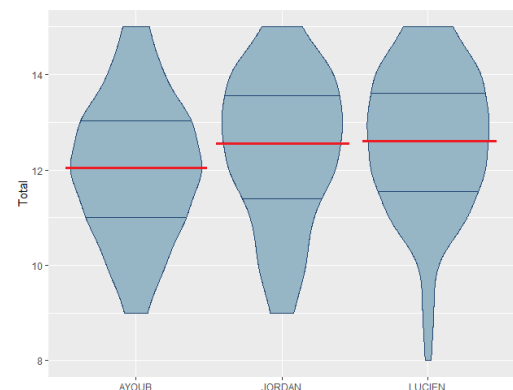
Figure 3. Répartition des scores attribués aux copies médianes



Pour les **copies médianes** ($N \approx 92$ par prénom), nous observons un constat similaire : pour une même copie, les différents quartiles de la distribution (percentiles 25, 50 et 75) ont des scores différents en fonction du prénom. Remarquons tout de même que les positions du 50^e percentile de Jordan et de Lucien sont assez proches.

Les **copies fortes** ($N \approx 92$ par prénom) obtiennent également des scores différents selon le prénom pour les percentiles 25, 50 et 75. Nous pouvons remarquer que les copies fortes et les copies faibles ont un point commun au niveau du 25^e percentile : Ayoub obtient le score le moins élevé, Lucien obtient le score le plus élevé et quant à Jordan, il se situe entre les deux.

Figure 4. Répartition des scores attribués aux copies fortes



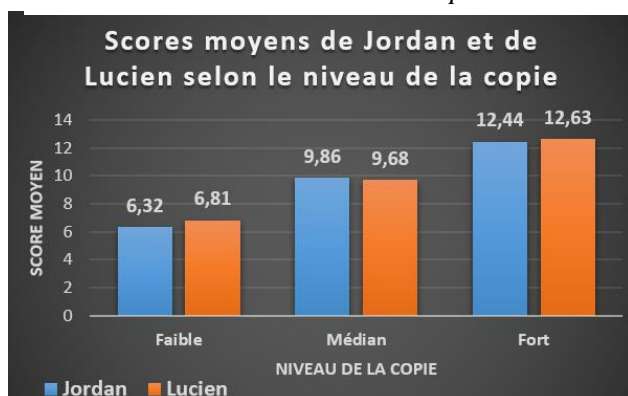
Pour clôturer l'analyse de cette première hypothèse, nous pouvons affirmer que dans notre échantillon, pour une même copie, Ayoub, Jordan et Lucien obtiennent des scores moyens différents. Nous avons constaté des variations dans les scores moyens obtenus par Ayoub, Jordan et Lucien dans notre échantillon. A travers les trois prochaines hypothèses, nous nous intéresserons aux sens de ces variations. Nous observerons également les ampleurs de l'effet qui détermineront la force des effets constatés dans notre échantillon. De plus, nous interpréterons l'analyse de la variance qui permettra de déterminer si ces résultats sont transférables dans la population de référence ou non.

5.2 Hypothèse n°2 : LUCIEN VS JORDAN

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »). (Autin et al., 2019 ; Rangvid, 2015)

Dans le but de mettre à l'épreuve cette deuxième hypothèse, nous allons recourir à différents types d'analyses. Pour commencer, analysons les scores moyens obtenus par les élèves.

Figure 5. Scores moyens de Jordan et de Lucien selon le niveau de la copie



Comme l'indique ce graphique, nous constatons que les copies faibles et fortes obtiennent un meilleur score à l'avantage de Lucien et nous observons le phénomène inverse pour les copies médianes. Ces différences de moyennes s'élèvent donc à 0.39 point à l'avantage de Lucien pour les copies faibles, à 0.18 point à l'avantage de Jordan pour les copies médianes et à 0.19 à l'avantage de Lucien pour les copies fortes.

L'analyse des ampleurs de l'effet, présentée dans le tableau suivant, témoigne de l'importance de ces différences et va également dans ce sens puisque pour les copies faibles et fortes, les copies de Jordan sont, en moyenne, moins bien évaluées que celles de Lucien, avec des ampleurs de l'effet de, respectivement, -0.29 et -0.12. Pour les copies faibles, nous obtenons une ampleur de l'effet assez importante, considérée comme effet moyen à fort.

Tableau 8. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Lucien selon le niveau de la copie

Prénom / Copie	Jordan (2)		Lucien (3)		2 VS 3 / Ampleur de l'effet
	N	Moyenne (ET)	N	Moyenne (ET)	
Faible	91	6.32 (1.57)	91	6.81 (1.81)	-0.29
Médiane	93	9.86 (1.57)	97	9.68 (1.94)	+0.10
Forte	91	12.44 (1.56)	87	12.63 (1.49)	-0.12

Les copies faibles attribuées à Jordan (M=6.32) ont été sous-évaluées comparativement aux copies faibles attribuées à Lucien (M= 6.81) avec une ampleur de l'effet de -0.29 qui est considérée comme un effet étant moyen à fort. Nous observons le même phénomène pour les copies fortes avec un score moyen de 12.44 pour Jordan et de 12.63 pour Lucien, avec une ampleur de l'effet relativement faible de -0.12. En ce qui concerne les copies médianes, nous

observons le phénomène inverse : la copie de Jordan est surévaluée comparativement à la copie de Lucien, avec une ampleur de l'effet de +0.10 qui est relativement faible.

Afin d'offrir une analyse la plus complète possible de cette deuxième hypothèse, nous avons réalisé un autre type d'analyse qui porte sur la variance simple (ANOVA). Dans le cadre de cette hypothèse, nous nous intéresserons seulement aux variables « prénom » (PRE) et « niveau de la copie » (NIV) de ce tableau. Nous poursuivrons l'analyse de la variance de la variable « visée de l'évaluation » (VIS) dans le point relatif à l'hypothèse n°5.

Tableau 9. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale

	DL	Carré moyen	Valeur de F	Pr > F
Prénom (PRE)	1	0.0001	0.00	0.99
Niveau de la copie (NIV)	1	4 835.61	1 539.73	< 0.0001
Visée de l'évaluation (VIS)	1	3.80	1.21	0.27

Précisons que plus la valeur de F s'éloigne de 1, plus l'hypothèse nulle est improbable et pourra être rejetée (Quertemont, 2023). Tout d'abord, ce tableau reprenant l'analyse de la variance simple sur la note totale indique un effet statistiquement non significatif de la variable « prénom » qui obtient une valeur F de 0 et qui ne diffère donc pas statistiquement de 0. Cette variable dépasse largement le seuil alpha de 5% avec une probabilité de dépassement de 0.99. En d'autres mots, nous ne pouvons pas affirmer qu'il existe un effet du prénom sur le score total obtenu par l'élève dans la population de référence. Les résultats de cette variable nous amènent donc à conserver l'hypothèse nulle ($0.99 > 0.05$).

Ensuite, en ce qui concerne l'influence de la variable indépendante « niveau de la copie » sur le score total de l'élève, nous obtenons une valeur F de 1 539.73 avec une probabilité de dépassement de <0.0001 inférieure au seuil alpha (5%). Ces résultats nous amènent à rejeter l'hypothèse nulle et à envisager l'hypothèse alternative selon laquelle l'effet du niveau de la copie sur le score total traduit un phénomène réel dans la population de référence. Nous pouvons donc affirmer que le niveau de la copie exerce une influence significative sur le score total attribué à la rédaction de l'élève. En effet, nous avons observé ce même résultat dans l'analyse des moyennes du tableau récapitulatif 7 présenté supra.

Dans notre échantillon, nous pouvons donc constater un score moyen plus élevé à l'avantage de Lucien pour les copies faibles et fortes, avec une ampleur de l'effet relativement importante pour les copies faibles (-0.29) et peu élevée pour les copies fortes (-0.12). Par ailleurs, les copies médianes reçoivent, en moyenne, un score plus élevé à l'avantage de Jordan avec une

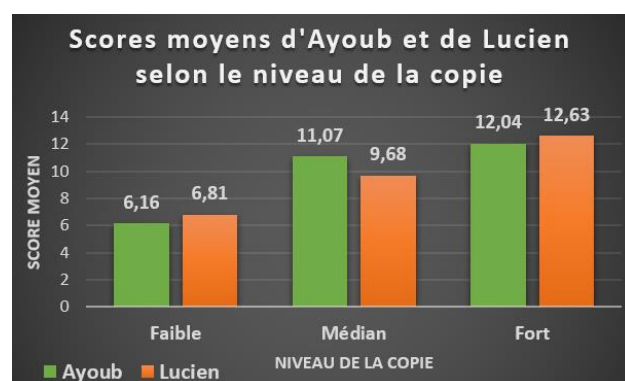
ampleur de l'effet assez faible (+0.10). Toutefois, l'analyse de la variance de la variable « PRE » révèle un résultat non significatif, ce qui ne nous permet pas d'affirmer que, pour une même copie, le prénom de l'élève a un effet sur son score moyen dans la population de référence.

5.3 Hypothèse n°3 : LUCIEN VS AYOUB

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique »). (Sprietsma, 2012 ; Rangvid, 2015)

En nous penchant tout d'abord sur l'analyse des moyennes présentées sur la figure 6, nous constatons que les copies fortes et faibles sont en moyenne mieux notées lorsqu'elles sont attribuées à Lucien (12.63 et 6.81) que quand elles sont attribuées à Ayoub (12.04 et 6.16). L'inverse est observé pour les copies considérées comme médianes : Lucien obtient une moyenne de 9.68 alors qu' Ayoub, lui, obtient 11.07.

Figure 6. Scores moyens d'Ayoub et de Lucien selon le niveau de la copie



Ces différences de moyennes s'élèvent donc à 0.65 point à l'avantage de Lucien pour les copies faibles, à 1.39 point à l'avantage d'Ayoub pour les copies médianes et à 0.59 à l'avantage de Lucien pour les copies fortes. Les résultats mis en avant au sein de cette figure rejoignent ceux développés dans l'hypothèse précédente quant aux sens des différences de moyennes.

Tableau 10. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Lucien et Ayoub selon le niveau de la copie

Prénom \ Copie	Ayoub (1)		Lucien (3)		1 VS 3
	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet
Faible	93	6.16 (1.96)	91	6.81 (1.81)	-0.34
Médiane	85	11.07 (1.75)	97	9.68 (1.94)	+0.75
Forte	97	12.04 (1.46)	87	12.63 (1.49)	-0.40

En ce qui concerne les amplitudes de l'effet qui témoignent de la grandeur de ces effets, les trois sont considérées comme étant des effets forts (-0.34 et -0.40) voire extrêmement fort (+0.75) dans le cas des copies médianes.

Par ailleurs, comme nous l'avons annoncé dans le cadre de l'analyse de l'hypothèse 2, l'influence de la variable « prénom » a un effet non significatif ($0.99 > 0.05$), ce qui nous empêche d'affirmer qu'il existe un effet du prénom sur le score total obtenu par l'élève dans la population de référence.

5.4 Hypothèse n°4 : JORDAN VS AYOUB

Il existe un biais de l'évaluation lié à l'origine ethnique des élèves indépendant de leur origine sociale et de leur niveau de performance. A copie identique, le score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique ») est inférieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »).

Les résultats présentés au sein de cette figure indiquent que les copies faibles de Jordan obtiennent un score moyen plus élevé (6.32) que celles d'Ayoub (6.16). Nous observons un constat similaire pour les copies fortes puisque Jordan obtient un score moyen de 12.44 alors qu'Ayoub obtient seulement 12.04 points. En revanche, l'inverse est observé pour les copies médianes. Ayoub obtient un meilleur score s'élevant à 11.07 comparativement à Jordan qui n'obtient que 9.86 points. Ces différences de moyennes s'élèvent donc à 0.16 point à l'avantage de Jordan pour les copies faibles, à 1.21 point à l'avantage d'Ayoub pour les copies médianes et à 0.40 à l'avantage de Jordan pour les copies fortes. À nouveau, les résultats mis en avant au sein de cette figure rejoignent ceux développés dans les deux hypothèses précédentes quant aux sens des différences de moyennes.

Figure 7. Scores moyens de Jordan et de Ayoub selon le niveau de la copie

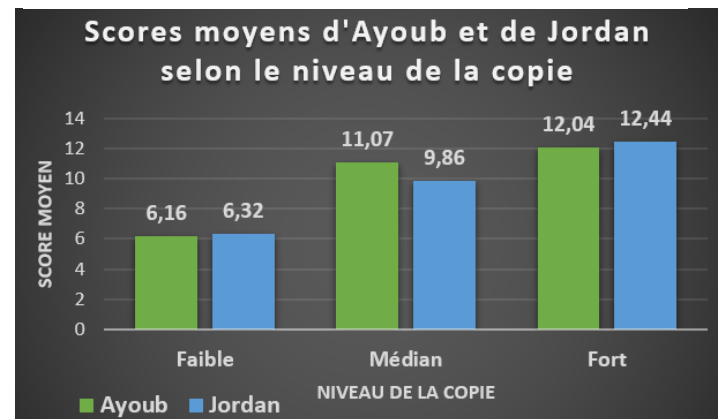


Tableau 11. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Ayoub selon le niveau de la copie

Prénom / Copie	Ayoub (1)		Jordan (2)		1 VS 2
	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet
Faible	93	6.16 (1.96)	91	6.32 (1.57)	-0.09
Médiane	85	11.07 (1.75)	93	9.86 (1.57)	+0.73
Forte	97	12.04 (1.46)	91	12.44 (1.56)	-0.26

En d'autres termes, les copies faibles et fortes obtiennent une moyenne plus élevée lorsqu'elles sont attribuées à Jordan plutôt qu'à Ayoub. Par contre, pour les copies médianes, la moyenne obtenue par Ayoub est supérieure à la moyenne obtenue par Jordan. Les ampleurs de l'effet présentées dans le tableau 11 suivent ce raisonnement et s'élèvent à -0.09 pour les copies faibles, à +0.73 pour les copies moyennes et à -0.26 pour les copies fortes.

Comme nous l'avons déjà expliqué dans le cadre de l'analyse des deux hypothèses précédentes, l'analyse de la variance simple sur la note totale révèle un effet non significatif de la variable « prénom ». Nous ne pouvons donc pas généraliser nos résultats à la population de référence.

5.5 Hypothèse n°5 : LUCIEN VS JORDANS VS AYOUB

A copie identique, l'évaluation en vue de la sélection (normative) peut induire une plus grande reproduction des inégalités (entre Lucien, Jordan et Ayoub) que l'évaluation en vue de l'apprentissage (formative). (Autin et al., 2019)

En vue de tester la cinquième hypothèse, nous avons calculé la moyenne des notes attribuées pour chacune des 9 copies tant pour l'évaluation formative que pour l'évaluation normative (3 prénoms x 3 niveaux x 2 visées de l'évaluation).

Tableau 12. *Moyenne (/15) et écart-type des notes et ampleur de l'effet pour chacune des copies selon la visée de l'évaluation*

Copie	Normative (1)		Formative (2)		Ampleur de l'effet (1) vs (2)
	Moyenne (ET)	N	Moyenne (ET)	N	
Ayoub faible	5.80 (2.08)	41	6.44 (1.84)	52	-0.33
Jordan faible	6.29 (1.49)	44	6.34 (1.66)	47	-0.03
Lucien faible	6.81 (1.97)	42	6.82 (1.68)	49	-0.005
Ayoub médian	11.07 (1.38)	40	11.07 (2.04)	45	0
Jordan médian	9.98 (1.76)	43	9.76 (1.39)	50	+0.14
Lucien médian	9.48 (1.99)	44	9.85 (1.89)	53	-0.19
Ayoub fort	11.83 (1.27)	46	12.23 (1.61)	51	-0.27
Jordan fort	12.35 (1.67)	40	12.51 (1.49)	51	-0.10
Lucien fort	12.73 (1.60)	41	12.54 (1.39)	46	+0.13

En analysant le tableau présenté ci-dessus, nous remarquons que pour une même copie, les scores obtenus dans le cadre d'une évaluation normative sont inférieurs aux scores obtenus dans le cadre d'une évaluation formative, à l'exception de deux copies (Jordan médian et Lucien fort). Cela étant, nous obtenons des ampleurs de l'effet très faibles voire inexistantes pour la majorité des copies, ce qui se traduit par une absence de grande différence de moyenne entre la

visée « normative » et la visée « formative ». Par contre, les copies « Ayoub faible » et « Ayoub fort » obtiennent des ampleurs de l'effet respectivement de -0.33 et de -0.27. Ces deux effets, considérés comme, respectivement, fort et moyen, témoignent d'une différence de scores moyens entre les deux visées de l'évaluation.

Maintenant que nous avons, pour une même copie, comparé les différences de moyennes entre les deux visées de l'évaluation, attardons-nous sur les différences de moyennes entre prénoms, pour une même copie et pour une même visée de l'évaluation.

Tableau 13. *Différence de moyenne et ampleur de l'effet pour les deux visées de l'évaluation*

	Normative		Formative	
	Différence de moyenne (1) – (2)	Ampleur de l'effet (1) VS (2)	Différence de moyenne (1) – (2)	Ampleur de l'effet (1) VS (2)
COPIES FAIBLES				
Jordan (1) VS Lucien (2)	-0.52	-0.30	-0.48	-0.29
Ayoub (1) VS Lucien (2)	-1.01	-0.50	-0.38	-0.21
Ayoub (1) VS Jordan (2)	-0.49	-0.24	+0.10	+0.06
COPIES MÉDIANES				
Jordan (1) VS Lucien (2)	+0.50	+0.27	-0.09	-0.05
Ayoub (1) VS Lucien (2)	+1.59	+0.94	+1.22	+0.62
Ayoub (1) VS Jordan (2)	+1.09	+0.69	+1.31	+0.77
COPIES FORTES				
Jordan (1) VS Lucien (2)	-0.38	-0.23	-0.03	-0.02
Ayoub (1) VS Lucien (2)	-0.90	-0.63	-0.31	-0.21
Ayoub (1) VS Jordan (2)	-0.52	-0.35	-0.28	-0.18

Tout d'abord, en situation d'évaluation normative, nous observons des différences de moyennes plus élevées comparativement aux différences de moyennes dans le cadre d'une évaluation formative pour toutes les comparaisons présentées dans ce tableau à l'exception d'une, celle qui concerne Ayoub et Jordan « médians ».

Ensuite, nous pouvons effectuer le même constat quant aux ampleurs de l'effet : elles sont plus élevées en situation d'évaluation normative qu'en situation d'évaluation formative, à l'exception de la comparaison qui concerne Ayoub et Jordan « médians ».

Dans le cadre de cette étude, deux résultats semblent intéressants à mettre en évidence. Observons deux comparaisons plus en détail : Jordan Fort VS Lucien Fort et Ayoub Faible VS

Jordan Faible. Pour ces deux duos de comparaison, nous obtenons des différences négligeables entre les copies dans un contexte formatif. En revanche, lorsque l'évaluation est à visée sélective, nous remarquons des différences de moyennes entre les deux copies avec des ampleurs de l'effet moyennes, au désavantage de Jordan Fort dans la première comparaison et au désavantage d' Ayoub Fort dans la seconde. Nous incluons ces résultats dans les hypothèses concernées dans la section suivante, consacrée à la discussion.

Enfin, remarquons que toutes les ampleurs de l'effet calculées dans le cadre d'une évaluation normative ont des effets moyens à extrêmement forts.

Poursuivons désormais l'analyse de ce tableau, débutée dans la section dédiée à l'hypothèse 2 pour les deux premières variables.

Tableau 9. *Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale*

	DL	Carré moyen	Valeur de F	Pr > F
Prénom (PRE)	1	0.0001	0.00	0.99
Niveau de la copie (NIV)	1	4 835.61	1 539.73	< 0.0001
Visée de l'évaluation (VIS)	1	3.80	1.21	0.27

Pour finir l'analyse de ce tableau, nous observons une valeur F s'élevant à 1.21 pour la variable indépendante « visée de l'évaluation » avec une probabilité de dépassement de 0.27 supérieure au seuil alpha fixé à 5%, ce qui nous amène à conserver l'hypothèse nulle. En d'autres termes, nous ne pouvons pas affirmer qu'il existe un effet de la visée de l'évaluation sur le score total obtenu par l'élève dans la population de référence. Les résultats de cette variable nous amènent donc à conserver l'hypothèse nulle ($0.99 > 0.05$).

Intéressons-nous désormais à la régression linéaire présentée dans le tableau 14, expliquant le score de l'élève par trois variables indépendantes introduites simultanément. La première concerne la variable « prénom » (PRE) dont la valeur 1 correspond à Ayoub, la valeur 2 correspond à Jordan et la valeur 3 correspond à Lucien. La deuxième désigne la valeur de la copie et correspond au « niveau de la copie » (NIV) dont la valeur 1 équivaut à la copie faible, la valeur 2 à la copie moyenne et la valeur 3 à la copie forte. Enfin, la troisième variable correspond à la « visée de l'évaluation » (VIS) dont la valeur 1 correspond à l'évaluation formative et la valeur 0 correspond à l'évaluation normative. Le score de l'élève est donc régressé sur trois variables indépendantes.

Tableau 14. Régression linéaire de la note totale selon le prénom, le niveau de la copie et la visée de l'évaluation

Paramètre	Estimation	Erreur standard	Valeur t	Pr > t
Intercept	3.65	0.23	15.62	< 0.0001
Prénom (PRE)	-0.0005	0.07	-0.01	0.99
Niveau de la copie (NIV)	2.96	0.07	39.24	< 0.0001
Visée de l'évaluation (VIS)	0.14	0.12	1.10	0.27

Le score de l'élève est donc expliqué par le prénom, le niveau de la copie et la visée de l'évaluation dans ce modèle. L'équation de la droite de régression se présente comme suit :

$$Y = \alpha + \beta_1.(X_1) + \beta_2.(X_2) + \beta_3.(X_3) + \epsilon$$

$$\text{Score de l'élève} = 3.65 - 0.0005 (\text{PRE}) + 2.96 (\text{NIV}) + 0.14 (\text{VIS})$$

Tout d'abord, cette régression linéaire indique qu'indépendamment du niveau de la copie et de la visée de l'évaluation, une augmentation d'une unité de la variable « prénom » est associée à une diminution minime du score de l'élève de 0.0005 point. Etant donné cet effet très faible et proche de 0, nous pouvons considérer que, dans le cadre de ce test statistique, le prénom de l'élève n'exerce pas d'influence sur le score total obtenu par l'élève. De plus, la statistique t pour cette variable vaut -0.01 avec une probabilité de dépassement s'élevant à 0.99, supérieure au seuil alpha (5%), ce qui nous amène à conserver l'hypothèse nulle selon laquelle le prénom n'influence pas le score attribué par les correcteurs.

Ensuite, le coefficient de régression associé au niveau de la copie (β_2) est égal à 2.96. Ainsi, indépendamment du prénom et de la visée de l'évaluation, une augmentation d'une unité de la variable niveau de la copie (NIV) est associée à un accroissement du score de l'élève de 2.96 points : plus le niveau de la copie est élevé, plus l'élève obtient de meilleurs points. A prénom équivalent et pour une même visée de l'évaluation, la copie considérée comme médiane a un score plus élevé que la copie faible. En effet, la copie médiane obtient 2.96 points de plus (2 x 2.96) que la copie faible (1 x 2.96). Le fait d'avoir une copie faible est donc associé à de moins bons scores. Ainsi, l'écart de score, sous contrôle du prénom et pour une même visée de l'évaluation, entre la copie faible et la copie forte s'élève à 5.92 (8.88 – 2.96) à l'avantage de la copie forte. Nous pouvons dès lors affirmer que cette différence de points est statistiquement significative puisque pour la variable NIV, la statistique T vaut 39.24 avec une probabilité de dépassement de < 0.0001, inférieure au seuil alpha de 5%. Ceci étant dit, nous pouvons donc rejeter l'hypothèse nulle. Il existe bien un effet réel, dans la population de référence, du niveau

de la copie sur le score de l'élève. Nous pouvons donc considérer que dans notre échantillon, le score total de l'élève est influencé, de manière significative, par le niveau de la copie.

Notons que, dans l'analyse des moyennes présentées dans le tableau 7, nous avons remarqué que les copies considérées comme faibles obtiennent des scores moyens inférieurs aux copies considérées comme médianes qui elles, obtiennent des score moyens inférieurs aux copies jugées comme fortes. Pour les deux visées de l'évaluation confondues, les copies faibles d'Ayoub, Jordan et Lucien, obtiennent respectivement des scores moyens de 6.16, 6.31 et 6.81. De même, pour les copies médianes, nous observons respectivement des scores moyens de 11.07, 9.86 et 9.68. Enfin, pour les copies fortes, les scores moyens s'élèvent respectivement à 12.04, 12.44 et 12.63.

Enfin, sous contrôle du prénom et du niveau de la copie, une augmentation d'une unité sur la variable « visée de l'évaluation » entraîne à un accroissement de 0.14 point sur le score de l'élève. Il apparait pour la statistique t une valeur de 1.10 avec une probabilité de dépassement de 0.27. Autrement dit, en ce qui concerne cette probabilité de dépassement, nous dépassons largement le seuil des 5% et sommes donc amenée à conserver l'hypothèse nulle.

Prenons l'exemple de la copie d'Ayoub Fort dans le cadre d'une évaluation formative et mettons-le en lien avec l'équation de notre droite de régression :

$$\text{Score de l'élève} = 3.65 - 0.0005 (\text{PRE}) + 2.96 (\text{NIV}) + 0.14 (\text{VIS})$$

$$\text{Score obtenu par Ayoub Fort} = 3.65 - (0.0005 \cdot 1) + (2.96 \cdot 3) + (0.14 \cdot 1) = 12.67$$

L'évaluation formative pour la copie d'Ayoub Fort obtient un score de **12.67/15**.

6 DISCUSSION

Dans cette partie qui concerne la discussion de ce mémoire, nous discuterons, dans un premier temps, les hypothèses formulées et les résultats obtenus au regard de la littérature sur les thématiques développées dans ce mémoire. Dans un second temps, nous présenterons les limites de cette recherche ainsi que les perspectives envisagées.

6.1 Discussion autour des différentes hypothèses

Selon les différentes hypothèses formulées au point 3.2. du présent mémoire, cette recherche permet donc de questionner l'existence d'un biais de l'évaluation lié à l'origine ethnique de l'élève (à copie équivalente, Ayoub devrait être sous-évalué comparativement à Lucien) indépendant du biais lié à son origine sociale (dans ce cas d'indépendance, les évaluations d'Ayoub et de Jordan devraient différer) en contextes d'évaluation sélective et d'évaluation formative. La mise en relation de ces trois éléments est d'ailleurs l'originalité de la présente recherche. Si les études antérieures ont déjà mis en évidence des biais de l'évaluation impactant les résultats d'élèves défavorisés (Rangvid, 2015), les résultats d'élèves issus de l'immigration (Sprietsma, 2012), ainsi que l'équité dans des contextes évaluatifs sélectifs (Autin et al., 2019), aucune d'entre elles ne permet de déterminer si le biais observé au désavantage des élèves issus de l'immigration est dépendant et/ou indépendant tant du statut socio-économique de cet élève que du contexte évaluatif.

L'analyse de la variance simple nous permet de considérer l'hypothèse alternative qui exprime l'influence de certaines variables du modèle sur la note totale obtenue par l'élève. En effet, les variables indépendantes présentes dans ce modèle expliquent 65% de la variance du score de l'élève, c'est-à-dire 65% des écarts de scores observés. Par ailleurs, lorsque nous avons analysé les résultats des sommes de carrés de type III de la variance simple sur la note totale, nous nous sommes rendu compte qu'en réalité, seulement la variable « NIV » (niveau de la copie) nous permettait de rejeter l'hypothèse nulle et d'envisager l'hypothèse alternative selon laquelle l'effet du niveau de la copie sur le score total traduit un phénomène réel dans la population de référence. En ce qui concerne les deux autres variables étudiées, le prénom et la visée de l'évaluation, celles-ci ne nous permettent pas d'étendre nos résultats à la population de référence puisque pour chacune d'elles, le seuil alpha est dépassé ($p > 0.05$) et la valeur F est de respectivement, 0 et 1.21. Les résultats de cette analyse ne nous permettent donc pas d'étendre les résultats de cette expérience sur l'influence du prénom de l'élève et de l'influence de la visée

de l'évaluation à la population de référence. En d'autres termes, les résultats qui seront discutés dans les prochains points concerneront uniquement notre échantillon et ne dépasseront pas ce cadre.

En outre, Quertemont (2023) explique que la mesure de la taille de l'effet est trop souvent négligée dans les conclusions de recherches qui se basent uniquement sur la signification statistique de l'effet. L'auteur rappelle qu'un effet peut être à la fois significatif d'un point de vue statistique et pourtant minime d'un point de vue pratique (ampleur de l'effet). C'est pourquoi, nous baserons principalement cette discussion sur les analyses des amplitudes de l'effet observées dans notre échantillon. Précisons tout de même que pour répondre aux hypothèses, nous ne tiendrons compte que des amplitudes de l'effet considérées comme étant, au minimum, un effet moyen (à partir de 0.16).

Nous mesurons la concordance inter-correcteurs puisqu'une même série de copies est évaluée par des correcteurs différents. Parfois, des différences marquantes peuvent être constatées entre plusieurs correcteurs pour une même copie (Lerclercq et al., 2004). Selon Demonty & Dupont (2022), une fidélité inter-correcteurs se traduit par un score identique, peu importe la personne qui a corrigé et représente un facteur d'équité.

6.1.1 Hypothèse n°1 : LUCIEN VS JORDAN VS AYOUB

Tous les élèves ne sont pas égaux face à l'évaluation. A copie identique, le score moyen obtenu par Ayoub, Jordan et Lucien diffère. (Autin et al., 2019 ; Sprietsma, 2012 ; Rangvid, 2015)

Sans accorder d'importance aux sens de ces différences de moyennes, il semblerait que les scores moyens obtenus par Ayoub, Jordan et Lucien, pour une même copie, diffèrent. En cohérence avec la littérature consultée, nous constatons que malgré la présence d'une grille de cotation, les scores attribués pour une même copie varient (Sprietsma, 2012). Dans la discussion des prochaines hypothèses, nous nous pencherons sur la comparaison de certaines copies, deux à deux, qui semblent intéressantes dans le cadre de cette étude. Nous irons également plus en profondeur quant à l'analyse de l'importance de ces différences grâce aux calculs du *g* de Hedges.

Par ailleurs, nous obtenons des résultats statistiquement significatifs quant à l'influence de la variable « niveau de la copie ». Nous pouvons en déduire que l'hypothèse alternative est démontrée d'autant plus que les résultats des scores moyens obtenus vont également dans ce

sens. Concrètement, ce résultat signifie que les copies faibles obtiennent des scores moyens inférieurs aux copies médianes qui elles, obtiennent des scores moyens inférieurs aux copies fortes. Nous nous attendions à ce résultat puisque les trois niveaux de copies ont été pensés dans ce même sens.

À ce stade, nous pouvons confirmer cette hypothèse dans le cadre de notre échantillon tout en précisant que certes, des différences de scores moyens pour une même copie sont présentes mais nous devons évaluer l'importance et le sens de ces différences à travers les prochaines hypothèses.

Cependant, comme expliqué supra, les variables « prénom de l'élève » et « visée de l'évaluation » n'affectent pas significativement le score total de l'élève, ce qui ne nous permet pas de rejeter l'hypothèse nulle. Dans ce cadre, il semble primordial de souligner l'importance de l'interprétation de cette hypothèse nulle. En effet, un résultat non significatif signifie que nous n'avons pas assez de preuves que pour réfuter l'hypothèse nulle, mais ne signifie pas que cette dernière est démontrée (Quertemont, 2023). De plus, un résultat statistiquement significatif ne veut pas forcément dire que l'effet d'une variable est important ou ne l'est pas. C'est la mesure de la taille de l'effet, qui est calculée par le g de Hedges dans notre cas, qui permet de se faire une idée de l'importance de l'effet mesuré (Quertemont, 2023). Nous l'utiliserons afin de déterminer l'importance des effets mesurés dans le cadre de notre échantillon.

Concernant la non-significativité observée, Quertemont (2022) annonce que la non-significativité d'un test signifie qu'il n'y a pas de garanties assez fortes que pour affirmer que les deux variables sont liées. L'auteur (2023) explique qu'un résultat peut être statistiquement non significatif pour trois raisons : (1) il n'y a réellement aucun effet dans la population ; (2) lors de la planification/exécution de l'étude, il y a eu des erreurs méthodologiques ; (3) la puissance statistique de l'étude n'est pas suffisante que pour mettre cet effet en évidence. La non-significativité de nos résultats pourrait provenir de deux raisons. La première concernerait une erreur méthodologique lors de la planification de l'étude. En effet, nous pensons que la copie médiane n'est pas adéquate pour ce test. Les résultats concernant cette copie vont systématiquement à contre sens des résultats obtenus pour les copies faibles et les copies fortes. En outre, les scores moyens attribués aux copies médianes et les scores moyens attribués aux copies fortes semblent assez proches. En d'autres termes, la copie médiane se rapprocherait d'une copie jugée comme étant forte par les correcteurs, ce qui fausserait donc nos trois catégories de niveaux de copies. La seconde raison qui affecterait la significativité de nos

résultats pourrait provenir de la petite taille de notre effectif par copie corrigée ($N \approx 92$), qui par ailleurs, varie d'une copie à l'autre (allant de 85 à 97). Lors de l'analyse de l'hypothèse 5, nous analysons 18 sous-groupes d'environ 46 sujets. En réalité, il se pourrait que le prénom de l'élève et la visée de l'évaluation ait un effet significatif dans la population de référence mais qu'il n'ait pas pu être démontré suite aux raisons évoquées ci-dessus.

6.1.2 Hypothèse n°2 : LUCIEN VS JORDAN

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »). (Autin et al., 2019 ; Rangvid, 2015)

Pour commencer, la comparaison des moyennes de Jordan et Lucien diffère pour les trois niveaux de copies. Dans notre échantillon, le seul résultat ayant une ampleur de l'effet considérée comme moyenne à forte concerne les copies faibles (-0.29). Pour ces copies, la comparaison des moyennes entre Jordan et Lucien témoigne d'une sous-évaluation de la copie attribuée à Jordan. Ensuite, en ce qui concerne les copies fortes, nos résultats vont dans ce même sens avec une ampleur de l'effet assez faible, ce qui engendre de faibles différences entre Lucien et Jordan. Par ailleurs, nous observons le phénomène inverse pour les copies médianes avec une légère surévaluation de la copie de Jordan et une ampleur de l'effet assez faible (+0.10). Nous ne discuterons pas ce résultat puisqu'il a une ampleur de l'effet trop faible que pour le considérer.

Si nous nous tenons à ces premiers résultats, il semblerait que Lucien et Jordan soient évalués sans différence de grandes ampleurs pour les copies médianes et fortes. Nous retrouvons ici un résultat similaire à celui de Van Ewijk (2011) qui n'a conclu aucun effet des prénoms sur la note attribuée. Par contre, les copies faibles semblent plaider en faveur de l'existence d'un effet de l'origine socio-économique dans notre échantillon. Ce résultat concernant les copies faibles rejoint en partie, les résultats mis en évidence dans différentes études (Autin et al., 2019 ; Sprietsma 2013 ; Lafontaine & Monseur 2009) qui témoignent d'inégalités de traitement concernant les différentes attentes des correcteurs selon certaines de leurs caractéristiques telles que leur genre, leur origine sociale, etc. Certains correcteurs ont tendance à attribuer des notes plus élevées aux élèves issus d'un milieu social favorisé (Pourtois et al., 1978 ; Leclercq et al., 2004), en l'occurrence à « Lucien Faible » dans le cadre de notre étude.

En cohérence avec la littérature consultée (Autin et al., 2019), nos résultats pour les copies faibles indiquent une surévaluation des élèves porteurs du stéréotype positif (niveau socio-économique élevé) et une sous-évaluation des élèves porteurs du stéréotype négatif (niveau socio-économique faible).

En d'autres termes, dans notre échantillon, à performances réelles identiques, pour les copies faibles, l'élève porteur du stéréotype « statut socio-économique élevé » (Lucien) est surévalué comparativement à l'élève qui porte le stéréotype « statut socio-économique faible » (Jordan) qui lui, est sous-évalué. Ce résultat est en cohérence avec d'autres recherches qui relèvent également une surévaluation des élèves ayant un prénom qui fait référence à un haut statut socio-économique et à une sous-évaluation des élèves porteurs du stéréotype défavorisé d'un point de vue socio-économique (Autin et al., 2019 ; Rangvid, 2015).

De plus, nous obtenons un autre résultat qui n'apparaissait pas dans les scores moyens qui impliquaient les deux visées de l'évaluation confondues. Dans l'analyse des résultats de l'hypothèse 5 où nous analysons chaque copie en fonction de la visée de l'évaluation, nous avons relevé la présence d'un effet du statut socio-économique au niveau des copies fortes et ce, pour la visée sélective. En effet, alors que les différences entre Lucien fort et Jordan fort dans un contexte formatif sont négligeables, nous relevons des différences de moyennes plus importantes au désavantage de Jordan dans un contexte normatif. En effet, dans ce contexte, Jordan fort est sous-évalué comparativement à Lucien fort, avec une ampleur de l'effet de -0.23 et une différence de moyenne de -0.38. Ces constats rejoignent donc également les propos tenus par Autin et ses collègues (2019) dans lesquels ils expliquent que les connaissances des évaluateurs sur la classe sociale de l'élève, par l'information du prénom dans notre cas, influencent l'évaluation de ses performances en faveur des élèves ayant un SES plus élevé.

Ces résultats attestent de l'existence d'un effet du statut socio-économique de l'élève sur son score. Cette deuxième hypothèse se voit confirmée au niveau des **copies faibles** dans notre échantillon (contextes sélectifs et formatifs confondus) ainsi qu'au niveau des **copies fortes en contexte sélectif** et ne peut être généralisée à la population de référence.

6.1.3 Hypothèse n°3 : LUCIEN VS AYOUB

A copie identique, le score moyen obtenu par Lucien (élève porteur du stéréotype « favorisé d'un point de vue socio-économique ») est supérieur comparativement au score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique »). (Sprietsma, 2012 ; Rangvid, 2015)

Notons tout d'abord que, dans notre échantillon, la comparaison des moyennes et le calcul de l'ampleur de l'effet témoignent d'une influence importante du prénom de l'élève sur le score attribué aux copies faibles et fortes de sorte que les copies d'Ayoub sont en moyenne, moins bien notées que les copies de Lucien. A performances réelles identiques, pour les copies jugées comme étant faibles et fortes, les correcteurs sous-évaluent les performances de l'élève ayant un prénom faisant référence à l'immigration (Ayoub) comparativement à l'élève autochtone (Lucien). Ces résultats vont dans la continuité de ceux présentés par d'autres chercheurs (Rangvid, 2015 ; Bonefeld & Dickhäuser, 2018). En effet, Rangvid (2015) soutient qu'un prénom faisant référence à l'immigration peut induire des différences de cotations, même si le correcteur ne connaît pas cet élève.

Par ailleurs, le phénomène inverse est observé pour les copies médianes. Pour ces copies, Ayoub obtient un score moyen supérieur à celui de Lucien, avec une ampleur de l'effet considérée comme très forte. Dans ce cadre, nous nous inspirons d'une explication soumise par Sprietsma (2012) à propos de l'étude de Van Ewijk (2011) qui n'avait conclu aucun effet du prénom (immigré ou natif) sur le score moyen. Nous pensons que les correcteurs qui ont surévalué Ayoub comparativement à Lucien pourraient être conscients d'une tendance à être influencés par l'ethnicité de l'élève et adaptent intentionnellement leurs notations. Cette justification reste très hypothétique et la correction de ces copies médianes mériterait d'être répliquée afin de vérifier si nous obtenons des résultats similaires à ceux-ci. De plus, comme nous l'avons déjà mentionné, la copie initialement jugée « médiane » se rapproche plutôt d'un niveau de copie « forte ».

Face à des copies absolument homogènes à l'exception du prénom, les sujets ont tendance à surestimer la copie « Lucien Fort » et à sous-estimer la copie équivalente « Ayoub Fort ». Il en est de même pour les copies faibles de notre échantillon. Ce troisième résultat semble rejoindre en partie ceux mis en évidence dans plusieurs études qui avaient déjà mis en évidence ce biais lié à l'origine ethnique par des différences de scores moyens, à copie égale, entre un élève natif et un élève d'origine immigrée (Sprietsma, 2012 ; Rangvid, 2015).

Autrement dit, à copie identique, Ayoub (défavorisé tant du point de vue ethnique que socio-économique) obtient un score inférieur à celui de Lucien (favorisé d'un point de vue socio-économique) pour les copies faibles et les copies fortes.

Ces résultats prouvent l'effet du statut socio-économique ainsi que de l'origine ethnique sur le score de l'élève dans notre échantillon. Cette troisième hypothèse se voit alors confirmée pour les **copies faibles** et **fortes** de notre échantillon (contextes sélectifs et formatifs confondus) mais, rappelons-le, invalidée pour la population de référence.

6.1.4 Hypothèse n°4 : JORDAN VS AYOUB

Il existe un biais de l'évaluation lié à l'origine ethnique des élèves indépendant de leur origine sociale et de leur niveau de performance. A copie identique, le score moyen obtenu par Ayoub (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique ») est inférieur comparativement au score moyen obtenu par Jordan (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »).

Les résultats obtenus suite à la comparaison des moyennes ainsi qu'au calcul du g de Hedges nous permettent de constater que, dans notre échantillon, les copies fortes obtiennent un score moyen moins élevé lorsqu'elles sont attribuées à Ayoub plutôt qu'à Jordan. Cette différence de moyenne est constatée avec une ampleur de l'effet considérée comme étant moyenne à forte. Ce même constat a été observé pour les copies faibles. Par contre, les différences de moyenne concernant ces copies obtiennent une ampleur de l'effet trop faible que pour en tirer des conclusions.

En revanche, en ce qui concerne les copies médianes, elles obtiennent des scores moyens plus élevés à l'avantage de Jordan avec une ampleur de l'effet très forte. Nous avons discuté un résultat similaire dans le cadre de l'hypothèse précédente.

En d'autres termes, dans notre échantillon, nous constatons des différences de scores moyens au niveau des copies fortes à l'avantage de Jordan. A copie identique, le score moyen obtenu par Ayoub Fort (élève porteur du stéréotype « défavorisé tant du point de vue ethnique que socio-économique ») est inférieur comparativement au score moyen obtenu par Jordan Fort (élève porteur du stéréotype « défavorisé d'un point de vue socio-économique »).

En outre, lorsque nous analysons chaque copie en fonction de la visée de l'évaluation, nous relevons la présence d'un effet de l'origine ethnique indépendant de l'effet du statut socio-économique au niveau des copies faibles et ce, pour la visée sélective. En effet, alors que les

différences entre Ayoub Faible et Jordan Faible dans un contexte formatif sont négligeables, nous relevons des différences de moyennes plus importantes au désavantage d'Ayoub dans ce contexte sélectif. Ayoub Faible est donc sous-évalué comparativement à Jordan Faible, avec une ampleur de l'effet de -0.24 et une différence de moyenne de -0.49.

Dans la discussion des hypothèses précédentes, nous avons relevé au niveau de notre échantillon, d'une part, un biais lié à l'origine sociale de l'élève pour certaines copies et d'autre part, un biais lié à l'origine ethnique de l'élève pour certaines copies. Comme nous l'avions signalé supra, à notre connaissance, aucune étude n'a étudié l'influence conjointe de ces deux biais. Les résultats obtenus dans le cadre de cette hypothèse sont sans doute les plus remarquables de cette étude sur l'effet du prénom. En effet, nous obtenons des résultats qui semblent plaider en faveur de l'existence d'un biais lié à l'origine ethnique de l'élève en partie indépendant de son origine socio-économique, dans notre échantillon.

Cette quatrième hypothèse se voit confirmée au niveau des **copies fortes** de notre échantillon (contextes sélectifs et formatifs confondus) ainsi qu'au niveau des **copies faibles uniquement dans un contexte sélectif** et ne peut être généralisée à la population de référence.

6.1.5 Hypothèse n°5 : LUCIEN VS JORDAN VS AYOUB

A copie identique, l'évaluation en vue de la sélection (normative) peut induire une plus grande reproduction des inégalités (entre Lucien, Jordan et Ayoub) que l'évaluation en vue de l'apprentissage (formative). (Autin et al., 2019)

Dans un premier temps, reprenons les différences de moyennes ainsi que les ampleurs de l'effet entre les deux visées de l'évaluation pour une même copie, c'est-à-dire pour chacune des neuf copies pour la visée formative et pour la visée sélective (Ayoub Faible normatif VS Ayoub Faible formatif).

Notons tout d'abord que les calculs du g de Hedges pour les comparaisons selon la visée de l'évaluation d'une même copie varient pour la plupart des copies. Par contre, pour Jordan Faible, Lucien Faible et Ayoub Médian, les ampleurs de l'effet sont inexistantes puisqu'elles avoisinent le zéro, ce qui se traduit par une absence de différence entre les deux visées de l'évaluation pour ces copies.

Remarquons ensuite que les comparaisons de moyennes pour lesquelles nous obtenons les ampleurs de l'effet les plus élevées concernent les copies « Ayoub Faible » (-0.33) et « Ayoub Fort » (-0.27). Nos résultats indiquent, pour ces deux copies, des scores plus élevés lorsque la

visée de l'évaluation est formative que quand elle est normative. Ces résultats interpellent puisque cela signifie, qu'Ayoub Faible et Ayoub Fort obtiennent davantage de points lorsqu'ils sont dans une évaluation formative. Rappelons que cette dernière a pour but d'informer l'élève quant au degré de maîtrise de son apprentissage (Detroz, 2021). Par contre, lorsque ces mêmes élèves se retrouvent en situation d'évaluation sélective, qui a pour but de trier les élèves (Autin et al., 2019), leurs moyennes subissent une diminution de points à hauteur de 0.40 point pour la copie forte et de 0.64 pour la copie faible.

Enfin, pour les autres copies (Jordan Médian, Lucien Médian, Jordan Fort et Lucien Fort), nous obtenons des ampleurs de l'effet relativement faibles et ne tiendrons donc pas compte de ces résultats.

Pour rappel, la consigne donnée aux participants devant corriger des évaluations normatives était la suivante : « Chaque rédaction que vous allez corriger permettra de définir si l'élève est promu ou bien s'il doit recommencer son année ». En d'autres termes, si l'élève obtenait un score inférieur à 7.5/15 à la rédaction, celui-ci devait recommencer son année. Or, la consigne donnée pour les évaluations formatives était, elle, très différente « Chaque rédaction que vous allez corriger permettra à l'élève de prendre conscience de ses acquis ainsi que des points à retravailler grâce aux commentaires que vous allez rédiger dans le but qu'il puisse s'améliorer ». Pour les deux visées de l'évaluation, toutes les copies faibles ont obtenu un score moyen en dessous de la moyenne et se retrouvent donc, en situation d'échec. Remarquons tout de même que pour ces copies faibles, elles obtiennent certes, des scores moyens plus élevés lorsqu'elles sont corrigées dans un contexte d'évaluation formative, mais ces différences sont insignifiantes dans les cas de Jordan et de Lucien. Par contre, comme nous l'avons souligné supra, Ayoub obtient un score moyen supérieur pour l'évaluation en vue de l'apprentissage avec une forte ampleur de l'effet. Nous pouvons donc en déduire qu'en contexte normatif, les performances de la copie Ayoub Faible sont davantage sous-évaluées alors que Jordan Faible et Lucien Faible ne subissent pas de grandes différences de moyennes entre les deux visées de l'évaluation. Il en est de même pour les copies fortes : alors que Jordan et Lucien subissent très peu de variations dans leurs scores moyens entre les deux visées de l'évaluation, Ayoub, lui, se voit attribuer un score plus inférieur en contexte sélectif qu'en contexte formatif.

En contexte de sélection, les copies **Ayoub Fort** et **Ayoub Faible** sont davantage sous-évaluées comparativement à leurs scores moyens en contexte de formation, dans notre échantillon.

Dans un second temps, discutons des différences de moyennes ainsi que des ampleurs de l'effet lorsque nous comparons les trois prénoms pour un même niveau de copie d'une part, dans la visée normative et d'autre part, dans la visée formative (Jordan Faible normatif vs Lucien Faible normatif VS Jordan Faible formatif vs Lucien Faible formatif).

Ainsi, à copies égales, nous remarquons de plus grandes différences de moyennes entre Ayoub, Jordan et Lucien dans les évaluations normatives que dans les évaluations formatives et ce, pour les trois niveaux de copies à l'exception d'une seule comparaison : Ayoub et Jordan Médians. Ces résultats corroborent avec ceux d'Autin et al. (2019) qui avaient conclu que lorsque l'évaluation adopte un rôle de sélection, à copie égale, l'écart de points entre les élèves de différentes classes sociales se marque davantage. En effet, alors que dans un contexte formatif, 0.38 point séparent les moyennes d'Ayoub et Lucien Faibles à l'avantage de ce dernier, nous observons, toujours à l'avantage de cet élève, une différence plus conséquente s'élevant à 1.01 point dans la visée normative. Pour l'évaluation dans un contexte de sélection, nous obtenons pour chacune des comparaisons entre élèves, des ampleurs de l'effet de moyenne à forte. Le calcul des ampleurs de l'effet de l'évaluation va dans ce même sens puisqu'en contexte de formation, les ampleurs de l'effet sont moindres comparativement à celles du contexte sélectif (à l'exception de la comparaison Ayoub Médian – Jordan Médian). Notons que toutes les ampleurs de l'effet calculées dans le contexte normatif sont considérées comme moyennes à fortes.

Par ailleurs, en ce qui concerne le sens de ces différences en contexte sélectif, nous formulons le même constat que celui formulé pour les précédentes hypothèses : les copies faibles et fortes sont sous-évaluées en défaveur des élèves porteurs des stéréotypes les plus négatifs (Ayoub < Jordan < Lucien). Ces effets sont constatés avec des ampleurs de l'effet considérables, allant de 0.24 à 0.63. Pour les copies médianes, nous constatons à nouveau l'effet inverse.

Nous pouvons donc affirmer que, pour les copies faibles et fortes, l'évaluation normative augmente considérablement la différence de moyennes entre élèves au désavantage des élèves porteurs de stéréotypes négatifs. En revanche, l'évaluation formative diminue ces différences de moyennes. Si nous prenons le cas de Jordan et Lucien Forts, nous observons une différence de moyenne s'élevant à 0.38 point à l'avantage de Lucien pour l'évaluation normative qui diminue fortement dans le cadre de l'évaluation formative puisqu'elle s'élève seulement à 0.03. Autrement dit, cette différence très proche du zéro est négligeable, avec une ampleur de l'effet de 0.02. Pour certaines comparaisons, ces différences restent d'ampleurs assez fortes mais dans tous les cas, elles diminuent lorsqu'elles sont en situation d'évaluation en vue de l'apprentissage

(à l'exception d'Ayoub et Jordan Médiants). En effet, ces différents résultats vont dans le même sens et attestent que les scores moyens attribués en contexte de sélection accroissent les différences de moyennes entre Lucien, Jordan et Ayoub.

Ainsi, les résultats de ces analyses statistiques nous permettent d'attester, de manière descriptive au niveau de notre échantillon, de la présence d'un effet entre les variables « prénom », « visée de l'évaluation » et la variable dépendante, le score de l'élève. Les résultats de cette cinquième hypothèse soutiennent donc certaines données de la littérature, selon lesquelles les auteurs ont remarqué un écart de points entre les élèves issus de différentes classes sociales lorsque l'évaluation adopte un rôle de sélection (Autin et al., 2019).

Cette cinquième hypothèse se voit confirmée au niveau des **copies faibles et fortes** de notre échantillon. Pour ces copies, l'évaluation en vue de la sélection induit une plus grande reproduction des inégalités entre Lucien, Jordan et Ayoub que l'évaluation en vue de l'apprentissage.

Nous terminerons cette partie dédiée à la discussion en reprenant les résultats les plus marquants et pour lesquels nous obtenons des tailles de l'effet relativement importantes dans le cadre de notre échantillon. **(1)** Pour une même copie, Jordan Faible obtient un score moyen inférieur à Lucien Faible (contextes sélectifs et formatifs confondus) et Jordan Fort obtient un score moyen inférieur à Lucien Fort uniquement en contexte sélectif ; **(2)** à copie identique, Ayoub Faible et Ayoub Fort obtiennent respectivement des scores moyens inférieurs à Lucien Faible et à Lucien Fort ; **(3)** à performances identiques, Ayoub Fort obtient un score moyen inférieur à celui de Jordan Fort (contextes sélectifs et formatifs confondus) et Ayoub Faible obtient un score inférieur à celui de Jordan Faible uniquement dans un contexte d'évaluation normative; **(4)** Ayoub Faible et Ayoub Fort sont davantage sous-évalués en contexte normatif ; **(5)** l'évaluation normative produit davantage d'inégalités entre Lucien, Jordan et Ayoub que l'évaluation formative pour les copies faibles et fortes.

Rappelons-le, nous ne pouvons pas transférer ces résultats à la population de référence puisque l'analyse de la variance révèle des influences non significatives du prénom et de la visée de l'évaluation sur le score de l'élève.

Ces résultats soutiennent les propos de Bressoux (2012) dans lesquels il explique que lorsqu'un correcteur évalue les performances d'un élève, il est biaisé par certaines de ses caractéristiques. Dans cette étude, le prénom de l'élève, renvoyant à son statut socio-économique et à son origine ethnique, était la seule information dont disposaient les correcteurs. Nous sommes donc en

mesure d'affirmer que le score moyen d'Ayoub et de Jordan, pour certains de ces résultats, a été contaminé par un effet de halo lié à l'origine sociale et à l'origine ethnique. Dutrévis (2015) ainsi que Leclercq et ses collègues (2004) soutiennent également cette idée selon laquelle certaines caractéristiques de l'élève influencent la note attribuée par l'évaluateur.

6.2 Limites de cette recherche et perspectives

Nous nous devons de poser un regard rétrospectif sur notre recherche et ainsi, mettre en évidence ses limites tout en présentant d'éventuelles perspectives.

Tout d'abord, pour être en parfaite cohérence avec la réalité de terrain, nous aurions dû demander à des enseignants de 5^e année de l'enseignement primaire d'effectuer la correction de ces copies. Lafontaine et Monseur (2009) avaient privilégié cette cohérence avec la réalité en ayant recours à un échantillon composé de 48 enseignants de mathématiques qui devaient corriger une épreuve qui concernait les mathématiques. Nous avons fait le choix d'ouvrir la récolte de données à tous les (futurs) enseignants dans le but de comptabiliser un plus grand nombre de sujets. En faisant ce choix, nous nous éloignons donc de la réalité de terrain.

Il est vrai que dans notre design de recherche, nous n'avons pas cherché à mesurer l'effet de fatigue ou encore l'effet de l'ordre des copies. Tous les ordres possibles n'ont d'ailleurs pas été envisagés dans l'organisation des douze groupes expérimentaux. Il aurait été intéressant de calculer la fidélité intra-correcteur en fonction de l'ordre des copies. Par ailleurs, le fait qu'un correcteur ait évalué 3 copies différentes de notre échantillon pourrait apporter certains biais à nos résultats. En effet, nous avons effectué nos analyses sur base des 825 copies qui en réalité, ont été corrigées par 275 participants. En fonction des analyses que nous avons réalisées, il aurait été préférable qu'un participant corrige une seule copie.

Ensuite, l'absence d'une réelle condition contrôle reste un point à remédier dans le cadre d'une prochaine étude. Dans ce cadre, nous proposons de prévoir une évaluation sans prénom. Selon Rangvik (2015), les conditions pour une correction objective sont une notation externe et à l'aveugle. Ainsi, nous obtiendrions un score moyen issu d'une correction totalement à l'aveugle : les correcteurs n'auraient aucune information sur l'élève rédacteur de la copie.

Cependant, les effets observés pour les copies médianes méritent d'être discutés. Tous les résultats des copies faibles et fortes vont dans un même sens (avec des tailles de l'effet de faible à forte). À contrario, les résultats des copies médianes ne suivent pas ce même raisonnement. Nous pensons qu'en réalité, la copie jugée initialement comme médiane se rapproche plutôt d'une copie forte. Nous soulevons donc une erreur méthodologique quant à l'attribution

« niveau médian » pour cette copie. Une perspective intéressante serait de sélectionner une autre copie médiane qui correspondrait réellement à un niveau de performances d'un élève moyen. Cette suggestion mériterait de faire l'objet de répliques ultérieures.

Enfin, d'autres perspectives découlent de notre méthodologie et de nos résultats. D'une part, pour de futures recherches semblables à celle menée dans ce travail, nous préconisons l'utilisation de questionnaire en ligne. En effet, cela permettrait un codage immédiat des données ainsi que l'accès à un plus grand nombre de participants (Demonty & Dupont, 2022). D'autre part, des recherches complémentaires sont nécessaires afin de comprendre les mécanismes psychologiques qui entrent en jeu dans le rôle qu'ont les évaluateurs lorsqu'ils ont un comportement discriminatoire envers les élèves défavorisés sur les plans ethniques et économiques.

7 CONCLUSION

Dans le présent mémoire, nos intentions étaient d'une part, d'étudier l'existence d'un biais lié à l'origine ethnique indépendant du biais lié à l'origine sociale de l'élève et d'autre part, d'observer les inégalités entre élèves tant dans l'évaluation en contexte sélectif que l'évaluation en contexte formatif.

Pour ce faire, nous avons formulé et testé différentes hypothèses. Nous avons tout d'abord étudié, de manière isolée, l'influence de ces deux biais en comparant d'une part, les évaluations de Lucien et Jordan et d'autre part, celles de Lucien et Ayoub. Nos résultats pour les copies faibles et fortes sont cohérents avec les recherches précédentes en sciences de l'éducation (Rangvid, 2015 ; Autin et al., 2019 ; Sprietsma, 2012 ; Bressoux, 2012). Ensuite, dans la continuité des travaux docimologiques présentés dans la revue de la littérature, nous avons étudié l'influence d'un biais lié à l'origine ethnique de l'élève indépendant du biais lié à l'origine sociale et ce, dans des contextes d'évaluation formative et d'évaluation sélective. Notre étude nous a mené à confirmer au niveau de notre échantillon, pour certains groupes de copies, un biais lié à l'origine ethnique de l'élève indépendant du biais lié à son origine sociale. Pour finir, nous nous sommes intéressée aux deux visées de l'évaluation présentes dans la récolte de données à savoir, l'évaluation en vue de la sélection et l'évaluation en vue de l'apprentissage.

Face à des copies strictement identiques à l'exception du prénom et ce, pour trois niveaux de copies différents, nous pouvons affirmer que toutes nos hypothèses ont été validées auprès de

certaines comparaisons de niveaux de copies, dans le cadre de notre échantillon. Les différences systématiques observées entre ces trois élèves pour les copies faibles et fortes, tantôt à l'avantage de Jordan (lorsqu'il est comparé à Ayoub), tantôt à l'avantage de Lucien (lorsqu'il est comparé à Jordan ou à Ayoub), confirment donc certaines données de la littérature préexistante (Rangvid, 2015 ; Autin et al., 2019 ; Sprietsma, 2012 ; Bressoux, 2012). Comme Autin et ses collègues (2019) le soulignent dans leur méta-analyse, même si certains résultats peuvent sembler faibles en termes de taille de l'effet, il est important de les interpréter sur la durée de la scolarité d'un élève. Sur une scolarité, de petites différences de cotations peuvent avoir de grandes conséquences sur les résultats scolaires de l'élève. Par ailleurs, certains auteurs ont également montré l'impact de ces biais de cotation dans l'avenir des élèves (Rangvid, 2015 ; Ichou, 2013 ; Sprietsma, 2012).

Nous constatons donc une grande contradiction entre d'une part, les missions de notre système scolaire, notamment stipulées dans l'article 6 du Décret Missions qui prévoient une égalité des chances d'émancipation et d'autre part, les discriminations systémiques produites à l'égard des élèves issus de l'immigration ainsi qu'à l'égard des élèves issus d'un milieu social défavorisé.

Cette étude menée auprès de 275 sujets ayant corrigé 3 copies (825 copies in fine) nous permet donc de répondre à notre question de recherche « *Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?* ». En conclusion, nous pouvons affirmer qu'au sein de notre échantillon, nous avons relevé l'existence d'un biais de l'évaluation lié à l'origine ethnique de l'élève indépendamment du biais lié à son origine sociale au niveau des copies fortes Ayoub-Jordan (contextes sélectifs et formatifs confondus) ainsi qu'au niveau des copies faibles dans un contexte sélectif. Quant aux deux visées de l'évaluation, nous pouvons soutenir l'idée selon laquelle l'évaluation en vue de la sélection induit une plus grande reproduction des inégalités que l'évaluation en vue de l'apprentissages pour les copies faibles et fortes de notre échantillon.

Le présent mémoire a permis de mettre en lumière l'influence de certaines caractéristiques relatives à l'évalué ainsi que l'influence de la visée de l'évaluation sur le jugement évaluatif des (futurs) enseignants de notre échantillon. Cependant, nous ne pouvons généraliser les résultats de cette recherche à la population de référence, c'est-à-dire à l'ensemble des (futurs) enseignants en FW-B, et devons limiter leurs interprétations à la taille de notre échantillon. Cette étude pourrait être répliquée en tenant compte des limites ainsi que des perspectives formulées à son égard.

8 BIBLIOGRAPHIE

Autin, F., Batruch, A. & Butera, F. (2019). The function of selection of assessment lead evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology, 111*, 717-735.

<https://doi.org/10.1037/edu0000307>

Baye, A. (2015, 3 mars). Le poids des origines. Influence du niveau socioéconomique et de l'origine culturelle sur les inégalités d'acquis et de parcours scolaires à partir des données PISA [Présentation PowerPoint]. Retrieved from :

https://orbi.uliege.be/bitstream/2268/179576/1/BAYE_Francqui.pdf

Bonefeld, M., & Dickhäuser, O. (2018). (Biased) Grading of Students' Performance : Students' Names, Performance Level, and Implicit Attitudes. *Frontiers in Psychology, 9*, 481.

<https://doi.org/10.3389/fpsyg.2018.00481>

Bressoux, P. (2012). L'influence des pratiques enseignantes sur les acquisitions scolaires des élèves. *Regards croisés sur l'économie, 12* (2), 208-217.

<https://doi.org/10.3917/rce.012.0208>

Crahay, M. (2013). L'école peut-elle être juste et efficace ? De l'égalité des chances à l'égalité des acquis. Bruxelles : De Boeck.

Crahay, M., Mottier Lopez, L., & Marcoux, G. (2019). L'évaluation des élèves : Docteur Jekyll and Mister Hyde dans l'enseignement. In M. Crahay (Ed.), *Peut-on lutter contre l'échec scolaire?*. (4^e édition revue et corrigée, pp.357-417). Louvain-la-Neuve, Belgique : de Boeck supérieur

Communauté française (1997). *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. Bruxelles : Communauté française de Belgique, Ministère de l'Éducation, de la Recherche et de la Formation. Retrieved from :

https://www.gallilex.cfwb.be/document/pdf/21557_052.pdf

Dachet, D., & Baye, A. (2021). Evidence-Based Education : The (Not So Simple) Case of French-Speaking Belgium. *ECNU Review of Education*, 4(1), 164-189.

<https://doi.org/10.1177/2096531120928086>

Dachet, D., & Pressia, F. (2022). Mémoire partim 1 : notes de cours. Unpublished document, Université de Liège, Liège.

Darley, J. M. & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44 (1), 20-33.

<https://doi.org/10.1037//0022-3514.44.1.20>

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review*, 95(2), 158–165. Retrieved from :

<http://www.jstor.org/stable/4132809>

De Ketele, J.-M. (1993). L'évaluation conjugquée en paradigmes. *Revue française de pédagogie*, 103(1), 59-80.

<https://doi.org/10.3406/rfp.1993.1298>

De Landsheere, G. (1992). Evaluation continue et examens : précis de docimologie. Bruxelles, Belgique : Labor.

Demeuse, M., & Baye, A. (2005). Pourquoi parler d'équité ? In M. Demeuse, A. Baye, M.H. Straeten, J. Nicaise, & A. Matoul (Eds.), *Vers une école juste et efficace. 26 contributions sur les systèmes d'enseignement et de formation*. Bruxelles, Belgique : De Boeck. (pp.150-170).

Demonty, I. & Dupont, V. (2022). Construction et analyse de tests et de questionnaires : notes de cours. Unpublished document, Université de Liège, Liège.

Detroz, P. (2021). Questions d'évaluation : notes de cours. Unpublished document, Université de Liège, Liège.

Dupriez, V. (2019). Face à l'origine sociale, les écoles et les enseignants sont impuissants. In P. Rayou (Ed.), *L'origine sociale des élèves*. (pp. 88-98). Editions Retz.

Dutrévis, M. (2015). Menace du stéréotype et groupes ethno-raciaux minoritaires. Quel poids sur les performances des élèves ? *Revue française de pédagogie*, 191, 61-72.

<https://doi.org/10.4000/rfp.4751>

Felouzis, G. (2003). La ségrégation ethnique au collège et ses conséquences. *Revue Française de Sociologie*, 44(3), 413.

<https://doi.org/10.2307/3323202>

Felouzis, G., Fouquet-Chauprade, B., & Charmillot, C. (2015). Les descendants d'immigrés à l'école en France : entre discontinuité culturelle et discrimination systémique. *Revue française de pédagogie*, 191, 11-27. Retrieved from :

<https://www.cairn.info/revue-francaise-depedagogie-2015-2-page-11.htm>

Felouzis, G., Fouquet-Chauprade, B., Charmillot, S., Imperiale-Arfaine, L. (2016). Inégalités scolaires et politiques d'éducation. Contribution au rapport du Cnesco. *Les inégalités scolaires d'origines sociales et ethnoculturelles*. CNESCO, Paris.

Frayssinhes, J. (2022). Les biais cognitifs : leurs rôles dans l'apprentissage. *Revue Education Permanente*, 233(4), 1-12. <https://hal.archives-ouvertes.fr/hal-03894576>

Fresson, M., & Dardenne, B. (2019). Les redoublants, victimes de stéréotypes dévalorisants ? In M. Crahay (Ed.), *Peut-on lutter contre l'échec scolaire?*. (4^e édition revue et corrigée, pp.231-258). Louvain-la-Neuve, Belgique : de Boeck supérieur

Goudeau, S., & Croizet, J.-C. (2017). Hidden Advantages and Disadvantages of Social Class : How Classroom Settings Reproduce Social Inequality by Staging Unfair Comparison. *Psychological Science*, 28(2), 162-170. <https://doi.org/10.1177/0956797616676600>

Goudeseune, D. (2018, juillet, 13). *L'ampleur d'effet, un indicateur statistique en éducation qui ne faut pas l'économie de la complexité !* Par temps clair. Retrieved May 04, 2023, from [Comprendre la signification d'une taille d'effet en sciences de l'éducation ~ Par temps clair \(par-temps-clair.blogspot.com\)](https://par-temps-clair.blogspot.com)

Ichou, M. (2013). Différences d'origine et origine des différences : les résultats scolaires des enfants d'émigrés/immigrés en France du début de l'école primaire à la fin du collège. *Revue française de sociologie*, 54 (1), p. 5-52.

<http://dx.doi.org/10.3917/rfs.541.0005>

Ichou, M. (2018). Les enfants d'immigrés à l'école : inégalités scolaires du primaire à l'enseignement supérieur.

Lafontaine, D., & Monseur, C. (2014). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et évaluation en éducation*, 32(2), 71-98.

<https://doi.org/10.7202/1024955ar>

Lafontaine, D., Crépin, F., & Quittre, V. (2017). Les compétences des jeunes de 15 ans en sciences, en mathématiques et en lecture. Résultats de l'enquête PISA 2015 en Fédération Wallonie-Bruxelles. *Les cahiers des sciences de l'éducation*, 37. Liège : aSPe ULiège.

Retrieved from :

http://www.aspe.ulg.ac.be/Files/cahier_pisa2015_37_.pdf

Lafontaine, D., Baye, A., & Monseur, C. (2019). Redoublement, efficacité et équité : l'éclairage des enquêtes internationales. In M. Crahay (eds.), *Peut-on lutter contre l'échec scolaire ?* (4th ed.) (pp. 111-154). Louvain-la-Neuve, Belgique : de Boeck supérieur

Leclercq, D., Nicaise, J., Demeuse, M. (Ed). (2004). Docimologie critique : Des difficultés de noter des copies et d'attribuer des notes aux élèves. In M, Demeuse (Ed.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation*. (pp.273-292)

Monseur, C., & Baye, A. (2015). Quels apports des données PISA pour l'analyse des inégalités scolaires ? Contribution au rapport du Cnesco : *Comment l'école amplifie-t-elle les inégalités sociales et migratoires ?*

Monseur, C. (2020). *Notions de statistiques appliquées à l'éducation (PEDA4054-1)*. Liège, Belgique : eCampus.

Pourtois, J-P., Bonacina, R., Delbecq, A., Segard, M. (1978). Le niveau d'expectation de l'examineur est-il influencé par l'appartenance sociale de l'enfant ? *Revue française de pédagogie*, 44 , 34-37.

Quertemont, E (2022). Psychostatistique descriptive et inférentielle partim 1 : notes de cours. Unpublished document, Université de Liège, Liège.

Quertemont, E (2023). Psychostatistique descriptive et inférentielle partim 2 : notes de cours. Unpublished document, Université de Liège, Liège.

Rangvid, B. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review*, 48, 41-55.

<http://dx.doi.org/10.1016/j.econedurev.2015.05.003>

Rot, N. & Butas, Z. (1959). Les distributions de notes scolaires comparées aux distributions des résultats obtenus dans les tests de connaissance. *Le Travail Humain*, 22,1/2. Retrieved from :

<http://www.jstor.org/stable/40659144>

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). « Unlearning » automatic biases : The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81(5), 856-868.

<https://doi.org/10.1037/0022-3514.81.5.856>

Sprietsma, M. (2012). Discrimination in grading : experimental evidence from primary school teachers. *Empirical Economics*, 45, 523-538.

<http://dx.doi.org/10.1007/s00181-012-0609-x>

Trouilloud, D. & Sarrazin, P. (2003). Note des synthèse [Les connaissances actuelles sur l'effet Pygmalion : Processus, poids et modulateurs]: Les connaissances actuelles sur l'effet Pygmalion : processus, poids et modulateurs. *Revue française de pédagogie*, 145(1), 89-119.

<https://doi.org/10.3406/rfp.2003.2988>

Van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review*, 30(5), 1045-1058.

<https://doi.org/10.1016/j.econedurev.2011.05.008>

Yzerbyt, V.(2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90-95.

<https://doi.org/10.1016/J.copsyc.2016>

Yzerbyt, V. & Klein, O. (2019). *Psychologie sociale*. De Boeck Supérieur.

Zafar, B. (2011). How Do College Students Form Expectations? *Journal of Labor Economics*, 29(2), 301-348.

<https://doi.org/10.1086/658091>

9 RÉSUMÉ

Différentes recherches en docimologie se sont intéressées aux biais impactant tout jugement évaluatif. Ainsi, certains chercheurs ont montré qu'à performances réelles identiques, des élèves porteurs du stéréotype « favorisé » sur le plan socio-économique sont surévalués comparativement aux élèves qui portent le stéréotype « défavorisé » qui eux, sont sous-évalués (Autin et al., 2019 ; Rangvid, 2015). Par ailleurs, Bonefeld et Dickhäuser (2018), Rangvid (2015) et Sprietsma (2013) ont mis en lumière des résultats similaires au désavantage des élèves d'origine immigrée. Ces constats interpellent puisque si ce phénomène se reproduit pour un grand nombre d'évaluations, la différence au bout du compte pourrait avoir des conséquences dans le parcours scolaire de ces élèves discriminés. La mise en parallèle de ces deux constats, à savoir, l'existence de biais de l'évaluation au détriment des élèves défavorisés (Autin et al., 2019) et au détriment des élèves d'origine immigrée (Rangvid, 2015) soulève une certaine interrogation. Les effets liés aux stéréotypes sur les élèves d'origine immigrée sont-ils dépendants ou non des effets liés au statut socio-économique ?

Le présent mémoire vise donc à étudier l'existence d'un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale. Dans ce cadre, une expérience docimologique sera conduite auprès d'enseignants et de futurs enseignants. L'objectif de la présente recherche est de mesurer l'effet de l'origine ethnique indépendamment de l'effet du statut socio-économique des élèves sur les notes qui leur sont attribuées par les (futurs) enseignants. Sur base d'un design expérimental, nous testons des enseignants et des futurs enseignants en position d'évaluateurs et nous observons comment ceux-ci se comportent face à la copie d'un élève issu de l'immigration (appelé Ayoub), à la copie d'un élève natif issu d'un milieu socio-économique défavorisé (appelé Jordan) et à la copie d'un élève autochtone issu d'un milieu socio-économique favorisé (appelé Lucien).

Nous analysons donc les comportements de nos (futurs) enseignants en Fédération Wallonie-Bruxelles (FW-B) en situation de correction et observerons s'ils (1) sont influencés dans leur cotation par l'origine ethnique et/ou l'origine sociale des élèves ; (2) font apparaître, dans leurs cotations, un biais lié à l'origine ethnique indépendant du biais lié à l'origine sociale ; (3) se comportent différemment en fonction de la visée de l'évaluation. Autrement dit, nous tentons de répondre à la question de recherche suivante : *Existe-t-il un biais de l'évaluation lié à l'origine ethnique d'un élève indépendant du biais lié à son origine sociale dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ?*

10 ANNEXES

TABLE DES ANNEXES

Annexe 0. La consigne de rédaction.....	70
Annexe 1. L'ensemble des rédactions d'une classe de P5	71
Annexe 2. La rédaction jugée comme étant « faible ».....	91
Annexe 2A. Ayoub faible	92
Annexe 2B. Jordan faible	93
Annexe 2C. Lucien faible	94
Annexe 3. La rédaction jugée comme étant « médiane ».....	95
Annexe 3A. Ayoub médian	96
Annexe 3B. Jordan médian.....	97
Annexe 3C. Lucien médian	98
Annexe 4. La rédaction jugée comme étant « forte ».....	99
Annexe 4A. Ayoub fort	100
Annexe 4B. Jordan fort.....	101
Annexe 4C. Lucien fort	102
Annexe 5. La grille d'évaluation formative.....	103
Annexe 6. La grille d'évaluation normative.....	105
Annexe 7. Les consignes générales (version formative) à destination des étudiants.....	106
Annexe 8. Les consignes générales (version normative) à destination des étudiants.....	107
Annexe 9. Les consignes générales (version formative) à destination des enseignants	108
Annexe 10. Les consignes générales (version normative) à destination des enseignants.....	109
Annexe 11. Courrier à l'intention des directions	110
Annexe 12. Annonce de recrutement sur les réseaux sociaux	111
Annexe 13. Avis de la commission de vigilance éthique du département des sciences de l'éducation	112
Annexe 14. Formulaire de consentement.....	113

TABLE DES TABLEAUX

Tableau 1. Organisation des douze groupes expérimentaux.....	116
Tableau 2. Choix des trois prénoms rédacteurs et les stéréotypes susceptibles d'être activés.....	116
Tableau 3. Les deux visées de correction d'évaluation.....	117
Tableau 4. Répartition des 275 sujets dans les douze groupes expérimentaux	117
Tableau 5. Répartition des sujets dans les différentes catégories de l'échantillon	117
Tableau 6. Analyse de la variance simple pour la note totale.....	117
Tableau 7. Tableau récapitulatif reprenant le nombre de copies, la moyenne, l'écart-type, la médiane, le minimum, le maximum et l'amplitude pour chacune des 9 copies	118
Tableau 8. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Lucien selon le niveau de la copie	118
Tableau 9. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale	118
Tableau 10. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Lucien et Ayoub selon le niveau de la copie	119

Tableau 11. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Ayoub selon le niveau de la copie	119
Tableau 12. Moyenne (/15) et écart-type des notes et ampleur de l'effet pour chacune des copies selon la visée de l'évaluation.....	119
Tableau 13. Différence de moyenne et ampleur de l'effet pour les deux visées de l'évaluation.....	120
Tableau 14. Régression linéaire de la note totale selon le prénom, le niveau de la copie et la visée de l'évaluation	120
Tableau 15. Moyenne (/15), écart-type et variance des neuf copies selon la visée de l'évaluation .	121
Tableau 16. Moyennes (/15) et écarts-types des notes pour chacune des neuf copies.....	121
Tableau 17. Moyenne, écart-type, et ampleur de l'effet des scores totaux obtenus par Ayoub, Jordan et Lucien selon le niveau de la copie.....	121

TABLE DES FIGURES

Figure 1. Présentation des analyses statistiques pour chacune des cinq hypothèses.....	122
Figure 2. Répartition des scores attribués aux copies faibles	122
Figure 3. Répartition des scores attribués aux copies médianes	122
Figure 4. Répartition des scores attribués aux copies fortes	123
Figure 5. Scores moyens de Jordan et de Lucien selon le niveau de la copie	123
Figure 6. Scores moyens d'Ayoub et de Lucien selon le niveau de la copie	124
Figure 7. Scores moyens de Jordan et de Ayoub selon le niveau de la copie	124

Annexe 0. La consigne de rédaction



Tu as décidé d'assister au spectacle proposé sur l'affiche ci-dessus, mais tu n'as pas envie d'y aller seul.

RÉDIGE une lettre à un ami ou une amie pour qu'il t'accompagne.

- Écris au moins deux arguments pour le **convaincre** de t'accompagner au spectacle en utilisant les informations présentes sur l'affiche.
- Pense à fournir tous les renseignements nécessaires à ton ami pour pouvoir assister au spectacle.

Annexe 1. L'ensemble des rédactions d'une classe de P5

Prénom :

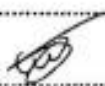
course sa l'intérêt de venir voir un
spectacle à 15h avec moi ?

C'est 1, rue de Charbonnage - Charleville, Tables de la
Fontaine, aff^é précieux fabuleux



Prénom :

Cécile je voudrais t'inviter à un spectacle de Jean de la Fontaine. Il paraît qu'il y aura des effets spéciaux et en plus ils passeront la Fable du loup et de l'agneau. Bref se serait chouette que tu viennes. C'est le 21 mai à 15 h la ville : Le petit Vanderille du Coron 1, rue du charbonnage - Charleroi. Ne t'inquiètes pas pour la réservation je m'en occupe.



Prénom :

Cher

Comme je sais que tu aimes les fables (surtout celle de Jean de La Fontaine) je t'invite au spectacle des Fables de La Fontaine le 21 mai à 15h, il se déroule à Petit Vanderille du Boson (1, rue du charbonnage - Charleville) tout près de chez toi. Il y aura plein d'effets ^{spéciaux} fabuleux comme tu les aimes. Voilà c'est tout pour moi, fait moi signe si ça t'intéresse.

À bientôt mon ami



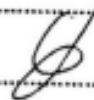
Prénom :

Cher Maillie,
je t'invite à venir au spectacle
Tables de la Fontaine avec moi, il
y aura des effets spéciaux fabuleux et
des animaux. Ça se passera le
21 mai - 15h

Adresse

Le Petit Vauverville du Caron
1, rue du Charbonnage - Charlevoix

J'espère que tu viendras lisais.



Prénom :

Chère Safia,
voudrais-tu m'accompagner à un
spectacle sur les fables de la Fontaine?
Apparemment, les effets spéciaux sont
fabuleux et beaucoup de fables seront
interprétées comme le loup et l'agneau.
Si tu veux venir avec moi, ça se passera
le 21 mai à 15h 1 rue du charbonnage au
Petit Vaudeville du Coron à Charleville.
J'attends ta réponse avec impatience.

Prénom :

Cher Martin

Je t'envoie cette lettre pour te demander si tu voulais
venir avec moi au spectacle de la Fables de la
Fontaine. Y'a des effets spéciaux fabuleux, c'est le 21
mars à 15h00, pour réserver : Le Petit Kouderville du Coran
: 071/310964. L'emplacement 1, rue du charbonnage - Ghent
Je sens que ça va être génial

A

Prénom :

Cher Oscar je t'écris cette lettre pour t'inviter à un spectacle sur les feux de la Fontaine. Il y aura des effets spéciaux folichons et des décors magnifiques. Si tu veux se sera le 21 mai à 15h, rue du Sabotage à Charlevoix dans le petit Vauclair du Garon.

Prénom :

Bonjour Lina je t'invite à aller au
spectacle des fables de la fontaine Je t'assure
ça sera fabuleux je te le promets je te donne
les coordonnées : Le Petit Pâquerette du Corson 1,
rue du charbonnage - Charleville le 21 mai
à 15 h les Effets spéciaux seront fabuleux
les fables aussi, la réservation : 071310964
j'espère que tu en auras Bises ♥

Prénom :

Cher Roger

13 mai 2022

Mon ami je t'invite le 21 mai pour les fêtes
de la fontaine au petit moulin de la corne.

Sur la réservation c'est au 71/31 22 64.

Apparemment les effets spéciaux sont fabuleux.

Et deux choses pour toi c'est à 15 h et tu n'a rien

de prévu à cette heure là 2^{ème} chose c'est tous près de

chez toi. ☺

Rue du charbonnage - charlevé

ton ami d'occasion

Prénom :

Cher, grand Ben, je t'invite à m'accompagner
au spectacle des Fables de la Fontaine.

Il y aura beaucoup d'effets spéciaux
fabuleux. C'est 1, rue du Charbonnage
Charleroi au Petit Vaudeville du Corin.

Le numéro de téléphone pour les réservations
est au 071310969. Il y aura des illustrations.
C'est le 21 mai à 15h.

✓

Prénom :

Cher Mille

Veux-tu venir avec moi aux Fables
de la Fontaine? ses le 21 mai à 15h,
en plus il pourrait qu'il aura des effets
spéciaux fabuleux! et ses a Charles
tout près de chez toi ses pas trop loin,
si tu veux ont se rejoint la - bas
vers 14 h 30 à l'avenue, l'affiche
est trop cool et donne envie!
tiens pour réserver ses aux: 071131 02
64, rue du Charbonnage je
t'attendrai la - bas
à plus tard!



Prénom :

Gauvain, sa te dit de venir avec moi à un spectacle ?

C'est le 21 mai à 15h

Le nom du spectacle, ses les Jolies de la fontaine.

C'est des effets spéciaux et folles.

L'endroit du spectacle ses Le petit Vaudouille du
Corm, Rue du Charbonnage à Chorberei numéro 1.

Il y aura des moutons, des grenouilles, des roche ect...

Dit moi si sa te dit de venir avec moi comme sa
je pourrai réserver merci d'avance ☺

Prénom :

Bonjour Liman, je ne veux pas aller
toute seule au spectacle (Fables
de la Fontaine) veut-tu bien venir
avec moi. d'adresse (1, rue du Charbonnage-
Baskrai) pour la reservation appelle-le
(071/310964) de 21 mai à 15 h.

Prénom :

Bonjour jardy,

Le 21 mai à 15h je vais voir un spectacle, tu
veux venir avec moi ?

Il y aura des effets spéciaux fabuleux.

Il y aura aussi des fables comme le loup et
l'agneau ou se passera au petit vaudeville à
charlerois 1, rue du charbonnage.

~~rejoins~~

Prénom :

Philippe, veut tu venir à un spectacle il y aura
des effets spéciaux fabuleux et aussi ça se déroulera
à la petit vaudeville du coron. Et c'est le 21 mai à
15h et c'est à Charleroi et aussi ça sera sur la scène
il aura des loups, des mastom, des grenouilles, des
vaches et c'est la fête de la fontaine et aussi
il faut réserver au 071/31 09 64 c'est la rue
9 rue du Charbonnage à merci d'avance j'ai hâte
!!!



Prénom :

Coucou Vincent, j'espère que tu vas bien.
Le 21 mai à 15h j'ai espéré que tu viennes avec
moi à un spectacle, il paraît très cool.
j'espère que tu viendras avec moi.

Les informations:

21 mai à 15h

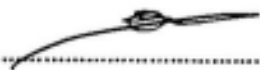
Le Petit Vaudeville du Corom.

Réservation : 071310964

1, rue du charbonnage - CHARLEROI.

C'est l'histoire du loup et de l'agneau.

A bientôt Vincent



Prénom :

Je t'invite pour m'accompagner au spectacle
(fables de la fontaine) le 15 mai à 15 h.
C'est : 1, rue du charbonnage - CHARLEROI -



Prénom :

Bonjour cher ami,

je me demandais si tu veux bien m'accompagner à
un spectacle sur la fête de la Fontaine. Il y aura des efforts
spéciaux Abulcux, une banderole d'annonce chapeaubable, et
je n'ai pas vraiment envie d'y aller seul. Si tu ne veux pas
en ville mais à 16h au "Le petit Vaudeville" du coron 1,
sur du charbonnage - charbonnier. S'il te plaît vien.
Merci beaucoup à bientôt.

AB

Prénom :

Bonjour mon amie, j'aimerais que tu viennes avec moi à un spectacle, qui aura lieu le 21 mai à 15 h, 1, rue du charbonnage à Charleroi. Viens voir les Fables de la Fontaine avec moi.

Il y aura des effets spéciaux fabuleux. Viens, ça peut être drôle, « le petit Vadeville du Coran ». Il faut réserver et je n'en perds rien si y allez seule.

ton amie

Prénom :

Tahira, je t'invite au spectacle des fables de la fontaine.
Le spectacle a l'air super! Il va avoir de super effets
spéciaux. Allez rendez vous au Petit Vaudeville du Louvre,
Si oui, 1, rue du Charbonnage - Charleroi, le 21 mai, 15h
• PS, j'ai hâte!

B

Annexe 2. La rédaction jugée comme étant « faible »

Prénom :

l'ouise sa l'intérêt de venir voir un
spectacle a 15h avec moi ?

c'est 1, rue de charbonnage - charbon, Tables de la
Fontaine, affⁱ plusieurs fabriques



Annexe 2A. Ayoub faible

n° de l'élève : 11

Prénom : Ayoub

Leona Sabahed sa l'intéresse de venir voir un
spectacle a 15h avec moi ?

C'est 1 rue du charbonnage - charbon, Fables de la
Fontaine, ~~flot~~ plusieurs fabuleux



Annexe 2B. Jordan faible

n° de l'élève: 15

Prénom : Jordan

l'ancien bassin sa l'intention de venir voir un
spectacle a 15h avec moi ?

c'est 1, rue du charbonnage - charbon, Tables de la
fontaine, ~~100~~ plusieurs fabriques



Annexe 2C. Lucien faible

n° de l'élève: 19

Prénom : Lucien

coucou Yvan sa t'intéresse de venir voir un
spectacle a 15h avec moi ?

c'est 1, rue du charbonnage - charlevi, Fables de la
Fontaine, ~~app~~ spectacles fabuleux



Annexe 3. La rédaction jugée comme étant « médiane »

Prénom :

Cocou je voudrais t'inviter à un spectacle de Jean de la Fontaine. Il paraît qu'il y aura des effets spéciaux et en plus ils passeront la Fable du loup et de l'agneau. Bref se sera chouette que tu viennes. C'est le 21 mai à 15 h la ville : Le petit Vaudeville du Coron 1, rue du charbonnage - Charleroi. Ne t'inquiètes pas pour la réservation je m'en occupe.

Annexe 3A. Ayoub médian

Prénom : Ayoub

n° de l'élève: 21

Coucou Stahed, je voudrais t'inviter à un spectacle de Jean de la Fontaine. Il paraît qu'il y aura des effets spéciaux et en plus ils passeront la Fable du loup et de l'agneau bref se sera chouette que tu viennes. C'est le 21 mai à 15h la ville : Le petit Vauderille du Coron 1, rue du charbonnage - Charleroi. Ne t'inquiète pas pour la réservation je m'en occupe.



Annexe 3B. Jordan médian

Prénom : Jordan

n° de l'élève : 25

Coucou Kevin, je voudrais t'inviter à un spectacle de Jean de la Fontaine. Il paraît qu'il y aura des effets spéciaux et en plus ils passeront la Fable du loup et de l'agneau bref se sera chouette que tu viennes. C'est le 21 mai à 15h la ville : Le petit Vaudeville du Coron 1, rue du charbonnage - Charleroi. Ne t'inquiète pas pour la réservation je m'en occupe.



Annexe 3C. Lucien médian

Prénom :

Lucien

n° de l'élève: 29

Coucou Ivan, je voudrais t'inviter à un spectacle de Jean de la Fontaine. Il paraît qu'il y aura des effets spéciaux et en plus ils passeront la Fable du loup et de l'agneau. Bref se serai chouette que tu viennes. C'est le 21 mai à 15h la ville : Le petit Vaudeville du Coron 1, rue du charbonnage - Charleroi. Ne t'inquiètes pas pour la réservation je m'en occupe.

Annexe 4. La rédaction jugée comme étant « forte »

Prénom :

Cher

Comme je sais que tu aimes les fables (notamment celle de Jean de La Fontaine) je t'invite au spectacle des Fables de La Fontaine le 21 mai à 15h, il se déroulera à Petit Vaudeville du Boron (1, rue du charbonnier - Charleville) tout près de chez toi. Il y aura plein d'effets ^{spéciaux} fabuleux comme tu les aimes. Voilà c'est tout pour moi, fait moi signe si ça t'intéresse.

A bientôt mon ami


★

Annexe 4A. Ayoub fort

n° de l'élève: 31

Prénom : Ayoub

Chère Hahed,
Comme je sais que tu aimes les fables (surtout celle de Zou de La Fontaine) je t'invite au spectacle des Fables de La Fontaine le 21 mai à 15h, il se déroule à Petit Vandœuvre du Bon (1, rue du charbonnier - Charleville) tout près de chez toi. Il y aura plusieurs effets fabuleux, comme tu les aimes. Voilà c'est tout pour moi, fait moi signe si ça t'intéresse.

A bientôt mon ami.


Annexe 4B. Jordan fort

Prénom : Jordan

n° de l'élève: 35

Cher Yéouing,

Comme je sais que tu aimes les fables (surtout celle de Jean de La Fontaine) je t'invite au spectacle des Fables de La Fontaine le 21 mai à 15h, il se déroule à Petit Vanderille du Bosc (1, rue du Châlonnais - Châlonnais) tout près de chez toi. Il y aura plusieurs effets fabuleux comme tu les aimes. Voilà c'est tout pour moi, fait moi signe si ça t'intéresse.

A bientôt mon ami.



Annexe 4C. Lucien fort

n° de l'élève: 39

Prénom : Lucien

Cher Yvan,
Comme je sais que tu aimes les fables (notamment celle de Yon de La Fontaine) je t'invite au spectacle des Fables de La Fontaine le 21 mai à 15h, il se déroule à Petit Vanderelle du Boron (1, rue du charbonnier - Charleville) tout près de chez toi. Il y aura plein d'effets ^{spéciaux} fabuleux comme tu les aimes. Voilà c'est tout pour moi, fait moi signe si ça t'intéresse.

A bientôt mon ami.



Annexe 5. La grille d'évaluation formative

Prénom de l'élève :

Numéro de l'élève :

Critère n° :	Indicateurs :	Score (à entourer) :	Commentaires :
Compétence I : Orienter son écrit en fonction de la situation de communication			
1	Texte facile à déchiffrer	1	
	Texte difficile à déchiffrer	0	
2	Quatre éléments de la lettre attendus : en tête – corps – formule de politesse - signature	2	
	Présence d'un ou deux éléments	1	
	Aucun élément spécifique à la mise en page d'une lettre	0	
3	Ton adapté à la situation (ton familier, tutoiement,...)	1	
	Ton inadapté à la situation	0	
Compétence II : Elaborer des contenus			
4	Présence des éléments essentiels à l'invitation (Où ? Quand ? Quoi ?)	2	
	Présence d'un ou deux éléments	1	
	Absence des éléments essentiels à l'invitation	0	
5	Présence de deux arguments pour convaincre le destinataire	2	
	Présence d'un seul argument	1	
	Absence d'argument	0	
6	Les arguments sont développés	1	
	Les arguments sont simplement recopiés de l'affiche	0	

Compétence III : Assurer l'organisation et la cohérence d'un texte			
7	Choix d'un système de mode et de temps approprié	2	
	Une ou deux erreurs dans le choix des temps	1	
	Système de mode et de temps inapproprié	0	
Compétence IV : Utiliser les unités grammaticales			
8	Pas d'erreur de syntaxe	2	
	Une ou deux erreurs de syntaxe	1	
	Plus de deux erreurs de syntaxe	0	
9	Pas d'erreur de ponctuation	2	
	Une erreur de ponctuation	1	
	Au moins deux erreurs de ponctuation	0	
SCORE FINAL (faire la somme de toutes les pondérations et inscrire le total ici)	 / 15	

Annexe 6. La grille d'évaluation normative

Prénom de l'élève :

Numéro de l'élève :

Critère n° :	Indicateurs :	Score (à entourer) :
Compétence I : Orienter son écrit en fonction de la situation de communication		
1	Texte facile à déchiffrer	1
	Texte difficile à déchiffrer	0
2	Quatre éléments de la lettre attendus : en tête – corps – formule de politesse - signature	2
	Présence d'un ou deux éléments	1
	Aucun élément spécifique à la mise en page d'une lettre	0
3	Ton adapté à la situation (ton familier, tutoiement,...)	1
	Ton inadapté à la situation	0
Compétence II : Elaborer des contenus		
4	Présence des éléments essentiels à l'invitation (Où ? Quand ? Quoi ?)	2
	Présence d'un ou deux éléments	1
	Absence des éléments essentiels à l'invitation	0
5	Présence de deux arguments pour convaincre le destinataire	2
	Présence d'un seul argument	1
	Absence d'argument	0
6	Les arguments sont développés	1
	Les arguments sont simplement copiés de l'affiche	0
Compétence III : Assurer l'organisation et la cohérence d'un texte		
7	Choix d'un système de mode et de temps approprié	2
	Une ou deux erreurs dans le choix des temps	1
	Système de mode et de temps inapproprié	0
Compétence IV : Utiliser les unités grammaticales		
8	Pas d'erreur de syntaxe	2
	Une ou deux erreurs de syntaxe	1
	Plus de deux erreurs de syntaxe	0
9	Pas d'erreur de ponctuation	2
	Une erreur de ponctuation	1
	Au moins deux erreurs de ponctuation	0
SCORE FINAL (faire la somme de toutes les pondérations et inscrire le total ici)	 / 15

Annexe 7. Les consignes générales (version formative) à destination des étudiants

Consignes générales

Chers étudiants,

Vous disposez des documents suivants :

- ✓ Les consignes de la rédaction à suivre par l'élève ;
- ✓ Trois rédactions d'élèves ;
- ✓ La grille de correction (en trois exemplaires) : une pour chaque rédaction.

Voici quelques consignes à suivre afin de mener à bien ce test.

✓ Corriger les rédactions à l'aide de la grille de correction

Chaque rédaction que vous allez corriger permettra à l'élève de prendre conscience de ses acquis ainsi que des points à retravailler grâce aux commentaires que vous allez rédiger dans le but qu'il puisse s'améliorer.

Dans un premier temps, il vous est demandé de prendre connaissance de la consigne de rédaction ainsi que des rédactions.

Dans un second temps, vous devez compléter la grille de correction pour chacun des trois élèves. N'oubliez pas d'inscrire le **prénom de l'élève** ainsi que son **numéro** dans les cases prévues. Lors de la correction, vous devez **entourer un et un seul** score par critère. Essayez autant que faire se peut de rédiger un commentaire dans la colonne prévue pour que l'enfant puisse comprendre ce qui est acquis et ce qu'il doit retravailler.

Une fois que vous aurez parcouru l'ensemble des critères, vous devez additionner les différents scores entourés et **indiquer le score /15** en bas du tableau.

Annexe 8. Les consignes générales (version normative) à destination des étudiants

Consignes générales

Chers étudiants,

Vous disposez des documents suivants :

- ✓ Les consignes de la rédaction à suivre par l'élève ;
- ✓ Trois rédactions d'élèves ;
- ✓ La grille de correction (en trois exemplaires) : une pour chaque rédaction.

Voici quelques consignes à suivre afin de mener à bien ce test.

✓ **Corriger les rédactions à l'aide de la grille de correction**

Chaque rédaction que vous allez corriger permettra de définir si l'élève est promu ou bien s'il doit recommencer son année.

Dans un premier temps, il vous est demandé de prendre connaissance de la consigne de rédaction ainsi que des rédactions.

Dans un second temps, vous devez compléter la grille de correction pour chacun des trois élèves. N'oubliez pas d'inscrire le **prénom de l'élève** ainsi que **son numéro** dans les cases prévues. Lors de la correction, vous devez **entourer un et un seul** score par critère.

Une fois que vous aurez parcouru l'ensemble des critères, vous devez additionner les différents scores entourés et **indiquer le score /15** en bas du tableau.

Annexe 9. Les consignes générales (version formative) à destination des enseignants

Consignes générales

Chers participants,

Tout d'abord, je tiens à vous remercier pour votre collaboration et votre aide qui me sera précieuse dans l'élaboration de mon mémoire.

Vous disposez des documents suivants :

- ✓ Les consignes de rédaction à suivre par les élèves ;
- ✓ Trois rédactions d'élèves ;
- ✓ Trois grilles de correction ;
- ✓ La lettre d'information et de consentement.

Voici quelques consignes à suivre afin de mener à bien ce test.

1. COMPLÉTER LE DOCUMENT DE CONSENTEMENT

Avant de commencer ce test, merci de lire et de compléter le document de consentement. Les parties du document à compléter se trouvent aux points B et C. En fin de document, vous devez dater, signer et indiquer votre nom et prénom.

2. CORRIGER LES RÉDACTIONS À L'AIDE DE LA GRILLE DE CORRECTION

Chaque rédaction que vous allez corriger permettra à l'élève de prendre conscience de ses acquis ainsi que des points à retravailler grâce aux commentaires que vous allez rédiger dans le but qu'il puisse s'améliorer.

Dans un premier temps, il vous est demandé de prendre connaissance de la consigne de rédaction ainsi que des rédactions.

Dans un second temps, vous devez compléter la grille de correction pour chacun des trois élèves. N'oubliez pas d'inscrire le **prénom de l'élève** ainsi que son **numéro** dans les cases prévues. Lors de la correction, vous devez **entourer un et un seul** score par critère. Essayez autant que faire se peut de rédiger un commentaire dans la colonne prévue pour que l'enfant puisse comprendre ce qui est acquis et ce qu'il doit retravailler.

Une fois que vous aurez parcouru l'ensemble des critères, vous devez additionner les différents scores entourés et **indiquer le score /15** en bas du tableau.

Annexe 10. Les consignes générales (version normative) à destination des enseignants

Consignes générales

Chers participants,

Tout d'abord, je tiens à vous remercier pour votre collaboration et votre aide qui me sera précieuse dans l'élaboration de mon mémoire.

Vous disposez des documents suivants :

- ✓ Les consignes de rédaction à suivre par les élèves ;
- ✓ Trois rédactions d'élèves ;
- ✓ Trois grilles de correction ;
- ✓ La lettre d'information et de consentement.

Voici quelques consignes à suivre afin de mener à bien ce test.

3. COMPLÉTER LE DOCUMENT DE CONSENTEMENT

Avant de commencer ce test, merci de lire et de compléter le document de consentement. Les parties du document à compléter se trouvent aux points B et C. En fin de document, vous devez dater, signer et indiquer votre nom et prénom.

4. CORRIGER LES RÉDACTIONS À L'AIDE DE LA GRILLE DE CORRECTION

Chaque rédaction que vous allez corriger permettra de définir si l'élève est promu ou bien s'il doit recommencer son année.

Dans un premier temps, il vous est demandé de prendre connaissance de la consigne de rédaction ainsi que des rédactions.

Dans un second temps, vous devez compléter la grille de correction pour chacun des trois élèves. N'oubliez pas d'inscrire le **prénom de l'élève** ainsi que **son numéro** dans les cases prévues. Lors de la correction, vous devez **entourer un et un seul** score par critère.

Une fois que vous aurez parcouru l'ensemble des critères, vous devez additionner les différents scores entourés et **indiquer le score /15** en bas du tableau.

Annexe 11. Courrier à l'intention des directions



Andenne, le 27 octobre 2022

Lettre d'information adressée aux directions du fondamental et du secondaire

Dans le cadre de la recherche intitulée : « Mesurer les facteurs qui interviennent dans l'évaluation d'une rédaction, en vue d'améliorer les pratiques évaluatives »

Étudiante : LABOUREUR Lola, étudiante au master en Sciences de l'Éducation à l'Université de Liège.
Adresse courriel : lola.laboureur@student.uliege.be

Promotrice : BAYE Ariane, professeure et directrice du service Analyses et Interventions dans les domaines du Décrochage et de l'Exclusion (AIDE) à l'Université de Liège.
Adresse courriel : ariane.baye@uliege.be

Madame la Directrice, Monsieur le Directeur,

Dans le cadre de mon mémoire en Sciences de l'éducation, je m'intéresse aux pratiques évaluatives. Ce projet vise à mesurer les différents facteurs qui interviennent dans l'évaluation d'une rédaction. Les résultats seront publiés dans un mémoire. Je souhaite recruter environ 180 participants durant les mois de décembre 2022 à juin 2023.

Je suis à la recherche d'enseignants du fondamental et de professeurs du secondaire (mathématiques – français – sciences – géographie – histoire) qui se porteraient volontaires pour m'aider dans ma recherche. Pourriez-vous distribuer la lettre d'information jointe à ce courrier à l'ensemble de vos enseignants ? De cette manière, les personnes intéressées pourront me contacter via mon adresse mail.

Je vous remercie d'avance pour votre collaboration et votre aide qui me sera précieuse.

Veillez agréer, Madame la Directrice, Monsieur le Directeur, l'expression de mes respectueuses salutations.

Lola Laboureur,
Étudiante à l'Université de Liège.

Annexe 12. Annonce de recrutement sur les réseaux sociaux



Bonjour à tous et toutes,

Dans le cadre de mon mémoire en Sciences de l'éducation, je m'intéresse aux facteurs qui interviennent dans l'évaluation d'une rédaction écrite.

Je suis à la recherche **d'instituteurs de l'enseignement fondamental et de professeurs de français de l'enseignement secondaire** qui seraient partants pour m'aider dans mon étude. Pour y participer, **merci de me contacter en privé** afin que je puisse vous envoyer les documents nécessaires.

Je vous rassure, cette tâche ne vous prendra pas plus de 40 minutes !

N'hésitez pas à partager pour m'aider à recruter un maximum de participants !

Annexe 13. Avis de la commission de vigilance éthique du département des sciences de l'éducation

AVIS DE LA COMMISSION DE VIGILANCE ÉTHIQUE DU DÉPARTEMENT DES SCIENCES DE L'ÉDUCATION

Après analyse du dossier d'éthique de sollicitude
présenté par LABOUREUR Lola, la commission de vigilance éthique
estime que le projet de recherche :

- N'entre pas dans le champ d'application de la loi sur l'expérimentation humaine.

La commission de vigilance éthique du département des sciences de l'éducation

- ~~valide les aspects éthiques de votre recherche.~~
 - valide les aspects éthiques de votre recherche, tout en vous invitant à prendre contact à votre promoteur, qui a reçu certaines précisions ou aménagements mineurs.
 - ~~refuse votre dossier en l'état actuel et vous demande d'apporter des précisions/modifications quant à votre projet de recherche (voir document attaché). Vous êtes donc invité à soumettre votre dossier à nouveau.~~
- Entre dans le champ d'application de la loi sur l'expérimentation humaine. Vous devez dès lors introduire une demande d'avis auprès du comité d'éthique facultaire, en suivant la procédure décrite sur l'Intranet du comité d'éthique facultaire (<https://intranet.fplse.uliege.be/Recherche/Ceth/ethique.php>).

DATE ET SIGNATURE
DE LA COMMISSION DE VIGILANCE ÉTHIQUE DU DÉPARTEMENT DES SCIENCES DE
L'ÉDUCATION

Le 11/11/2022



Annexe 14. Formulaire de consentement

Lettre d'information et de consentement adressée aux (futurs) enseignants du préscolaire, du primaire et du secondaire

Dans le cadre de la recherche intitulée : « Mesurer les facteurs qui interviennent dans l'évaluation d'une rédaction, en vue d'améliorer les pratiques évaluatives »

Étudiant.e : LABOUREUR Lola, étudiante au master en Sciences de l'Éducation à l'Université de Liège.
Adresse courriel : lola.laboureur@student.uliege.be

Promotrice : BAYE Ariane, professeure et directrice du service Analyses et Interventions dans les domaines du Décrochage et de l'Exclusion (AIDE) à l'Université de Liège.
Adresse courriel : ariane.baye@uliege.be

Vous êtes invité à participer à un projet de recherche. Votre participation est volontaire. Avant d'accepter, veuillez prendre le temps de lire ce document présentant les conditions de participation au projet. N'hésitez pas à poser toutes les questions que vous jugerez utiles à la personne qui vous a transmis ce document.

A. RENSEIGNEMENTS AU PARTICIPANT

1. Objectifs du projet de recherche

Ce projet vise à mesurer les différents facteurs qui interviennent dans l'évaluation d'une rédaction en vue d'améliorer les pratiques évaluatives. Les résultats seront publiés dans un mémoire. Nous souhaitons recruter environ 180 participants durant les mois de décembre 2022 à juin 2023.

Ce projet a été validé et analysé par la commission de vigilance éthique du département des Sciences de l'éducation de l'Université de Liège le 11 novembre 2022.

2. Participation à la recherche

Vous êtes sollicité pour participer à ce projet, car vous êtes un enseignant du préscolaire ou du primaire. Votre participation au projet de recherche est entièrement volontaire. Si vous y consentez, votre participation consiste à corriger la rédaction d'un élève à l'aide d'une grille d'évaluation qui sera complétée par vos soins à un moment et dans un lieu qui vous conviendront. Cette correction sera réalisée de façon anonyme, cela veut dire que nous ne vous demanderons pas de vous identifier lors de la réalisation de cette activité. La correction de cette évaluation durera entre 20 et 45 minutes.

3. Avantage et bénéfice

Vous contribuerez à une meilleure compréhension des critères évaluatifs dans le cadre d'une rédaction.

4. Risque et inconvénient

À notre connaissance, il n'y a pas de risque particulier associé à votre participation à ce projet.

5. Confidentialité et anonymat

L'étudiant prendra les mesures nécessaires afin que les renseignements personnels que vous lui donnerez demeurent confidentiels. Les moyens mobilisés pour ce faire sont ceux de la loi européenne du règlement général de protection des données (RGPD) :

- *Toutes les données récoltées dans le présent formulaire seront rendues anonymes, et cela dès que possible, en utilisant un code alphanumérique pour chaque participant et/ou institution. Il n'y aura aucune possibilité d'identifier des personnes, dans aucune des données traitées ni dans les publications des résultats.*
- *Les documents papier liés aux participants ou aux institutions où se déroule ma recherche, seront scannés et ne seront conservés que dans leur version numérique, en l'enregistrant sur un environnement informatique sécurisé.*
- *Les correspondances entre les codes/pseudonymes et les participants seront centralisées dans un fichier isolé et seront stockées uniquement sur une interface sécurisée par l'Université de Liège. Seuls le promoteur et l'étudiant responsable de la recherche ont un accès à cet espace, via un identifiant ULiège.*
- *Tous les renseignements personnels seront détruits de façon irrévocable après la collecte des données.*

6. Accessibilité et utilisation des données de recherche

Seul l'étudiant réalisant la recherche présentée plus haut, sa promotrice) et éventuellement les membres du jury de mémoire (pour validation de la démarche scientifique) auront accès aux données à caractère personnel. Aucune des données récoltées ou traitées ne sera transférée à un tiers hors Université.

Les données de recherche ne seront utilisées qu'aux fins de la présente recherche.

7. Conservation du dossier de recherche

En vertu des mesures de sécurité imposées par le département des sciences de l'éducation de l'Université de Liège, fondées sur le RGPD, la conservation des documents de recherche est fixée à deux ans. Cela signifie que les données d'identification de ce projet sont conservées jusqu'au 30 juin 2025.

8. Transmission des résultats aux participants

Aucun résultat ne sera communiqué aux participants.

9. Droit des participants

Votre participation à ce projet est entièrement volontaire et vous pouvez à tout moment vous retirer de la recherche sur simple avis verbal et sans devoir justifier votre décision, sans conséquence pour vous.

À votre demande, tous les renseignements personnels et les données déjà collectées pourront être détruits. Comme le prévoit le RGPD (Art. 15 à 23), chaque personne concernée par le traitement de données peut, en justifiant de son identité, exercer une série de droits :

- Obtenir, sans frais, une copie des données à caractère personnel la concernant faisant l'objet d'un traitement dans le cadre de la présente étude et, le cas échéant, toute information disponible sur leur finalité, leur origine et leur destination ;
- Obtenir, sans frais, la rectification de toute donnée à caractère personnel inexacte la concernant ainsi que les données incomplètes soient complétées ;
- Obtenir, sous réserve des conditions prévues par la réglementation et sans frais, la limitation du traitement de données à caractère personnel la concernant ;
- S'opposer, sous réserve des conditions prévues par la réglementation et sans frais, pour des raisons tenant à sa situation particulière, au traitement des données à caractère personnel la concernant;
- Introduire une réclamation auprès de l'Autorité de protection des données (contact@apd-gba.be, <https://www.autoriteprotectiondonnees.be>)

Pour exercer ces droits, vous pouvez vous adresser au Délégué à la protection des données de l'Université, soit par courrier électronique (dpo@uliege.be), soit par lettre datée et signée à l'adresse suivante :

Université de Liège
M. le Délégué à la protection des données,
Bât. B9 Cellule "GDPR",
Quartier Village 3,
Boulevard de Colonster 2,
4000 Liège, Belgique.

10. Assurance

Puisque la recherche à laquelle vous vous apprêtez à prendre part ne relève pas du champ d'application de la loi sur l'expérimentation humaine du 7 mai 2004, vous ne bénéficiez pas de l'assurance souscrite pour vous, liée à votre participation à cette recherche (assurance sans faute, couvrant les dommages directs ou indirects liés à la participation à votre étude). En revanche, l'étudiant-chercheur est couvert par une assurance en responsabilité civile en cas de dommages causés à un tiers ou aux biens d'un tiers.

B. DÉCLARATION DU PARTICIPANT

- Je reconnais qu'on m'a expliqué clairement la nature de ma participation à la recherche.

- Je comprends que je peux prendre mon temps pour réfléchir avant de donner mon consentement à participer à la recherche aux conditions énoncées dans le présent formulaire.

C. CONSENTEMENT DU PARTICIPANT

J'ai pris connaissance du présent document d'information et de consentement et, en posant ma signature, je consens à participer aux activités de recherche présentées dans la rubrique « 2. Participation à la recherche ».

J'accepte de participer de façon anonyme au projet de recherche selon les informations énoncées

Oui Non

Je consens à ce que l'étudiant utilise, les **données dépersonnalisées** pour d'autres projets de recherche de même nature et dans le respect des mêmes principes de confidentialité et de protection des informations.

Oui Non Non applicable

Signature du participant : _____ Date : _____

Nom : _____ Prénom : _____

Tableau 1. Organisation des douze groupes expérimentaux

Groupe expérimental n°	Formative (F) ou Normative (N)	Élève n°1	Élève n°2	Élève n°3
1	F	Ayoub FO 31	Lucien M 29	Jordan FAI 15
2	F	Ayoub FO 31	Jordan M 25	Lucien FAI 19
3	F	Lucien FAI 19	Ayoub M 21	Jordan FO 35
4	F	Lucien FO 39	Jordan FAI 15	Ayoub M 21
5	F	Jordan FO 35	Lucien M 29	Ayoub FAI 11
6	F	Jordan M 25	Ayoub FAI 11	Lucien FO 39
7	N	Ayoub FO 31	Lucien M 29	Jordan FAI 15
8	N	Ayoub FO 31	Jordan M 25	Lucien FAI 19
9	N	Lucien FAI 19	Ayoub M 21	Jordan FO 35
10	N	Lucien FO 39	Jordan FAI 15	Ayoub M 21
11	N	Jordan FO 35	Lucien M 29	Ayoub FAI 11
12	N	Jordan M 25	Ayoub FAI 11	Lucien FO 39
FO = Fort		M = Médian	FAI = Faible	

Tableau 2. Choix des trois prénoms rédacteurs et les stéréotypes susceptibles d'être activés

Prénoms des rédacteurs :	Stéréotypes susceptibles d'être activés :
Jordan	Défavorisé d'un point de vue socio-économique Élève autochtone
Ayoub (prénom à consonance maghrébine)	Défavorisé tant du point de vue ethnique que socio-économique Élève issu de l'immigration
Lucien	Favorisé d'un point de vue socio-économique Élève autochtone

Tableau 3. Les deux visées de correction d'évaluation

VISÉES :	CONSIGNES :
VISÉE FORMATIVE	L'ENSEIGNANT EST INVITÉ À CORRIGER UNE ÉVALUATION DANS LE BUT D'AIDER L'ÉLÈVE À S'AMÉLIORER.
VISÉE SÉLECTIVE	L'ENSEIGNANT EST INVITÉ À CORRIGER UNE ÉVALUATION DANS LE BUT DE DÉTERMINER SI L'ÉLÈVE EST EN MESURE DE RÉUSSIR SON ANNÉE SCOLAIRE OU BIEN S'IL DOIT RECOMMENCER SON ANNÉE.

Tableau 4. Répartition des 275 sujets dans les douze groupes expérimentaux

<i>Groupe n°</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>Nombre de sujets par groupe</i>	27	24	25	20	26	26	24	22	20	20	20	21

Tableau 5. Répartition des sujets dans les différentes catégories de l'échantillon

AESS	ISEF	ENSEIGNANTS	TOTAL
N = 106	N = 49	N = 120	275

Tableau 6. Analyse de la variance simple pour la note totale

Source	DL	Somme des carrés	Carré moyen	Valeur de F	Pr > F
Modèle (trait)	3	4 840.46	1 613.49	513.76	< 0.0001
Erreur	821	2 578.39	3.14		
Total corrigé	824	7 418.85			

R-carré	Coefficient de variation	Ecart quadratique moyen = Root MSE	Note moyenne
0.65	18.36	1.77	9.65

Tableau 7. Tableau récapitulatif reprenant le nombre de copies, la moyenne, l'écart-type, la médiane, le minimum, le maximum et l'amplitude pour chacune des 9 copies

9 copies	N=	Moyenne (écart-type)	Médiane	Minimum	Maximum	Amplitude/ Etendue
Ayoub Faible	93	6.16 (1.96)	6	1	12	11
Ayoub Moyen	85	11.07 (1.75)	11	5	14	9
Ayoub Fort	97	12.04 (1.46)	12	9	15	6
Jordan Faible	91	6.31 (1.57)	6	2	11	9
Jordan Moyen	93	9.86 (1.57)	10	6	14	8
Jordan Fort	91	12.44 (1.56)	13	9	15	6
Lucien Faible	91	6.81 (1.81)	7	2	10	8
Lucien Moyen	97	9.68 (1.94)	10	2	14	12
Lucien Fort	87	12.63 (1.49)	13	8	15	7

Tableau 8. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Lucien selon le niveau de la copie

Prénom Copie	Jordan (2)		Lucien (3)		2 VS 3
	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet
Faible	91	6.32 (1.57)	91	6.81 (1.81)	-0.29
Médiane	93	9.86 (1.57)	97	9.68 (1.94)	+0.10
Forte	91	12.44 (1.56)	87	12.63 (1.49)	-0.12

Tableau 9. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale

	DL	Carré moyen	Valeur de F	Pr > F
Prénom (PRE)	1	0.0001	0.00	0.99
Niveau de la copie (NIV)	1	4 835.61	1 539.73	< 0.0001
Visée de l'évaluation (VIS)	1	3.80	1.21	0.27

Tableau 10. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Lucien et Ayoub selon le niveau de la copie

Prénom Copie	Ayoub (1)		Lucien (3)		1 VS 3
	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet
Faible	93	6.16 (1.96)	91	6.81 (1.81)	-0.34
Médiane	85	11.07 (1.75)	97	9.68 (1.94)	+0.75
Forte	97	12.04 (1.46)	87	12.63 (1.49)	-0.40

Tableau 11. Moyenne, écart-type et ampleur de l'effet des scores totaux obtenus par Jordan et Ayoub selon le niveau de la copie

Prénom Copie	Ayoub (1)		Jordan (2)		1 VS 2
	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet
Faible	93	6.16 (1.96)	91	6.32 (1.57)	-0.09
Médiane	85	11.07 (1.75)	93	9.86 (1.57)	+0.73
Forte	97	12.04 (1.46)	91	12.44 (1.56)	-0.26

Tableau 12. Moyenne (/15) et écart-type des notes et ampleur de l'effet pour chacune des copies selon la visée de l'évaluation

Copie	Normative (1)		Formative (2)		Ampleur de l'effet (1) vs (2)
	Moyenne (ET)	N	Moyenne (ET)	N	
Ayoub faible	5.80 (2.08)	41	6.44 (1.84)	52	-0.33
Jordan faible	6.29 (1.49)	44	6.34 (1.66)	47	-0.03
Lucien faible	6.81 (1.97)	42	6.82 (1.68)	49	-0.005
Ayoub médian	11.07 (1.38)	40	11.07 (2.04)	45	0
Jordan médian	9.98 (1.76)	43	9.76 (1.39)	50	+0.14
Lucien médian	9.48 (1.99)	44	9.85 (1.89)	53	-0.19
Ayoub fort	11.83 (1.27)	46	12.23 (1.61)	51	-0.27
Jordan fort	12.35 (1.67)	40	12.51 (1.49)	51	-0.10
Lucien fort	12.73 (1.60)	41	12.54 (1.39)	46	+0.13

Tableau 13. Différence de moyenne et ampleur de l'effet pour les deux visées de l'évaluation

	Normative		Formative	
	Différence de moyenne (1) – (2)	Ampleur de l'effet (1) VS (2)	Différence de moyenne (1) – (2)	Ampleur de l'effet (1) VS (2)
<u>COPIES FAIBLES</u>				
Jordan (1) VS Lucien (2)	-0.52	-0.30	-0.48	-0.29
Ayoub (1) VS Lucien (2)	-1.01	-0.50	-0.38	-0.21
Ayoub (1) VS Jordan (2)	-0.49	-0.24	+0.10	+0.06
<u>COPIES MÉDIANES</u>				
Jordan (1) VS Lucien (2)	+0.50	+0.27	-0.09	-0.05
Ayoub (1) VS Lucien (2)	+1.59	+0.94	+1.22	+0.62
Ayoub (1) VS Jordan (2)	+1.09	+0.69	+1.31	+0.77
<u>COPIES FORTES</u>				
Jordan (1) VS Lucien (2)	-0.38	-0.23	-0.03	-0.02
Ayoub (1) VS Lucien (2)	-0.90	-0.63	-0.31	-0.21
Ayoub (1) VS Jordan (2)	-0.52	-0.35	-0.28	-0.18

Tableau 14. Régression linéaire de la note totale selon le prénom, le niveau de la copie et la visée de l'évaluation

Paramètre	Estimation	Erreur standard	Valeur t	Pr > t
Intercept	3.65	0.23	15.62	< 0.0001
Prénom (PRE)	-0.0005	0.07	-0.01	0.99
Niveau de la copie (NIV)	2.96	0.07	39.24	< 0.0001
Visée de l'évaluation (VIS)	0.14	0.12	1.10	0.27

Tableau 15. Moyenne (/15), écart-type et variance des neuf copies selon la visée de l'évaluation

Visée de l'évaluation	N	Moyenne (écart-type)	Variance	Ampleur de l'effet (1) vs (2)
Normative (1)	381	9.58 (3.04)	9.24	-0.05
Formative (2)	444	9.72 (2.97)	8.81	

En moyenne, les copies corrigées dans le cadre d'une évaluation normative sont évaluées plus négativement que celles corrigées dans le cadre d'une évaluation formative. Cependant, cet effet est très faible puisque l'ampleur de l'effet s'élève seulement à -0.05 et est proche de 0. La différence entre ces deux groupes est donc négligeable.

Tableau16. Moyennes (/15) et écarts-types des notes pour chacune des neuf copies

	Copie faible	Copie médiane	Copie forte
Ayoub	6.16 (1.96)	11.07 (1.75)	12.04 (1.46)
Jordan	6.32 (1.57)	9.86 (1.57)	12.44 (1.56)
Lucien	6.81 (1.81)	9.68 (1.94)	12.63 (1.49)

Tableau 17. Moyenne, écart-type, et ampleur de l'effet des scores totaux obtenus par Ayoub, Jordan et Lucien selon le niveau de la copie

Prénom \ Copie	Ayoub (1)		Jordan (2)		Lucien (3)		1 VS 2	1 VS 3	2 VS 3
	N	Moyenne (ET)	N	Moyenne (ET)	N	Moyenne (ET)	Ampleur de l'effet Hedges'g		
Faible	93	6.16 (1.96)	91	6.32 (1.57)	91	6.81 (1.81)	-0.09	-0.34	-0.29
Médiane	85	11.07 (1.75)	93	9.86 (1.57)	97	9.68 (1.94)	+0.73	+0.75	+0.10
Fort	97	12.04 (1.46)	91	12.44 (1.56)	87	12.63 (1.49)	-0.26	-0.40	-0.12

Figure 1. Présentation des analyses statistiques pour chacune des cinq hypothèses

Hypothèse n°1	Hypothèses n°2,3,4 et 5
<ul style="list-style-type: none">✓ Indices de tendance centrale✓ Indices de dispersion	<ul style="list-style-type: none">✓ Indices de tendance centrale✓ Indices de dispersion✓ Ampleur de l'effet✓ Analyse de la variance simple (ANOVA)

Figure 2. Répartition des scores attribués aux copies faibles

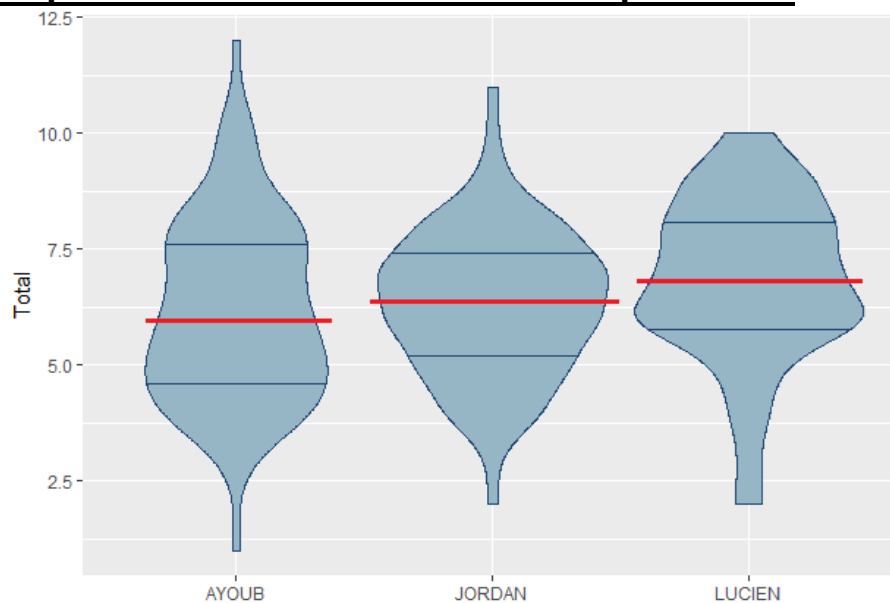


Figure 3. Répartition des scores attribués aux copies médianes

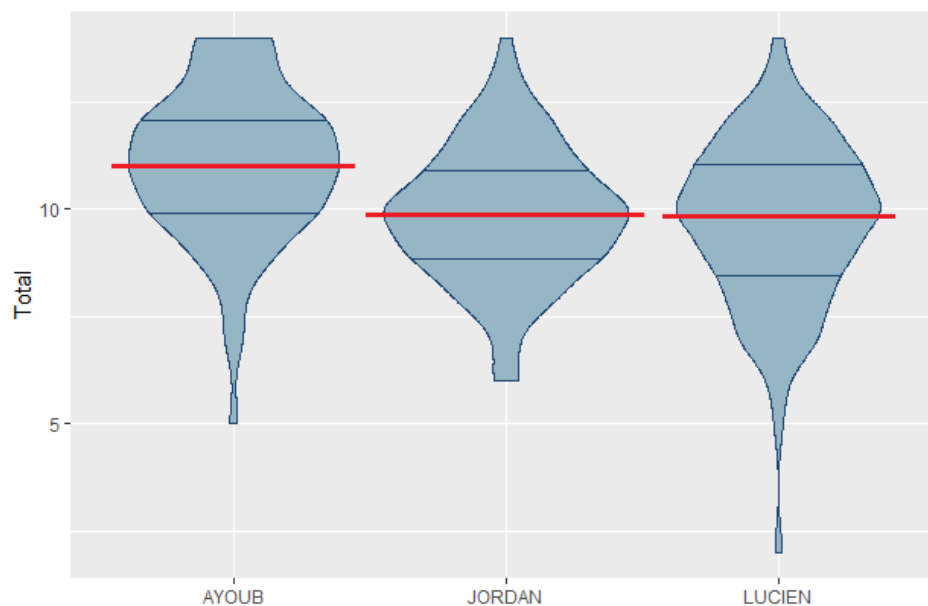


Figure 4. Répartition des scores attribués aux copies fortes

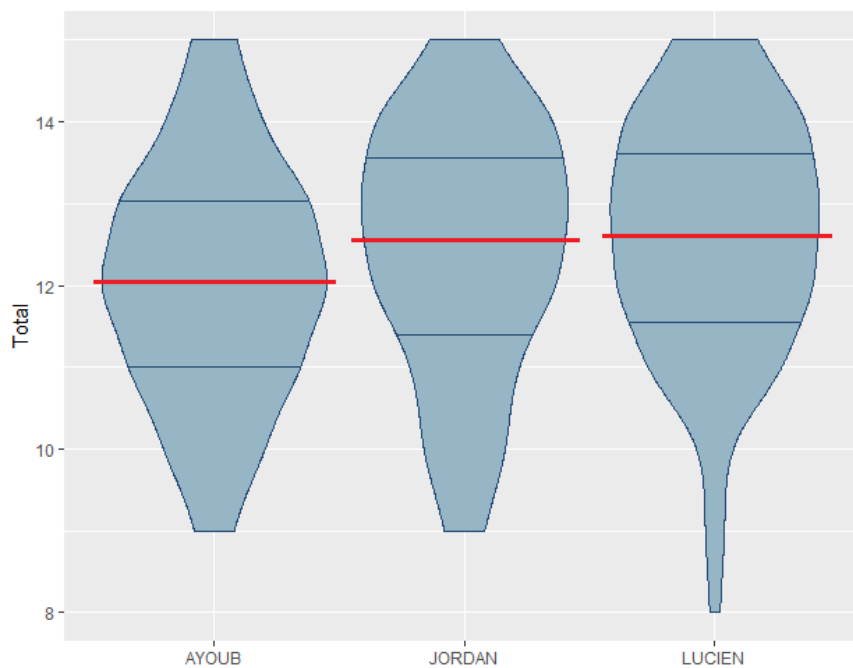


Figure 5. Scores moyens de Jordan et de Lucien selon le niveau de la copie

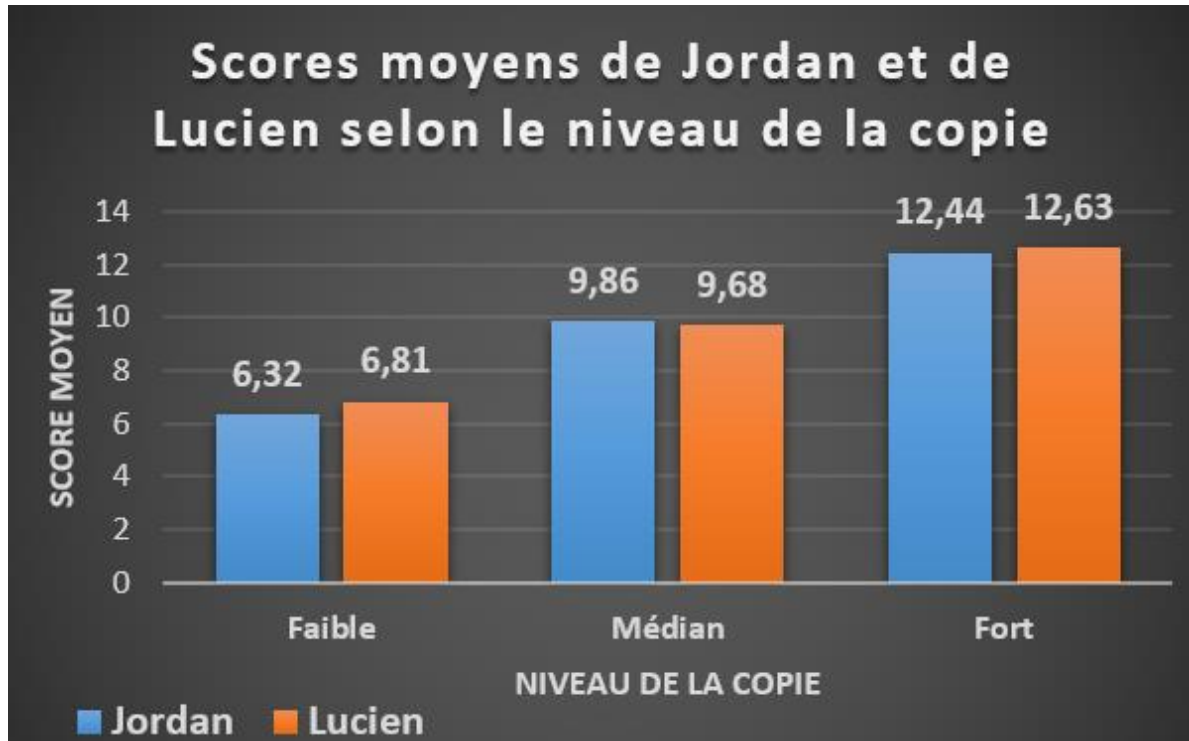


Figure 6. Scores moyens d'Ayoub et de Lucien selon le niveau de la copie

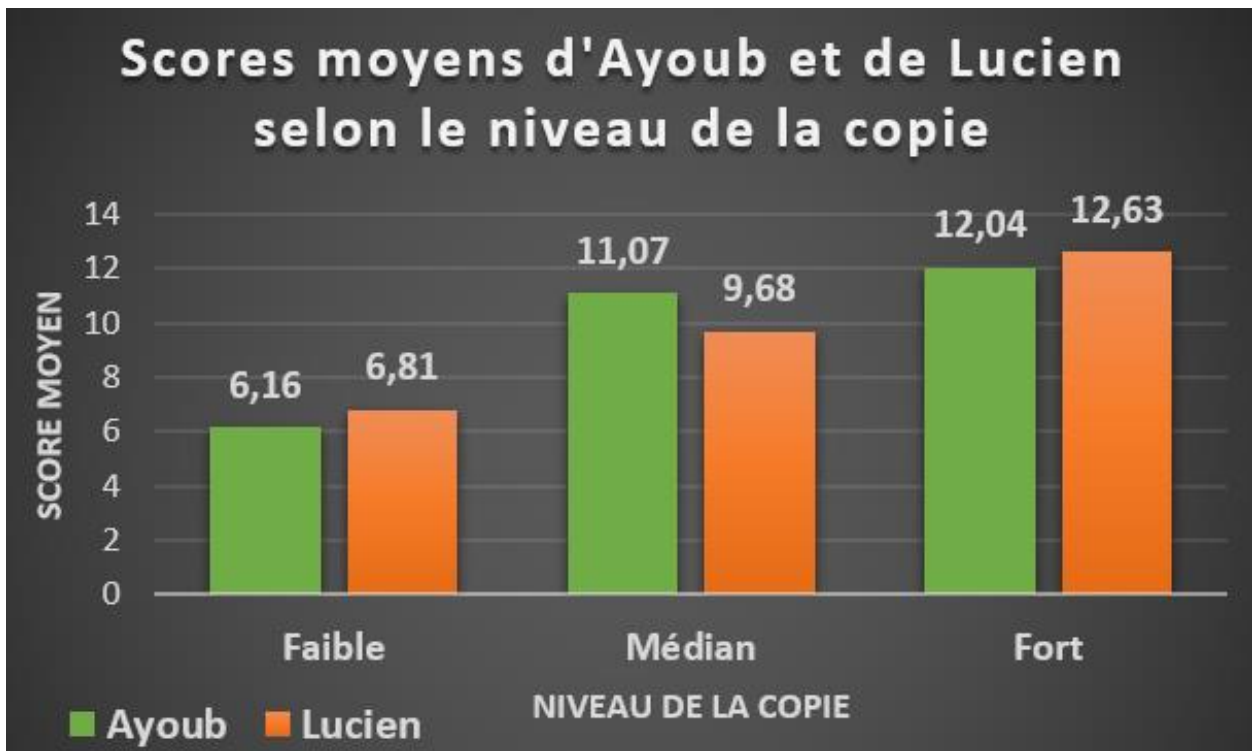


Figure 7. Scores moyens de Jordan et de Ayoub selon le niveau de la copie

