



Université de Liège
Faculté de psychologie, de Logopédie et des Sciences de l'Éducation
Département des Sciences de l'Éducation

Les évaluations et les stéréotypes liés au redoublement

Dans quelle mesure, l'ancrage du stéréotype du redoublement dans la copie d'un élève de 3^e année secondaire, accompagnée ou non d'une grille d'évaluation, va-t-il influencer l'évaluation donnée par des enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles sur cette copie ?

Étude quantitative menée auprès de 140 enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles.

Promotrice : Ariane Baye

Assistant doctorant : Dylan Dachet

Lecteurs : Christian Monseur et Quitte Valérie

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

Année académique 2021-2022



Université de Liège
Faculté de psychologie, de Logopédie et des Sciences de l'Éducation
Département des Sciences de l'Éducation

Les évaluations et les stéréotypes liés au redoublement

Dans quelle mesure, l'ancrage du stéréotype du redoublement dans la copie d'un élève de 3^e année secondaire, accompagnée ou non d'une grille d'évaluation, va-t-il influencer l'évaluation donnée par des enseignants de l'enseignement secondaire ordinaire en Fédération

Wallonie-Bruxelles sur cette copie ?

Étude quantitative menée auprès de 140 enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles.

Promotrice : Ariane Baye

Assistant doctorant : Dylan Dacht

Lecteurs : Christian Monseur et Quittre Valérie

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

Année académique 2021-2022

Remerciements

La rédaction de ce mémoire n'aurait pu aboutir sans l'aide de plusieurs personnes. Je souhaite vivement remercier.

Avant tout, je tiens à remercier chacune des personnes qui ont participé à ma recherche, sans qui ce mémoire n'aurait pu être réalisé.

Ensuite, je tiens à exprimer ma reconnaissance envers ma promotrice, Madame Baye, pour sa confiance en mon projet de recherche ainsi que pour sa supervision tout au long de la réalisation de celui-ci.

Je remercie ensuite chaleureusement, Monsieur Dachet, assistant doctorant de Madame Baye, pour sa disponibilité, ses relectures, ses encouragements et commentaires qui m'ont grandement aidée dans la mise en œuvre de mon projet. Sans sa précieuse aide, mon mémoire n'aurait pu être celui qu'il est actuellement.

Je voudrais également remercier chacun des membres de ma famille et de ma belle-famille pour leur soutien tout au long de la réalisation de ce mémoire. Un merci tout particulier à ma maman et à ma grande sœur pour leur positivisme, leurs encouragements et pour avoir toujours cru en moi et en mes capacités à réaliser ce Master en Sciences de l'Éducation.

Je remercie également mes amies Pauline Otten et Marie Dandoy pour leurs encouragements et pour la solidarité dont elles ont fait preuve car « On est toutes dans le même bateau, et il y en a plus dans trois têtes que dans une ».

Pour terminer, un énorme merci à mon compagnon Anthony Libert pour la patience dont il a fait preuve à mon égard et pour son amour inconditionnel.

Table des matières

1. Introduction	1
2. Revue de la littérature	2
2.1 Le redoublement	2
2.1.1 L'origine du redoublement	3
2.1.2 Le redoublement	4
2.1.3 Les effets du redoublement – la stéréotypisation	6
2.1.4 Pourquoi continue-t-on de faire redoubler malgré les effets négatifs collatéraux?	10
2.2 Les effets délétères des stéréotypes	14
2.2.1 Le phénomène de confirmation des attentes	14
2.2.2 La stigmatisation	16
2.2.3 La menace du stéréotype et l'effet solo	16
2.2.4 Le caractère inévitable des stéréotypes	18
2.3 L'évaluation	19
2.3.1 Le jugement professionnel et le jugement évaluatif	19
2.3.2 Le jugement subjectif	19
2.3.3. Les biais liés à l'évaluation	21
2.4 Les grilles d'évaluation	32
3. Objet de recherche et hypothèses	34
3.1 Objet de recherche	34
3.2 Hypothèses	35
3.2.1 Hypothèse n°1 :	35
3.2.2 Hypothèse n°2 :	36
3.2.3 Hypothèse n°3 :	36
3.2.4 Hypothèses relatives aux quatre variables sociodémographiques	37
3.2.5 Étude confirmatoire et exploratoire	39
4. Méthodologie	40
4.1 Design de la recherche	40
4.2 Instruments de mesure et données récoltées	42
4.2.1 Instruments de mesure	42
4.2.2 Données récoltées	45
4.3 Public cible et échantillon	45
4.4 Procédure de recrutement	47
4.5 Laboratoire cognitif	48
4.5.1 Résultats du laboratoire cognitif	48
4.6 Aspects éthiques et codage	51
4.7 Les analyses statistiques	53
4.7.1 Hypothèse n°1 et hypothèse n°3	53
4.7.2 Hypothèse n°2	55
4.7.3 Hypothèses relatives aux quatre variables sociodémographiques	56
5. Résultats et interprétations	56
5.1 Hypothèse n°1	58
5.2 Hypothèse n°2	62
5.3 Hypothèse n°3	63
5.4 Les 4 variables sociodémographiques	65
6. Discussion	68

6.1 Données sociodémographiques	68
6.2 Discussion autour de nos différentes hypothèses	69
7. Limites et perspectives	76
8. Conclusion	79
9. Bibliographie.....	82
10. Résumé.....	89
11. Annexes.....	90

1.Introduction

L'égalité est une valeur centrale en éducation. Tous les élèves, peu importe leurs caractéristiques, devraient pouvoir être traités de manière égale à l'école. Néanmoins, lorsque l'on s'intéresse à la qualité du jugement évaluatif des enseignants, les travaux en docimologie mettent en évidence quatre sources de biais susceptibles d'influencer le jugement évaluatif de l'évaluateur (Detroz, 2021) : 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. Ainsi, le jugement évaluatif de l'évaluateur peut être biaisé par une série de facteurs, dont les caractéristiques de l'élève. L'étude de Sprietsma (2013), par exemple, a permis de mettre en évidence l'influence que pouvait avoir l'origine ethnique de l'élève sur le jugement évaluatif de l'évaluateur dans le cadre d'une épreuve de français : les enseignants ont tendance à évaluer plus négativement une copie lorsque celle-ci est attribuée à un élève ayant un prénom qui ne correspond pas à la nationalité de l'enseignant. L'étude de Autin et ses collègues (2019), quant à elle, a mis en évidence que, dans une épreuve à visée sélective, les élèves de faible statut socio-économique ont tendance à être, en moyenne, moins bien évalués que les élèves de statut socio-économique élevé. Concernant l'effet de genre, l'étude de Lafontaine et Monseur (2009) met en lumière cette tendance selon laquelle les filles, dans une épreuve de mathématiques, sont, en moyenne, légèrement moins bien évaluées que les garçons.

Dans la continuité de ces travaux en docimologie, il nous semble opportun d'étudier l'influence que peut avoir l'existence d'un autre biais lié à l'évalué, à savoir l'effet que peut avoir le statut de **redoublant** sur le jugement évaluatif d'un évaluateur.

Un tel sujet d'étude prend, par ailleurs, d'autant plus son sens s'il est traité dans un système éducatif qui pratique massivement le redoublement, à l'instar du système éducatif de la FW-B (Boraita, 2016). En effet, malgré les résultats qui attestent de la faible efficacité du redoublement et de son caractère préjudiciable à plusieurs égards, des enseignants de nombreux pays continuent de recourir à cette pratique, comme si c'était le seul moyen pour gérer l'hétérogénéité scolaire. C'est notamment le cas dans notre pays, où le redoublement semble être considéré comme normal, comme la solution par défaut pour faire face aux difficultés d'apprentissage et comme faisant partie intégrante du mode de fonctionnement de notre système éducatif (Boraita, 2015; Goffin et al., 2013).

Dès lors, au même titre que les élèves migrants dans une épreuve de français, que les filles dans une épreuve de mathématiques, et que les élèves de statut socio-économique faible dans une épreuve à visée sélective, les élèves redoublants subissent-ils aussi des inégalités de traitement dans les évaluations ? En d'autres termes, **l'objectif** de cette recherche est de répondre à la question suivante : *Dans quelle mesure, l'ancrage du stéréotype du redoublement dans la copie d'un élève de 3^e année secondaire, accompagnée ou non d'une grille d'évaluation, va-t-il influencer l'évaluation donnée par des enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles sur cette copie ?*

C'est au travers d'une étude expérimentale, recrutant entre minimum 120 et maximum 200 enseignants du secondaire en FW-B, que nous tenterons de répondre à cette question. Plus précisément, pour tester l'influence que peut avoir le statut d'élève redoublant sur le jugement évaluatif d'un évaluateur, nous soumettrons à chacun des participants une rédaction à évaluer, avec ou sans grille d'évaluation, au sein de laquelle l'élève sera soit présenté comme étant redoublant, soit comme étant un élève promu. La méthodologie de cette étude sera plus amplement détaillée dans la quatrième partie de ce travail.

La présentation du présent mémoire s'articulera autour de 8 parties dont la première est la présente introduction. Celle-ci sera suivie d'une **revue de littérature** au sein de laquelle les notions suivantes seront développées : le redoublement, ses origines et ses conséquences, les stéréotypes et leurs effets délétères, les biais d'évaluation et les grilles d'évaluation. S'enchaînera à cette revue de la littérature, une partie dédiée à la **présentation de l'objectif et des hypothèses** de notre recherche. Ensuite, la **méthodologie** de notre étude sera présentée : les instruments de mesure, l'échantillon, la procédure de recrutement, les analyses statistiques, etc. Puis, nos **résultats** seront exposés et ensuite **discutés** à la lumière de la littérature de recherche. Enfin, nous terminerons ce travail par la présentation de la **conclusion générale**, des **limites et perspectives** de notre recherche.

2. Revue de la littérature

2.1 Le redoublement

Avant d'approfondir le concept du redoublement, il serait judicieux de comprendre son origine ; comment le redoublement est-il né au sein de notre société ? Dès lors, dans un premier temps, nous expliciterons la notion d'échec scolaire.

2.1.1 L'origine du redoublement

Isambert-Jamati (1992), intéressée par l'émergence du **concept d'échec scolaire** dans les milieux pédagogiques français, y a consacré une étude (Crahay, 2019). Toutefois, il faudra attendre les années 60 pour voir apparaître plus largement des articles à ce sujet. La sociologue française fait remarquer qu'avant les années 60, le terme « échec scolaire » n'était que rarement employé dans les articles, et quand il s'y trouvait, était utilisé pour désigner les enfants de bonne famille échouant dans leur perspective de poursuite de longues études. Hutmacher (1993, cité par Crahay, 2019) met en évidence que l'apparition du terme « d'échec scolaire » marque une transformation du sens donné à un phénomène déjà connu auparavant, un progrès dans les conceptions pédagogiques. En effet, à la fin du XIXe siècle et au début du XXe siècle, les élèves en difficulté étaient désignés comme des « vicieux, réfractaires, ascolaires, rebelles, asociaux, paresseux ou encore peu doués, inintelligents, débiles » (Hutmacher, 1992, cité par Crahay, 2019, pp. 24-25). À cette époque, l'explication des différences entre élèves n'envisageait pas la notion d'échec scolaire car « [...] il n'y a pas d'échec tant que l'on ne s'attend pas à une réussite, tant que l'explication renvoie à l'incapacité d'une fraction d'élèves, à des « tares » ou des « déficits » individuels sur lesquels l'école ne se donne pas prise » (Hutmacher, 1993, p.30). Aujourd'hui, dans la perspective d'une réussite pour tous, la place de l'échec scolaire prend une part plus importante. Par ailleurs, de nos jours, plus personne n'attribue l'échec scolaire à un « don » inné ou encore à une absence de don (Best, 1996). De Landersheere définit la notion d'échec scolaire comme une « Situation où un objectif éducatif n'a pas été atteint » (1992, p.91). Cette définition est encore valable à l'heure actuelle ; on la retrouve dans des textes officiels de différents pays et, notamment, en Belgique où le concept se définit par « des objectifs cognitifs non atteints. » (Eurydice, cité par Crahay, 2019). Dans cet esprit, la non-réussite renvoie donc à l'échec scolaire (Isambert-Jamati, 1992, citée par Crahay, 2019).

Si chaque élève a pour mission de réussir l'enseignement fondamental, il n'est pas concevable de les exclure de l'école comme cela a été fait par le passé (Isambert-Jamati, 1992, citée par Crahay, 2019). À l'époque, Bourdieu nommait ces élèves exclus les « exclus de l'intérieur » (Bourdieu & Champagne, 1992). En revanche, le **redoublement** apparaît comme un remède acceptable lorsque l'élève n'entre pas dans le canevas proposé par l'institution scolaire. Il redouble en vue de réussir ultérieurement (Isambert-Jamati, 1992, citée par Crahay, 2019).

Bien que l'« échec scolaire » soit devenu un terme employé par tout le monde, on ne connaît pas réellement toutes les notions cachées qui le constituent (Crahay 2019).

A ce terme « fourre-tout », Mérieu (2005) distingue deux types d'élèves vivant des réalités complètement différentes : « l'élève en difficulté » chez qui des remédiations ponctuelles permettraient de surmonter ses difficultés, et « l'élève en échec scolaire » nécessitant une prise en charge spécifique par des professionnels (logopèdes, enseignants supplémentaires, etc).

Perrenoud (1989), quant à lui, distingue trois registres qui interviennent dans la fabrication de l'échec scolaire. Le premier concerne le système scolaire qui, selon lui, génère ses propres représentations de réussite ou d'échec. Le second dénonce un phénomène inégalitaire selon lequel la réussite ou l'échec dépendrait de normes d'excellence. Et enfin, le dernier fait référence à l'école en mettant en évidence la contradiction qu'il y a entre la volonté de faire apprendre et une incapacité à y parvenir. Ce dernier registre sous-entend donc que l'école serait elle-même à l'origine de l'échec scolaire. En adhérant à ces registres, une culture de l'échec se verrait installée, et un système éducatif, pratiquant de manière massive le redoublement, se verrait d'application dans les écoles (Crahay, 2019), comme c'est le cas en Fédération Wallonie-Bruxelles (FW-B) (Boraita, 2016).

Ainsi, nous pouvons constater le lien étroit qui existe entre échec scolaire et redoublement ; c'est sur ce dernier concept que nous allons, dès à présent, nous pencher.

2.1.2 Le redoublement

Commençons par définir ce terme : Selon Gary-Bobo et Robin (2012, p.98) le redoublement correspond à « [...] recommencer une année scolaire entière, en répétant les mêmes enseignements, lorsque son niveau est jugé insuffisant pour passer en classe supérieure. »

De plus, d'après Galand et al. (2019) le redoublement peut être vu comme une mesure permettant de venir en aide aux élèves en difficulté, et c'est le cas notamment en FW-B, où le redoublement est vu comme une solution par défaut pour faire face aux difficultés d'apprentissage. Toutefois, comparée aux autres dispositifs de soutien, cette mesure s'avère être drastique dans ses coûts financiers, dans sa durée (une année scolaire minimum) et également dans son impact sur le parcours scolaire de l'élève étant donné, a minima, la séparation que cela engendre avec ses camarades. En FW-B, et de manière traditionnelle, le redoublement propose aux élèves de revivre une année scolaire dans la même classe que celle de l'année antérieure en suivant le même curriculum (Galand et al., 2019).

Alors que les données internationales PISA démontrent que les pays avec promotion automatique ont de meilleures performances que les pays pratiquant le redoublement (Crahay, 2004), les indicateurs de l'enseignement en FW-B de 2020 (p.38), paradoxalement, mettent en évidence le recours massif au redoublement dans le système éducatif belge. Plus précisément, en FW-B, malgré les restrictions imposées par les différents décrets mis en place, par exemple le décret mission (1997) qui avait notamment pour objectif de réduire progressivement les pratiques du redoublement, ce dernier est tout de même présent en primaire, et dans une large mesure en secondaire (Baye & Dacht, 2020). À titre d'exemple, les résultats de PISA 2015 indiquent, qu'en FW-B, à peu près 50% des élèves de 15 ans ont au moins connu une fois le redoublement, taux qui s'avère supérieur à celui des autres pays de l'OCDE (Lafontaine et al., 2019). Plus particulièrement, au niveau des secondaires, le taux de retard scolaire augmente de manière croissante au fil des années : une première augmentation est constatée entre la 2^e et 3^e secondaire, suivie d'une deuxième en 5^e année secondaire (Goffin et al., 2013). En 2015-2016, dans le secondaire ordinaire (toutes sections confondues), le taux de retard s'élève à 29% en 1^{ère} année, à 36% en 2^e année, à 50% 3^e année, et enfin à 61% en 5^e année secondaire (Baye, 2019). Dès lors, en FW-B, finir ses secondaires en ayant connu le redoublement fait partie de la norme.

Par ailleurs, les enquêtes internationales réalisées sur le redoublement montrent, tant dans le primaire que dans le secondaire, des inégalités socioculturelles plus marquées dans les systèmes scolaires qui recourent davantage à la pratique du redoublement (Galand et al., 2019). Pour expliquer ce résultat, des analyses réalisées en profondeur sur les données PISA montrent, qu'à performance équivalente à ce test, le risque d'avoir connu le redoublement est plus important chez les élèves défavorisés. Ainsi, le redoublement amplifie les écarts de performances en fonction de l'origine sociale de l'élève, et donc accroît les inégalités sociales (Galand et al., 2019). Ces constats rendent d'autant plus cohérents les résultats des tests PISA qui montrent que la FW-B se classe toujours parmi les systèmes éducatifs où les inégalités sociales sont les plus marquées (Lafontaine et al., 2019).

Dans la décision du redoublement, on retrouve également une iniquité de genre : les résultats PISA 2015 montrent qu'il y a plus de probabilités de redoubler lorsque l'on est un garçon que lorsque l'on est une fille (Lafontaine et al., 2019). En FW-B, le redoublement est donc pratiqué de manière inéquitable entre les filles et les garçons (Baye, 2019). De manière plus récente, ces résultats rejoignent ceux développés dans les indicateurs de l'enseignement de la FW-B 2020, qui montrent notamment que le taux de redoublement généré en 2017-2018 est systématiquement plus élevé pour les garçons que pour les filles, que ce soit en 3^e année ou en 5^e année secondaire, et ce, quelle que soit la forme d'étude (Indicateurs de l'enseignement FWB, 2020, p.38).

2.1.3 Les effets du redoublement – la stéréotypisation

L'intérêt pour les effets du redoublement n'est pas récent, de nombreuses études s'y intéressent depuis près d'un siècle, sans pour autant que la problématique soit considérée comme définitivement résolue (Crahay, 2004).

Tout d'abord, Holmes et Matthews (1984) dans leur méta-analyse, mettent en lumière les effets négatifs que peut avoir le redoublement, notamment sur le concept de soi des élèves ayant été retenus. La méta-analyse de Jimerson (2001), reprenant 20 études sur 400 concernant le redoublement entre 1990 et 1999, démontre également les effets négatifs du redoublement, notamment sur la réussite scolaire des élèves retenus ainsi que sur leur adaptation socio-émotionnelle.

Ensuite, Crahay (2019) mentionne les effets délétères que peuvent avoir les échecs antérieurs sur les performances futures des élèves. Les étudiants qui ravivent des souvenirs d'échec avant de réaliser une tâche obtiennent des résultats plus faibles à l'épreuve, comparés aux étudiants qui évoquent des souvenirs de réussite. Cette diminution de performance s'explique par la crainte qu'ils ont d'échouer (Selimbegovic et al., 2011, cités par Crahay, 2019).

L'étude réalisée par Grisay (1993) soutient également cette idée d'effet négatif lié au redoublement en mettant en évidence que les élèves faibles, qui en font l'objet, progressent moins que les élèves ayant des difficultés comparables, mais ayant été promus.

De plus, la méta-analyse de Allen et ses collègues (2009), reprenant 22 études sur 199, met en évidence que, plus le nombre d'années après la rétention augmente, plus les effets de la rétention deviennent négatifs. Cet effet négatif du nombre d'années après la rétention est d'autant plus fort pour les études utilisant des comparaisons au niveau scolaire (grade constant), que pour celles utilisant des comparaisons d'âge (âge constant). Ainsi, les élèves retenus bénéficieraient, par rapport à leurs camarades de classe plus jeunes, d'une amélioration à court terme de leurs résultats scolaires, et ensuite cet avantage se perdrait au fur et à mesure des années.

Si on compare, cette fois-ci, le niveau cognitif atteint par les redoublants et les non-redoublants à degré scolaire constant¹, les performances sont équivalentes et ce, peu importe après combien d'années (1 an, 2 ans...après le redoublement) est effectuée cette comparaison (Crahay, 2019).

¹ Comparaison a degré scolaire constant : élèves redoublants vs non-redoublants au terme de la même année scolaire

Ainsi, malgré qu'ils aient sacrifié une année, les élèves redoublants n'ont pas de meilleures performances que celles de leurs condisciples promus. Cette absence d'effets permet d'illustrer la faible efficacité du redoublement dans sa prétention à pouvoir aider les élèves en difficulté (Crahay, 2004).

Par ailleurs, contrairement à une opinion largement répandue chez les enseignants, selon Crahay (2004) le redoublement précoce² n'a pas d'effet préventif. Mais, à long terme, il peut entraîner des conséquences dramatiques et conduire l'élève au décrochage scolaire, c'est-à-dire, à se désintéresser progressivement de l'école suite à l'accumulation de facteurs internes et externes au système scolaire (Lambillotte & Leclerf, 1996, cités par Barthélémy, 2020). Ce constat est d'autant plus flagrant que le redoublement s'est effectué de manière **précoce** (Crahay, 2004).

Ces résultats rejoignent, d'une part, ceux de Witte, Nicaise, Lavrijsen, Van Landeghem, Lamote et Van Damme (2013, cités par Galand et al., 2019) qui déclarent que le taux de redoublement est associé de manière systématique au décrochage scolaire, de sorte que, le redoublement accroît les risques de décrochage scolaire. D'autre part, ces résultats se retrouvent dans la méta-analyse réalisée par Jimerson et ses collègues (2002), qui, au travers d'un examen systématique de 17 études sur le décrochage, mettent en évidence que le redoublement est l'un des facteurs prédictifs le plus puissant du décrochage scolaire ultérieur. Il ajoute à cela l'idée selon laquelle la probabilité de décrochage scolaire s'avère d'autant plus importante que les élèves ont été retenus **plusieurs fois**. Toutefois, malgré les études ayant attesté de cette association entre redoublement et décrochage scolaire, certains enseignants n'ont pas connaissance de l'existence d'un tel lien. Ainsi, cette sensibilisation permettrait d'influencer les stratégies d'intervention des professionnels de l'éducation, en dépassant l'utilisation de la rétention pour mettre en place des stratégies dont l'efficacité a été démontrée (Jimerson et al., 2002).

Aussi, il apparaît que les élèves redoublants sont perçus négativement par leurs pairs et les enseignants (Crahay, 2004). En effet, les non-redoublants associent une série de caractéristiques négatives au groupe des enfants redoublants telles que : distraits, malpolis, désobéissants, mauvais, fainéants, bagarreurs, paresseux etc. Ils attribuent aux enfants redoublants plus de caractéristiques négatives que de caractéristiques positives (Crahay, 2019).

² Redoublement précoce : redoublement avant d'entrer à l'école primaire ou en 1^{er} année primaire

On parlera ici de **stéréotypes** envers les élèves redoublants, c'est-à-dire des «[...] croyances des personnes concernant les attributs, les traits de personnalité typiques qui définissent un groupe » (Yzerbyt, 2016, p.90). De plus, dans cette théorisation du concept de stéréotype, Crahay (2019) distingue le concept d'endogroupe, groupe d'appartenance d'un individu, du concept d'exogroupe, tous les autres groupes auxquels l'individu ne s'identifie pas et dont il ne se sent pas faire partie. Néanmoins, il est possible que les membres de l'endogroupe adhèrent également aux stéréotypes associés à leur groupe. Dans le cas des enfants redoublants, ils pourraient attribuer aux membres de leur groupe, le manque d'intelligence, la paresse, etc (Crahay, 2019).

Toujours en lien avec les stéréotypes négatifs associés aux redoublants, Crisafulli et ses collègues (2002, cités par Crahay, 2019) ont montré que nombreux étaient les enfants qui associaient à la définition de « enfant redoublant » des réponses faisant référence au « manque d'efforts » et aux « comportements en classe ». Ils tentent d'expliquer le redoublement au travers d'une série de comportements caractéristiques relevés en rapport avec les mauvais comportements en classe, le manque d'efforts et, dans une moindre mesure, mentionnent « un manque d'intelligence » (Crahay, 2019).

Toutefois, Jussim et al. (2015), évoquent l'idée que le contenu des stéréotypes n'est pas toujours faux, mais qu'il peut parfois refléter les caractéristiques réelles d'un groupe. Par exemple, pour décrire l'enfant redoublant, le sexe masculin est davantage choisi que le sexe féminin, et en réalité, il s'avère que les garçons redoublent plus que les filles (Crahay, 2019).

Ainsi, et compte tenu de toutes ces caractéristiques, Crahay (2019) met en évidence que, lorsqu'ils apprennent leur redoublement, les enfants concernés ressentent souvent de la tristesse, la sensation d'être punis, la peur d'être la cible de moqueries et de rejet de la part de leurs condisciples. De plus, ils associent leur redoublement à des causes internes : s'ils ont redoublé, c'est leur faute, notamment à cause de leur manque d'intelligence. Penser de la sorte peut engendrer un sentiment d'incompétence et d'impuissance face à la tâche à réaliser (Crahay, 2019).

Par ailleurs, le stéréotype associé au redoublement, sera présent dans le milieu scolaire mais généralement de manière implicite. En effet, souvent, l'équipe pédagogique ne parlant pas à l'enfant de son redoublement, et l'enfant n'en parlant pas avec ses condisciples, il n'est pas toujours évident pour un élève redoublant d'identifier les raisons de son redoublement (Crahay, 2019).

Cependant, Crahay (2019) nuance les résultats précédents en déclarant que le redoublement n'aurait **pas un effet négatif direct** sur l'élève. L'influence négative sur ces certaines dimensions³ serait due en partie à la menace du stéréotype. Nous reviendrons sur ce point dans la partie relative aux « effets délétères des stéréotypes sur le redoublement ». En outre, dans sa méta-analyse, Allen et ses collègues (2009) soulèvent l'absence d'effet positif de la rétention sur la réussite scolaire et remet donc en question les avantages éducatifs des politiques de rétention. En effet, même les études qui fournissent des contrôles solides pour limiter les effets de sélection, ne parviennent pas à trouver des avantages au maintien dans l'année scolaire. Ce qui souligne l'importance de se concentrer sur les stratégies employées, lorsque les enfants ne parviennent pas à atteindre les compétences du niveau de l'année scolaire, afin de soutenir la réussite scolaire des élèves.

Dès lors les enseignants qui décident de faire redoubler, le font sans disposer de preuves valables attestant que cette pratique sera plus porteuse que la promotion dans la classe supérieure.

Avant de conclure ce point, il semble également important de mettre en évidence que les études n'ont pas seulement retenu des effets négatifs du redoublement, mais certaines ont permis de relever un effet positif du redoublement à un moment donné. Par exemple, et comme énoncé précédemment, la méta-analyse de Allen et ses collègues (2009) a permis de démontrer un effet positif du redoublement à court terme sur les résultats scolaires des élèves retenus, et de manière plus récente, l'étude réalisée par Goos et al. (2013) rejoint également ces résultats. Plus précisément, dans l'étude quasi-expérimentale réalisée par Goos et al. (2013), 4000 enfants belges ont été suivis de la 1^{ère} primaire jusqu'au début des secondaires, dont certains ont connu le redoublement après 1 an (298 élèves), afin de mettre en évidence les effets du redoublement. Pour ce faire, ils ont comparé les résultats scolaires des élèves promus à ceux des élèves redoublants. Les résultats mettent en évidence qu'à court terme, les élèves retenus obtiennent de meilleurs résultats en lecture et en mathématiques comparativement à leurs camarades promus. Toutefois, avec le temps, ces effets positifs s'estompent et c'est la tendance inverse qui s'observe ; dès la 2^e primaire, les élèves promus obtiennent des résultats supérieurs à ceux des élèves redoublants.

³ Ces dimensions concernent la performance, l'estime de soi, le sentiment de compétence et la motivation scolaire.

Pour conclure ce point, nous retiendrons que, dans le meilleur des cas, les élèves redoublants auront perdu au moins un an de leur scolarité pour atteindre un niveau semblable aux élèves promus et qui avaient des difficultés comparables. Dans le pire des cas, ils auront des résultats scolaires d'un niveau inférieur, un risque de décrochage scolaire plus important et davantage de problèmes psychosociaux (Galand et al., 2019). On peut donc réfuter l'idée selon laquelle le redoublement agirait comme une remédiation permettant aux élèves de surmonter leurs lacunes et de repartir dans le cursus scolaire sur de meilleures bases (Crahay, 2004). Nous terminerons cette partie par cette phrase: « [...]le redoublement est une fausse solution à un vrai problème. » (Galand et al., 2019, p.21).

2.1.4 Pourquoi continue-t-on de faire redoubler malgré les effets négatifs collatéraux?

Malgré tous ces résultats qui attestent de la faible efficacité du redoublement et de son caractère préjudiciable à plusieurs égards, des enseignants de nombreux pays continuent de recourir à cette pratique, comme si c'était le seul moyen pour remédier aux difficultés scolaires des élèves. C'est notamment le cas en Belgique, où le redoublement est considéré comme normal, faisant partie intégrante du mode de fonctionnement du système scolaire (Boraita, 2015 ; Goffin et al., 2013). Dès lors nous sommes encline à nous poser la question suivante : pourquoi continue-t-on de faire redoubler malgré la connaissance des effets négatifs de cette pratique?

Le rapport d'Eurydice (2011, cité par Goffin et al., 2013) déclare que, dans les pays où le redoublement est pratiqué de manière récurrente, il est vu comme étant positif pour les apprentissages et ce essentiellement en Belgique, Espagne, France, Luxembourg, Pays-Bas et Portugal. Mais à quoi cette vision positive est-elle associée ?

Décider du redoublement d'un élève relève d'une certaine complexité (Marcoux & Crahay, 2008). En effet, c'est l'interaction d'un faisceau de facteurs qui conduit les enseignants à considérer cette mesure comme la plus opportune.

Premièrement, dans ce processus décisionnel intervient la note scolaire de l'élève. Toutefois, les notes n'interviennent pas de façon unique et pas toujours de façon prioritaire : elles permettent de soutenir un jugement en voie de formation. Dès lors, en étant convaincus qu'un élève qui ne maîtrise pas les bases n'est plus en mesure de construire des connaissances et compétences solides, les enseignants pensent agir dans son intérêt en le faisant redoubler (Marcoux & Crahay, 2008).

Deuxièmement, le jugement des enseignants s'appuie essentiellement sur des croyances conditionnelles quant aux bienfaits du redoublement. Ils n'évoquent pratiquement jamais dans leurs explications leur connaissance des recherches sur le redoublement. C'est en fonction de ces croyances que les enseignants décèlent l'évolution des redoublants et orientent leurs questionnements et réflexions quant aux élèves en difficulté. Plus précisément, ce jugement porte davantage sur des dimensions psychologiques comme la maturité de l'élève, son attitude générale, plutôt que sur des dimensions pédagogiques telles que la maîtrise de compétences (Marcoux & Crahay, 2008).

Effectivement, les enseignants auraient plus tendance à recourir au redoublement s'ils ont une conception maturationaliste du développement. Ils penseraient que faire répéter une année, fournirait à l'élève le temps nécessaire pour développer une maturité suffisante pour aborder, avec de meilleures capacités cognitives, les apprentissages et les causes de ses difficultés ou de blocages (Marcoux & Crahay, 2008).

Cette conception maturationaliste se retrouve également dans une étude de Byrnes (1990, citée par Boraita, 2015), qui souligne qu'un des éléments décisionnels les plus importants est le manque de maturité. Il ajoute également à cela l'idée de basse motivation et de faible estime de soi.

Troisièmement, et comme le met en évidence Draelants (2006), le redoublement répond notamment à une rationalité pragmatique. Il permet d'aboutir à une reconfiguration de classe mieux adaptée au rythme des progrès de la majorité en évitant que les progrès des « meilleurs » soient ralentis par les élèves en difficulté. Le redoublement permettrait donc de gérer l'hétérogénéité, de faire un tri des élèves dans l'établissement, de se positionner de manière stratégique et symbolique par rapport aux établissements environnants, et de maintenir l'autonomie professionnelle des enseignants (Goffin et al., 2013). Galand et ses collègues (2019) rejoignent ces constats en mettant en évidence que le redoublement plus fréquent permettrait une homogénéisation des élèves à l'intérieur des écoles, et que l'enseignement serait d'une meilleure efficacité avec des élèves plus semblables ou avec des classes plus homogènes.

Quatrièmement, et comme le montre Boraita (2015), la croyance en faveur du redoublement est liée à la non-connaissance des recherches sur le redoublement, aux conceptions de l'évaluation normative, à la justice méritocratique, à la vision d'une intelligence innée et de l'apprentissage par répétition et par déclic.

Malheureusement, nombreux sont les futurs enseignants qui n'ont pas vraiment connaissance des recherches sur le redoublement et qui sont favorables à sa pratique (Boraita, 2015). Ces résultats sont en cohérence avec ceux rapportés dans l'étude de Goffin et ses collègues (2013), au sein de laquelle il a été soumis à 196 futurs enseignants en Belgique francophone, avant et après leur formation, un questionnaire portant sur leur sentiment d'auto-efficacité, leur conception de l'intelligence ainsi que sur leurs croyances relatives au redoublement. En début de formation, la croyance quant aux bienfaits du redoublement s'explique en partie par un sentiment d'efficacité générale et par les croyances relatives au développement de l'intelligence, vision de l'intelligence comme innée. En fin de formation, après avoir suivi des cours portant sur le redoublement, on constate notamment une évolution des croyances des enseignants sur le redoublement. Ils remettent en cause les bienfaits du redoublement et semblent mieux percevoir les effets négatifs reliés à celui-ci .

Cinquièmement, les enseignants ont des jugements normatifs envers leurs élèves ou leur famille. Ils regardent les élèves au travers d'une vision idéalisée de ces derniers (ex : motivés, actifs, enthousiastes, participatifs, etc.) et les environnements familiaux en fonction d'une représentation normative de la famille idéale (ex : des parents qui peuvent apporter un accompagnement éducatif, qui ne sont pas divorcés, qui parlent la langue d'enseignement à la maison, etc.). Cette double perception peut déboucher sur des raisonnements de compensation, ou à l'opposé, de cumulation. Par exemple, des enseignants vont décider de promouvoir un élève qui présente plusieurs difficultés, mais qui a une famille avec un potentiel de ressources éducatives importantes (Hutmacher, 1993).

Les enseignants se retrouvent également face à des dilemmes moraux lorsque mérites et talent ne concordent pas. C'est le cas, par exemple, lorsqu'ils se retrouvent face à un élève ayant des difficultés mais qui est méritant. Toutes ces incertitudes, ces dilemmes exprimés par certains enseignants, soutiennent l'idée selon laquelle la croyance en les bienfaits du redoublement n'est pas absolue (Marcoux & Crahay, 2008).

Par contre si le redoublement n'a pas les effets attendus, son inefficacité est soit refoulée, soit expliquée par des causes externes (ex: l'élève n'a pas fait les efforts nécessaires s'il n'a pas reçu de soutien de la part de ses parents, etc.).

Parfois même, les enseignants soutiennent l'idée selon laquelle "[...] revoir une seconde fois la matière ne peut faire de tort" (Marcoux & Crahay, 2008, p.513).

Dès lors, compte tenu de tous ces éléments, nous sommes encline à nous demander s'il serait possible de modifier les croyances des enseignants quant à l'efficacité du redoublement.

A cette question, et comme énoncé précédemment, Boraita (2013, 2015) fournit une réponse favorable. En effet, deux de ses études, l'une réalisée en 2013 et l'autre en 2015, démontrent l'influence positive que peut avoir, sur les croyances des futurs enseignants, le fait de suivre une formation théorique et pratique sur le redoublement. Cela permet de faire évoluer, changer et/ou évoluer les croyances sur le redoublement de sorte que, « plus les étudiants ont connaissance des recherches sur le redoublement, plus ils ont tendance à être contre la pratique. » (Boraita, 2013, p.364). Goffin et ses collègues (2013) ajoutent à cette idée que les cours théoriques affectent beaucoup moins les croyances des étudiants que les expériences pratiques et les stages. Cette évolution pourrait donc s'expliquer davantage par les cours pratiques qu'ont suivis ces élèves plutôt que par les cours théoriques. Plus particulièrement, Boraita (2013) met en évidence que l'évolution des croyances des futurs enseignants relatives au redoublement est en relation avec les croyances concernant l'évaluation et la justice. Ainsi, on retrouve une adhésion à l'égalité des acquis et aux évaluations formatives chez les futurs enseignants qui ne sont pas pour le redoublement.

Ces résultats s'avèrent essentiels pour la formation initiale ou continue des enseignants, puisqu'ils témoignent de la nécessité et de l'importance des savoirs théoriques (Boraita, 2015). Toutefois, ils imposent la prudence, car on ne peut garantir que ces changements et évolutions observés sont des transformations durables des conceptions des enseignants (Boraita, 2013). L'entrée en fonction de ces futurs enseignants peut bouleverser leurs convictions développées suite à leur formation initiale (Hausoul, 2005, cité par Boraita, 2013). Effectivement, Marcoux et Crahay (2008, p.512) déclarent que la croyance en les bienfaits du redoublement « [...] est aussi une représentation sociale : se nourrissant du fait qu'une majorité d'enseignants la partage, au moins à des degrés divers, elle se perpétue, voire se renforce. »

En outre, la pratique du redoublement ne dépend pas uniquement des croyances des enseignants. Même si un changement au niveau des croyances est plus facile à mettre en exécution qu'un changement au niveau des structures du système scolaire, changer les croyances des enseignants sur le redoublement n'enrayera pas automatiquement ces fonctions latentes qui rattachent les enseignants à la pratique (Boraira, 2013). Dès lors, « [...] un changement des croyances devrait plutôt accompagner et influencer le changement des pratiques, et par-delà celui du système. » (Boraira, 2013, p.366).

On retrouve également cette idée chez Goffin et ses collègues (2013) qui envisagent le changement au niveau des conceptions comme un préalable à un changement de pratiques, même si, parfois, ces changements peuvent s'accompagner, malgré tout, d'une conservation de l'attachement à la pratique du redoublement.

2.2 Les effets délétères des stéréotypes

Les effets délétères des stéréotypes sont nombreux, mais dans le cadre de ce mémoire nous nous focaliserons sur ceux qui nous intéressent, à savoir : le phénomène de confirmation des attentes (prophétie autoréalisatrice, effet pygmalion), la stigmatisation, la menace du stéréotype et l'effet solo.

2.2.1 Le phénomène de confirmation des attentes

Toute personne ayant besoin de comprendre, évaluer, expliquer et prédire les comportements des autres, va interagir avec autrui en ayant certaines **attentes** concernant la manière dont ils vont se comporter. Néanmoins, toutes ces croyances, impressions, perceptions ou attentes, que nous pouvons avoir envers une personne, peuvent influencer involontairement nos pensées et comportements envers cette dernière, et en retour influencer les pensées et comportements de celle-ci (Snyder & Stukas, 1999). Toutefois, il est prouvé que la prise de conscience de l'existence d'un biais implicite peut en limiter les effets (Rudman et al., 2001).

Le phénomène le plus largement étudié est celui que Robert Merton a appelé dès 1948 « *self-fulfilling prophecy* », c'est-à-dire **prophétie autoréalisatrice** (Trouilloud & Sarrazin, 2003). Selon Merton (1948, p.195), « Une prophétie autoréalisatrice est une définition d'abord erronée d'une situation qui suscite un nouveau comportement qui rend *exacte* cette conception initialement fausse. » [Notre traduction]. On parlera de prophétie autoréalisatrice lorsqu'une croyance ou attente d'un enseignant à l'égard d'un élève influence l'attitude du premier à l'égard du second de manière à ce que l'élève se conforme à la croyance de l'enseignant, voire la renforce (Trouilloud & Sarrazin, 2003).

Il faudra attendre la sortie du livre *Pygmalion à l'école* de Rosenthal et Jacobson (1968) pour que « les attentes des enseignants » fassent véritablement leur apparition dans la sphère éducative (Trouilloud & Sarrazin, 2003). Plus précisément, en 1965, Rosenthal et Jacobson (cités par Rosenthal & Jacobson, 1968) ont réalisé une expérience dans une école élémentaire publique, au cours de laquelle ils ont fait croire à des enseignants que certains élèves de leur classe (les élèves « promus ») avaient une forte probabilité de faire des progrès significatifs durant l'année.

En réalité, les élèves avaient été sélectionnés de manière aléatoire et ne présentaient pas de différence particulière avec les autres élèves de la classe.

Toutefois, d'après les auteurs de cette étude, les enseignants se comportaient de manière différente avec les élèves pour lesquels ils attendaient de meilleurs résultats. Ils les traitaient de manière « plus agréable, plus amicale et plus enthousiaste » (Lecompte, 1997, p.38) et leur offraient plus d'opportunités de répondre. Effectivement, ils leur laissaient davantage de temps pour organiser leurs réponses, pour travailler en autonomie et leur posaient plus souvent des questions. A l'inverse, ils interrogeaient moins souvent les élèves non promus et, quand ils le faisaient, leur posaient des questions faciles et avaient tendance à leur donner plus rapidement la réponse. Toutefois, cet effet n'apparaît pas systématiquement, son importance varie selon les enseignants et les élèves (Lecomte, 2019).

Traités de la sorte, ces élèves présentaient une augmentation significative de leur quotient intellectuel (QI). Cette augmentation peut s'expliquer par le fait que, encouragés par l'attention dont ils faisaient l'objet, ils ont développé une meilleure estime d'eux-mêmes et se sont mis à travailler davantage (Lecompte, 1997). En d'autres termes, en progressant au niveau du QI, les espérances placées en eux se sont vues confirmées. La prophétie ainsi accomplie a été nommée par les auteurs « **effet Pygmalion** » (Trouilloud & Sarrazin, 2003).

Cette expérience de Rosenthal et Jacobson (1968, cités par Crespo, 1988) permet donc d'identifier l'influence que peuvent avoir les évaluations positives des enseignants, sur les capacités intellectuelles et les rendements de leurs élèves. Toutefois, l'impact des perceptions n'est pas unidirectionnel, les élèves semblent également exercer une influence sur les attitudes et comportements des enseignants (Crespo, 1988).

Par ailleurs, les attentes des enseignants peuvent également **affecter la façon dont ils perçoivent et évaluent les performances des élèves** (Sprietsma, 2013). Les notes, principaux indicateurs des capacités et performances de l'élève pour l'enseignant, peuvent entraîner des conséquences, à long terme, sur les résultats de l'élève et donc, sur son avenir. Cependant, les enseignants ont tendance à utiliser les notes à d'autres fins que mesurer la performance de l'élève dans une tâche donnée (ex : punir/récompenser le comportement en classe, etc.), bien que celles-ci ne permettent pas toujours une mesure objective des performances.

Quand bien même l'enseignant utiliserait un ensemble de critères d'évaluation spécifiques, de multiples impressions et attentes subjectives pourraient venir influencer l'estimation de la qualité du travail des élèves, et donc biaiser son jugement évaluatif (Sprietsma, 2013). Nous reviendrons sur les biais de l'évaluation plus loin dans ce travail.

Pour résumer ce point, même si ce n'est pas systématique, il existe, à l'heure actuelle, des preuves importantes pouvant attester de l'existence d'un traitement différentiel des enseignants, en concordance avec leurs attentes à l'égard de leurs élèves (Trouilloud & Sarrazin, 2003).

2.2.2 La stigmatisation

Les enfants redoublants sont influencés par la mauvaise réputation qui pèse sur leur groupe de sorte qu'ils se sentent stigmatisés. Un stigmatisme désigne la « 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 a aura "eux", qui ont une mauvaise réputation, et "nous", les normaux. » (Croizet & Leyens, 2003, cités par Crahay, 2019, p.233). La conscience du stigmatisme s'opère très tôt, dès le début des primaires (vers 6 ans) (Cvencek et al., 2011).

Se sentir stigmatisé entraîne diverses conséquences (ex : de l'anxiété), qui, souvent, sont négatives, perturbatrices et déstabilisantes, et peuvent entraîner une baisse d'estime de soi, un désinvestissement dans les domaines concernés, du désespoir ou encore de la colère (Crahay 2019).

Plus particulièrement, les enfants qui sont porteurs du stéréotype, à force d'y être confrontés, peuvent l'internaliser, de sorte qu'il finirait par faire partie intégrante de leur identité. Dans ce cas, petit à petit, les individus stigmatisés finiraient par adhérer au stéréotype qui porte sur eux, orienteraient leurs attitudes et comportements en classe et ailleurs en fonction de ce dernier, et se construiraient progressivement une image négative de leur identité scolaire (Crahay, 2019).

2.2.3 La menace du stéréotype et l'effet solo

D'après Crahay (2019) la menace du stéréotype est un phénomène situationnel au cours duquel un individu appartenant à un groupe stigmatisé, va avoir peur de confirmer le stéréotype qui porte sur son groupe. En retour de cette crainte de confirmer le stéréotype, on peut constater une baisse de performance chez l'individu stigmatisé.

En effet, les études réalisées sur ce sujet ont permis de mettre en évidence que, dans une situation menaçante, les sujets obtenaient de moins bons résultats dans les épreuves ciblées par le stéréotype. L'influence négative du stéréotype sur les performances de l'élève sera d'autant plus grande que l'enfant percevra l'épreuve à réaliser comme évaluant de manière précise le secteur dans lequel il est moins bon. Cette référence à un phénomène situationnel provient du fait que, dans la menace du stéréotype, l'environnement est d'une grande importance puisque c'est sur base de la situation, plus précisément sur base des indices de la situation, que l'individu va se rappeler le stéréotype qui le concerne (Crahay, 2019). Ces indices peuvent être plus ou moins flagrants, mais d'après certains auteurs, il s'avère que plus ils sont subtils plus ils auraient un effet délétère comme déclencher des émotions et des pensées négatives (Dardenne et al., 2007 ; Freesson et al., 2018 ; Wheeler & Petty, 2001, cités par Crahay, 2019). Cela pourrait s'expliquer par le fait que, en situation ambiguë (indices subtils), il serait plus difficile pour les personnes d'identifier l'origine de leur inconfort et de leurs émotions négatives. Alors que, quand les stéréotypes sont présentés de manière explicite, il est plus facile de les rejeter (Crahay, 2019).

Dans le cas du redoublement, les enfants redoublants seront fréquemment confrontés au stéréotype qui concerne leur groupe d'appartenance. Plus précisément, dès qu'ils seront confrontés à une situation où leurs compétences scolaires seront amenées à être évaluées, ils craindront de confirmer le stéréotype qui porte sur eux. Cette crainte à son tour, par l'intermédiaire de différents mécanismes (ex : augmentation des émotions négatives, diminution de la mémoire de travail, etc.), entrainera une diminution de leurs résultats scolaires (Shmader et al., 2008). D'après le modèle de Shmader et al. (2008), les pensées et émotions négatives qu'entraînerait la menace du stéréotype, amèneraient les individus qui les subissent à tenter de les supprimer ou de les réguler. Toutefois, les stratégies régulatrices nécessitant beaucoup de ressources cognitives, finiraient par les épuiser et les rendre indisponibles pour la réalisation de la tâche en cours. Ainsi, les performances des élèves dans l'activité s'en verraient réduites. La confrontation aux stéréotypes négatifs peut entraîner, à long terme, des conséquences au niveau des émotions, de la cognition et de la réalisation des individus stigmatisés (Crocker & Major, 1989 ; Croizet & Martinot, 2003). En effet, à force d'être confrontés à ces sentiments de menace, les individus vont, progressivement, se désengager et se désinvestir du domaine dans lequel ils sont impliqués. La désidentification permet à l'élève de préserver son image de soi en abandonnant progressivement le domaine scolaire. Elle correspond donc à un mécanisme de protection du self. (Crahay, 2019).

Pour terminer ce point, il nous semble également intéressant de souligner ce que les études sur les stéréotypes ont mis en évidence concernant **l'effet solo**, c'est-à-dire le fait d'être le seul représentant de son groupe dans une situation sociale donnée. Elles ont montré qu'être l'unique représentant de son groupe avait tendance à exacerber les effets de la menace du stéréotype (Crahay, 2019). Les élèves redoublants, se trouvant en minorité dans certains groupes classes et particulièrement en primaire, subiraient de manière plus conséquente les effets délétères liés à la menace du stéréotype et auraient d'autant plus peur de le confirmer. Un redoublant fait donc partie d'une minorité sociale qui est stéréotypée de manière négative dans un groupe classe (Crahay, 2019).

Toutefois, malgré tous ces effets délétères, les stéréotypes sont inévitables et relèvent d'une certaine utilité, et c'est ce que nous allons mettre en évidence dans le point suivant.

2.2.4 Le caractère inévitable des stéréotypes

Leyens et Scaillet (2012), relatant l'expérience réalisée par Patricia Devine (1989) démontrent le caractère inévitable et spontané des stéréotypes. Dans cette expérience, il a d'abord été demandé à des étudiants américains blancs d'énoncer librement une série d'informations se rapportant aux stéréotypes culturels concernant les Noirs-Américains, à savoir que la majorité d'entre eux associe l'agressivité aux Noirs-Américains. Sur base de toutes ces croyances et traits de personnalité négatifs et positifs au sujet de ce groupe, Devine a pu distinguer les sujets racistes des non racistes.

Ensuite, Devine a confronté d'autres sujets (racistes et non racistes) à une exposition subliminale (80 ms) de 100 mots dans deux conditions : 80 mots stéréotypes envers les noirs (ex : crime, ghetto, etc.) 20 mots neutres pour une moitié des participants, et pour l'autre moitié 20 mots stéréotypes et 80 mots neutres. La vitesse de présentation de ces mots était telle qu'ils ne pouvaient avoir conscience des mots qu'ils voyaient. Dans ce cas, le stéréotype était donc activé sans qu'ils ne s'en rendent compte.

Puis il a été présenté à tous les participants un personnage fictif dénommé Donald. Suite à cette présentation, un lien entre le nombre d'amorces présentées et l'évaluation donnée à Donald a été établi. De sorte que, plus les sujets ont été amorcés par des mots stéréotypés, plus ils percevaient Donald comme agressif ; résultat d'autant plus vrai pour les sujets racistes que pour les non-racistes. Par conséquent, il n'est pas indispensable d'être en accord avec le stéréotype, mais le connaître est suffisant pour que celui-ci soit « utilisé » sans qu'on ait conscience de son activation.

Ainsi, « Les stéréotypes sont là pour définir des catégories et le tout a pour but de nous guider dans l'action. L'internalisation des stéréotypes peut faire en sorte que leur simple confrontation suffise à nous faire agir d'une certaine façon.» (Leyens & Scaillet, 2012, p.111).

Compte tenu de tous ces effets délétères liés aux stéréotypes, on est en position de se demander quel serait l'impact des stéréotypes liés au statut scolaire de l'élève (redoublant ou non) sur l'évaluation donnée par un enseignant. Ce qui nous amène aux deux points suivants, qui concernent le jugement évaluatif et les biais à l'évaluation.

2.3 L'évaluation

2.3.1 Le jugement professionnel et le jugement évaluatif

Comme le met en évidence Leveault (2008), le jugement professionnel correspond à une démarche au terme de laquelle, sur base de l'expertise professionnelle (expériences et formation) une prise de décision est effectuée. Mottier Lopez et Allal (2008) ajoutent à cela que le jugement professionnel est à la fois un processus cognitif (de mise en relation de différentes informations, de raisonnement, d'interprétation, d'appréciation, de prise de décision) et une pratique sociale qui repose sur des normes, des langages, des outils, des cultures propres aux contextes sociaux dans lesquels le jugement de l'enseignant se situe.

Dans le domaine des évaluations des apprentissages, la notion de jugement est omniprésente si bien que les concepts « évaluer » et « juger » sont souvent considérés comme étant des synonymes (Mottier Lopez & Tessaro, 2016). Le jugement professionnel est donc au centre des pratiques d'enseignement et relève d'un enjeu capital en termes d'évaluation formative et certificative des apprentissages des élèves (Mottier Lopez & Tessaro, 2016). Il permet de se centrer sur les compétences professionnelles essentielles à la mise en place d'une évaluation pédagogique de qualité, inscrite dans un cadre éthique et déontologique. Plus particulièrement, et comme le mettent en évidence Mottier Lopez et Tessaro (2016), dans un contexte d'évaluation des apprentissages, pour désigner le jugement professionnel on parlera plutôt de **jugement évaluatif**.

2.3.2 Le jugement subjectif

Concernant le concept d'évaluation, celui-ci peut être défini de plusieurs manières mais la définition qui est la plus couramment utilisée est celle proposée par Madaus, Scriven et Stufflebeam (1986, cités par Romainville, 2011).

D'après ces auteurs,

[...]évaluer consiste à mesurer puis à apprécier, à l'aide de critères, l'atteinte des objectifs d'enseignement, en trois étapes. La première étape réside dans le recueil systématique, valide et fidèle d'informations appropriées aux objectifs d'enseignement. C'est la phase d'observation et/ou de recueil de données. La deuxième étape consiste à interpréter les informations recueillies à l'aide de critères. C'est la phase d'analyse. Celle-ci débouche sur la troisième étape, à savoir l'établissement de conclusions et/ou la prise de décisions. C'est la phase du jugement à proprement parler. (p.1)

L'évaluation des apprentissages est donc un aspect complexe de l'enseignement (Leroux & Bélair, 2015) et le modèle de Pellegrino, Chudowsky et Glaser (2001, cités par Berthiaume et al., 2011) permet de l'illustrer en divisant le processus d'évaluation des apprentissages en trois étapes : « [...]la clarification des apprentissages visés par l'évaluation, le choix de la méthode pour observer la " preuve " de ces apprentissages et le développement d'un mécanisme d'interprétation visant à analyser cette " preuve ".» (p.2) des apprentissages réalisés.

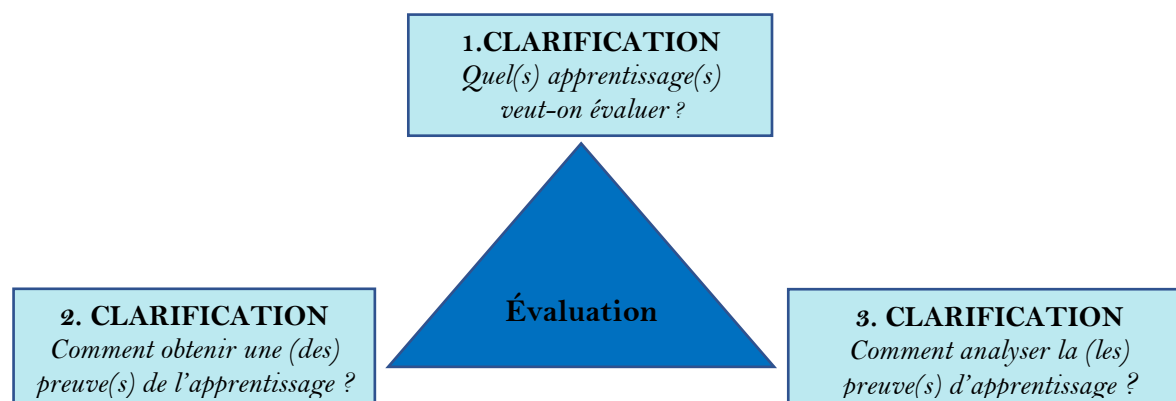


Figure 1. les étapes du processus d'évaluation des apprentissages (Pellegrino et al., 2001, cités par Berthiaume et al., 2011)

Tout processus d'évaluation se devrait de respecter les 5 principes suivants : la fidélité, l'équité, l'efficacité, la sensibilité et la validité (Mémo pédagogique, 2016). Néanmoins, malgré la minutie de l'évaluateur, le jugement de l'enseignant est influencé par des aspects inconscients, difficilement contrôlables (Leveault, 2005, cité par Mottier Lopez & Allal, 2008).

Nombreux sont les enseignants qui déclarent que la partie la plus difficile dans l'évaluation des apprentissages des élèves réside dans leur interprétation (Berthiaume et al., 2011). C'est d'ailleurs dans cette étape que la subjectivité, implicite dans le processus d'apprentissage, influence considérablement l'évaluation (Chirart et al., 2014). En effet, dans l'interprétation des preuves d'apprentissages, les enseignants peuvent être influencés par des facteurs ou des dimensions qui ne sont pas en lien avec les apprentissages visés par l'évaluation (Berthiaume et al., 2011), ce qu'appelle Romainville (2011, p.3) « des caractéristiques externes à la cible de l'évaluation ».

Il s'agit par exemple de caractéristiques propres aux étudiants telles que : leur aspect physique, leur tenue, leur prononciation ou leur accent, leur attitude face à l'école, etc., ou encore de caractéristiques relatives à la présentation du texte, comme l'écriture, la propreté ou l'orthographe. C'est ce qu'on appelle plus communément **l'effet halo** (Leclerq et al., 2004 ; Romainville, 2011).

Cette influence, qui engendre une subjectivité dans le processus d'évaluation, est inévitable puisqu'il s'agit d'êtres humains évaluant des phénomènes humains. En effet, lorsque nous observons des phénomènes humains, nous ne pouvons pas nous empêcher de les mettre en lien avec nos propres connaissances et/ou croyances (Berthiaume et al., 2011). Gérard (2002, p.26) va dans le sens de ces propos en déclarant que « [...] l'objectivité de l'évaluation est impossible, parce que la subjectivité est inévitablement présente dans tous les processus d'évaluation [...] de plus cette subjectivité est nécessaire pour que l'on puisse parler d'un processus d'évaluation. ». Tout comme on ne peut échapper à la subjectivité, il est difficile d'éviter les biais cognitifs qui se mettent en travers du processus d'évaluation (Chaumont & Leroux, 2018).

Avant de débiter cette section relative aux biais d'évaluation, il nous semble important de préciser que, lorsque nous utilisons l'appellation « biais d'évaluation », nous faisons référence au système de croyances et de valeurs, propre à chaque enseignant, qui vient exercer une influence sur leurs réflexions et leurs prises de décision dans le contexte évaluatif.

2.3.3. Les biais liés à l'évaluation

Au cours de plusieurs décennies, des études ont été réalisées sur les facteurs qui contribuent au manque d'objectivité dans l'évaluation des résultats scolaires (Klein & Taub, 2005).

La docimologie, cette discipline scientifique qui est dédiée à l'étude des examens scolaires, et plus particulièrement à la manière dont sont attribuées les notes par les correcteurs, permet de mettre en évidence un certain nombre de facteurs pouvant influencer notre jugement évaluatif (Detroz, 2021). Ces éléments susceptibles de biaiser l'évaluation peuvent être répartis selon quatre 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.

Dans le cadre de cette section, nous avons fait le choix de nous focaliser essentiellement sur les sources de biais qui nous intéressent, à savoir les trois dernières. Plus précisément, sont développés dans chacune de ces trois sources de biais, uniquement les éléments sur lesquels nous portons un certain intérêt dans le cadre de cette recherche.

Commençons par nous centrer sur les biais liés à la copie.

2.3.3.1 Les biais liés à la copie

Nombreuses sont les études qui attestent de la problématique liée à fiabilité du score (Charney, 1984 ; Odell & Cooper, 1980). Les biais en évaluation sont des problématiques qui persistent, et qui peuvent se présenter sous différentes formes (Craig, 2001). Dans les sources de biais d'évaluation liés à la copie, on retrouve notamment les biais liés à l'orthographe, la propreté, la lisibilité de la copie, etc. (Detroz, 2021). Dans le cadre de ce mémoire, nous avons choisi de développer uniquement les biais liés à la copie qui nous intéressent, autrement dit celui relatif à la relation entre l'écriture d'un élève (lisibilité) et la note reçue, ainsi que celui portant sur la comparaison entre la méthode de transcription relative à la production manuscrite et celle dactylographiée.

2.3.3.1.a Écriture manuscrite « soignée » ou « désordonnée » ?

Plusieurs études démontrent que l'écriture soignée est mieux notée que l'écriture désordonnée (Craig, 2001 ; Hughes et al., 1983 ; Markham, 1976 ; Powers et al., 1994, etc.). L'étude plus récente de Craig (2001) s'est intéressée à la question de la qualité de l'écriture manuscrite et du traitement de texte en tant que facteurs de biais dans les tests d'anglais langue seconde. Pour ce faire, 4 évaluateurs (experts en tests de placement anglais langue seconde) ont évalué 40 rédactions : 20 originales et 20 recopiées (de manière soignée ou désordonnée) ou dactylographiées. Les résultats attestent que les transcriptions réalisées par traitement de texte ont été moins bien évaluées que leurs homologues manuscrites et ce, indépendamment de la lisibilité de l'originale. Cette étude conclut donc que la simple transcription d'une dissertation manuscrite en version informatisée peut avoir un effet sur la note.

Seule l'étude de Brown (in press) atteste d'un effet contraire, à savoir que l'écriture désordonnée est mieux notée que l'écriture soignée. Dans cette étude expérimentale, 4 juges (des enseignants de l'écriture dans le contexte de 1^{ère} ou 2^e langue) ont été amenés à porter un jugement sur la lisibilité de copies manuscrites en langue seconde en se positionnant sur une échelle allant de 1 à 6. Les résultats, contre toute attente, ont mis en évidence qu'une mauvaise lisibilité entraîne des notes plus élevées. Autrement dit, dans cette étude, la "mauvaise" écriture s'est donc avérée un avantage dans la notation.

2.3.3.1.b Texte manuscrit ou texte dactylographié ?

Compte tenu de ce constat, on pourrait s'attendre à ce que les dissertations rédigées par traitement de texte (sur ordinateur) soient mieux notées que les dissertations manuscrites.

Toutefois, la littérature sur le sujet ne va pas dans ce sens. En effet, la majorité des études menées sur le sujet mettent l'accent sur le fait que les dissertations manuscrites sont mieux notées que celles rédigées par un traitement de texte (Arnold et al., 1990 ; Bridgeman & Cooper, 1998 ; Brown, in press ; Powers et al., 1994 ; Peterson et Lou, 1991, cités par Craig, 2001). Si on prend l'exemple de l'étude de Peterson et Lou (1991, cités par Craig, 2001) réalisée à Portland, il a été demandé à 103 élèves de 9^e année⁴ de réaliser une évaluation d'écriture. Ensuite, leurs copies ont été recopiées à l'ordinateur en veillant à bien reproduire les erreurs commises. Une note allant de 1 à 5 selon certains critères prédéfinis a ensuite été attribuée à ces copies. Globalement, bien qu'il n'y ait pas de différences, en termes de contenu entre les copies écrites à la main et celles rédigées par traitement de texte, il en ressort que les copies manuscrites ont été mieux notées que celles réécrites à l'ordinateur.

Dans leurs études, Powers et al. (1994) et Brown (in press) rejoignent ces constats, et ajoutent l'idée selon laquelle l'écart de notes entre les dissertations manuscrites et celles dactylographiées était plus important pour les dissertations de qualité inférieure (désordonnées) que pour celles de qualité supérieure (soignées).

Les résultats de l'étude de Bridgeman et Cooper (1998) rejoignent également ces constats. Plus particulièrement, cette étude basée sur une analyse des données de plusieurs programmes de tests nationaux, s'est intéressée à l'influence que pouvait exercer le passage d'un test papier-crayon à un test informatisé sur la performance des élèves. Les résultats indiquent que les dissertations écrites à la main obtiennent, en moyenne, des scores plus élevés que celles rédigées par traitement de texte et ce, peu importe le sexe ou l'origine ethnique de l'élève.

Différents facteurs pourraient expliquer ces écarts de notes entre les copies manuscrites et celles dactylographiées, et en voici quelques-uns :

- Les erreurs : elles sont plus faciles à repérer sur ordinateur alors qu'une mauvaise écriture, à l'inverse, peut « masquer » une multitude de fautes (Craig, 2001 ; Powers, et al., 1994 ; Arnold et al., 1990) ;
- Les attentes : les évaluateurs auraient des attentes plus élevées vis-à-vis des évaluations traitées par ordinateur comparativement à celles écrites à la main (Craig, 2001 ; Arnold et al., 1990 ; Powers et al., 1994).

⁴ 9th Grade aux États-Unis = 3^e année secondaire en Belgique

- Les préférences : les évaluateurs auraient une préférence pour les dissertations manuscrites plutôt que pour celles dactylographiées et ce, même si les premières s'avéraient plus difficiles à lire (Arnold et al., 1990)
- La longueur perçue : une copie qui semble plus courte, comme c'est le cas lorsqu'elle est dactylographiée, sera moins bien évaluée qu'une copie plus longue (Powers et al., 1994 ; Peterson & Lou, 1991, cités par Craig, 2001 ; Arnold et al., 1990). Cela sous-entend que si une copie est plus courte, c'est parce que l'élève a fait moins d'efforts (Klein & Taub, 2005).

Malgré ces éléments allant en défaveur des copies dactylographiées, Bridgeman et Cooper (1998) mettent également en évidence que les évaluateurs sembleraient noter avec une plus grande fiabilité les dissertations rédigées par traitement de texte que les dissertations manuscrites. Le passage de la copie manuscrite à la copie rédigée par traitement de texte paraîtrait donc être associé à une plus grande fiabilité.

Dans son étude, McColly (1970, cité par Charney 1984) révèle également les avantages liés à une copie dactylographiée, si bien qu'il recommande de dactylographier chaque copie avant de l'évaluer afin de minimiser les effets liés à l'écriture et à la propreté. En effet ce dernier part du principe qu'il existe un lien entre l'apparence de l'écriture - comme la qualité de l'écriture - et la note totale que recevra la copie de sorte qu'une copie écrite avec une mauvaise écriture aura tendance à recevoir une note plus basse, comparativement à une copie avec une belle écriture.

2.3.3.2 Les biais liés à l'évaluateur

Les enseignants jouent un rôle important en évaluation. En effet, c'est grâce à leur jugement concernant une copie que la performance de l'élève va pouvoir être traduite sous forme de résultat (Lim, 2011). Par résultat, dans le cadre de ce mémoire, nous faisons référence à la (aux) note(s) attribuée(s) à une copie par un évaluateur. Il est donc crucial que les enseignants puissent fournir des évaluations appropriées et ce, de manière cohérente (Lim, 2011). En effet, la validité de la notation occupe une place essentielle (Shaw & Weir, 2007).

Toutefois, divers auteurs mettent en évidence que certaines caractéristiques propres aux évaluateurs peuvent venir influencer les résultats scolaires des élèves (Darling-Hammond, 2000 ; Detroz, 2021). Par exemple, **les compétences** des évaluateurs pourraient expliquer, en partie, les différences de résultats entre élèves.

En effet, dans leur méta-analyse, Nye et al. (2004) soulignent notamment que les compétences des enseignants expliqueraient entre 7% et 21% de la variance des gains de rendement chez les élèves. Or les compétences d'une personne se fondent notamment sur sa formation et sur son expérience, ce qui nous amène aux deux points suivants (Johansson et al., 2013).

Si on en croit l'étude de Shohamy et ses collègues (1992, cités par Lim, 2011), la **formation** des évaluateurs exercerait également une influence sur leur capacité à évaluer et donc, sur les notes attribuées aux élèves. Plus précisément, au travers de leur étude, ils ont comparé l'évaluation de 50 copies réalisées par 10 évaluateurs formés en évaluation (sur les échelles/critères holistique, communicatif et précision) avec l'évaluation de 50 copies réalisées par 10 évaluateurs non formés. Il en est ressorti que, même si les coefficients de fiabilité inter-évaluateurs étaient généralement élevés, les évaluateurs formés étaient plus fiables que les évaluateurs non formés.

L'expérience des enseignants est également connue pour être un facteur important pour la réussite des élèves. En effet, plusieurs auteurs (Darling-Hammond, 2000) ont montré que, comparativement à leurs collègues ayant trois ans d'expérience voire plus, les enseignants ayant moins de un à deux ans d'expérience sont moins efficaces. Néanmoins, l'effet de l'expérience semble s'amoinrir avec le temps (après +/- cinq ans).

S'il existe des différences entre les évaluateurs expérimentés et ceux qui ne le sont pas, il s'avère qu'il existe également des différences entre les évaluateurs expérimentés, et l'étude de Wolfe et ses collègues (1998) permet d'en témoigner. Au travers de cette étude, dans un premier temps, 36 correcteurs ont été classés en 3 groupes d'expertise : compétents, intermédiaires et experts et, dans un deuxième temps, il a été question d'analyser leur manière d'évaluer. Plus précisément, pendant que chacun des correcteurs évaluait 24 dissertations, les comportements cognitifs de ceux-ci ont été examinés au travers d'un protocole de réflexion à voix haute, ainsi qu'au travers de quatre variables : 1) l'orientation de la notation, 2) le degré de spécificité de la notation, 3) le degré d'adoption des critères de notation de la grille et 4) les actions de traitement. Les résultats obtenus indiquent l'existence de différences qualitatives dans la manière d'évaluer des différents groupes. En effet, comparativement aux évaluateurs compétents, les évaluateurs experts étaient : moins souvent interrompus dans leur processus de lecture de la dissertation, capables de ne pas porter de jugement sur la dissertation avant la fin de la lecture de celle-ci, en mesure de faire des commentaires plus généraux et non centrés sur une partie spécifique de la dissertation, capables de considérer tous les critères de la même manière et avaient tendance à se baser davantage sur les critères de la grille d'évaluation pour justifier leur notation.

Par conséquent, bien que tous les participants à cette étude soient formés et expérimentés, il existait tout de même des différences qualitatives dans leur approche de la tâche d'évaluation.

Par ailleurs, toujours concernant l'expertise des évaluateurs, plusieurs études ont démontré des différences entre les évaluateurs inexpérimentés et les évaluateurs expérimentés en ce qui concerne la sévérité générale de la notation et leurs stratégies de notation (Shohamy et al., 1992, cités par Lim, 2011 ; Weigle 1999).

Pour exemple, dans l'étude de Weigle (1999), il a été demandé à 16 évaluateurs d'évaluer des essais écrits par des étudiants en anglais langue seconde sur 2 questions différentes, de difficulté plus ou moins équivalente et ce, à l'aide d'une grille de notation. Dans ces 16 participants se trouvaient 8 évaluateurs expérimentés, c'est-à-dire des évaluateurs qui ont déjà eu l'occasion d'évaluer des compositions ESLPE (English as Second Language Placement Exam) et 8 évaluateurs inexpérimentés qui, à l'inverse, n'ont aucune expérience en matière d'évaluation de composition ESLPE. Les résultats indiquent, au travers d'une analyse quantitative (mesure de Rash) et qualitative (analyse verbale des protocoles) des données, que les évaluateurs inexpérimentés ont tendance à être partiellement (sur une question mais pas sur l'autre) plus sévères que les évaluateurs expérimentés. Mais cette différence a tendance à disparaître après que les évaluateurs inexpérimentés aient suivi des formations (sur la normalisation de l'ESLPE, sur la notation opérationnelle des compositions de l'ESLPE, et des formations sur des procédures de réflexion à voix haute). Cette différence peut être expliquée par l'analyse qualitative des données qui démontre l'existence d'un lien entre ces différences et la facilité avec laquelle la grille de notation pouvait être appliquée aux deux questions, ainsi qu'un lien avec la façon dont les deux groupes percevaient les questions. Face à cet écart entre les évaluateurs novices et expérimentés, **la formation** des évaluateurs peut donc s'avérer d'une certaine utilité pour réduire ces différences (Calfee, 1983, cité par Weigle, 1999 ; Weigle, 1999). En effet, la formation des évaluateurs est fréquemment présentée comme étant un moyen de compenser les différents antécédents des évaluateurs ainsi que d'ajuster les attentes de façon à réduire la variabilité de la notation fondée sur des attentes disparates (Charney, 1984 ; Huot, 1990b, cité par Weigle, 1999).

L'effet de l'expérience des évaluateurs étant démontré par plusieurs études, nous sommes encline à nous demander si cet effet persiste au fur et à mesure que les années d'expérience s'accumulent. En effet, le développement de l'évaluation (formation, expérience, etc.) comporte également une dimension temporelle. Toutefois, peu d'études se sont intéressées aux effets à plus long terme (Lim, 2011).

A cette question, l'étude longitudinale de Lim (2011) permet de fournir une réponse. Au travers de cette étude, les performances des évaluateurs, qu'ils soient nouveaux (novices) ou expérimentés, ont été examinées à plusieurs reprises. Plus précisément, dans cette étude, une série de copies écrites en langue seconde (issues des données opérationnelles de la section d'écriture du MELAB, un examen international de compétences en anglais) ont été fournies à des évaluateurs afin d'examiner, sur 3 périodes de temps de 12 à 21 mois, la qualité de l'évaluation des évaluateurs novices et des évaluateurs expérimentés. Par évaluateur expérimenté, Lim faisait référence soit à la durée pendant laquelle l'évaluateur a évalué, soit à la quantité d'évaluations qu'il a effectuée. Les résultats démontrent que, dans ce contexte d'étude, la sévérité et la cohérence des évaluateurs novices ne se différencient pas toujours de celles de leurs homologues expérimentés. Dans le cas où l'on pouvait les distinguer, bien que la qualité de leur évaluation reste toujours inférieure à celle des évaluateurs expérimentés, la qualité d'évaluation des évaluateurs novices s'est améliorée relativement rapidement au cours du temps et ils ont été capables de maintenir cette qualité d'évaluation dans la durée. Ces constats pourraient être expliqués par la fréquence ou la quantité des évaluations qui ont été effectuées par les évaluateurs. Cette étude fournit donc des preuves concernant le développement initial et à long terme des évaluateurs et concernant la qualité des évaluations.

Outre les compétences, l'expérience, et la formation des évaluateurs, les biais liés au **genre de l'évaluateur** exercent également un effet sur le jugement évaluatif.

Dans l'appréhension du monde social, les catégories de sexe occupent une place importante car elles sont à la fois des catégories cognitives et sociales (Mosconi, 1999). De ces catégories découlent une série d'inférences permettant d'attribuer des traits stéréotypés incluant des traits de personnalités, des professions, et des rôles. C'est ce que Mosconi (1999, p.89) appelle les « effets d'étiquetage ». Ces stéréotypes féminins et masculins impliquent un système de bi-catégorisation dans lequel les comportements, les capacités, les intérêts et les attitudes d'une personne sont catalogués comme soit féminins soit masculins et classés comme étant conformes ou déviants par rapport à ces stéréotypes (Mosconi, 1999).

En outre, les stéréotypes de sexe exercent des effets sur nos modes de pensée ainsi que sur nos manières de nous comporter, Pichevin (1995, citée par Mosconi 1999) déclare que les stéréotypes de sexe

[...]donnent forme et contenu à nos perceptions, normalisent nos jugements, nos évaluations, nos interprétations, nos attentes relatives aux hommes et aux femmes, guident nos comportements, modifient nos rapports à autrui et génèrent un monde à leur image. En orientant et en altérant le regard sur autrui et sur soi, ils contribuent non seulement à créer des différences entre les sexes mais aussi à hiérarchiser les deux sexes. (p.90).

Par ailleurs, il s'avère que la catégorisation de sexe est plus prégnante en situation mixte qu'en situation non-mixte. Ainsi, la mixité a tendance à accentuer les stéréotypes de sexe (Mosconi, 1999).

Dans le contexte de l'évaluation, les effets de genre ont été abordés à plusieurs reprises. Comparativement aux effets liés au genre de l'évalué, les études réalisées sur les effets liés au genre de l'évaluateur sont moins nombreuses. Lorsqu'elles s'y intéressent, elles traitent principalement de l'analyse des évaluations des enseignants par les élèves (Basow, 1995 ; Basow & Silberg, 1987 ; Martin, 1984). Le rôle de l'évaluateur dans ce type d'étude est donc incarné par les élèves et les évalués sont les enseignants.

Dans l'étude de Basow (1995), par exemple, une série d'évaluations réalisées par des professeurs et évaluées par des élèves du secondaire, au cours d'une période de 4 ans, ont été analysées. Les analyses des résultats témoignent de l'existence d'une interaction significative entre le sexe des enseignants et celui des élèves. Globalement, il en ressort que les professeurs féminins ont tendance à recevoir des notes plus élevées de la part des élèves de sexe féminin et des notes plus faibles de la part des élèves de sexe masculin. Les notes des professeurs masculins, quant à elles, ne semblent pas être affectées par le sexe des étudiants.

Cette étude rejoint en certains points celle réalisée par Basow et Silberg (1987), dans laquelle plus de 1000 étudiants du premier cycle ont évalué 16 professeurs hommes et femmes en termes d'efficacité de l'enseignement et de caractéristiques liées au sexe. Les résultats obtenus indiquent notamment que les étudiants masculins ont attribué aux professeurs féminins des notes significativement moins bonnes comparativement aux notes attribuées aux professeurs de sexe masculin sur les six mesures d'évaluation de la qualité de l'enseignement. De même, les étudiants de sexe féminin ont évalué moins favorablement les professeurs de sexe féminin comparativement aux professeurs de sexe masculin et ce, sur trois mesures.

Enfin, de manière plus générale, dans son étude Martin (1984) indique que les professeurs qui correspondent aux stéréotypes sociaux reçoivent de meilleures évaluations que les professeurs qui s'écartent des attentes stéréotypées.

Outre le genre de l'évaluateur, dans les biais cognitifs de l'évaluation des apprentissages figurent également les biais liés à la **fatigue**.

Des recherches relatives aux habitudes d'évaluation, à l'aspect humain de correction ainsi qu'aux éléments exerçant une influence sur la correction ont été effectuées par Roberge (2006).

En observant des enseignants en situation d'évaluation, elle a pu relever que des éléments, relatifs au contexte de l'évaluation, pouvaient influencer le travail intellectuel et physique des évaluateurs durant la correction. Un des éléments relevés concernait l'effet lié à la fatigue des enseignants. Elle distingue notamment deux effets de fatigue, à savoir la fatigue physique qui peut être par exemple associée à des yeux qui se ferment ou qui piquent. La seconde est celle relative à la fatigue mentale, c'est-à-dire celle due à la complexité de la tâche. En effet, il arrive parfois que les corrections à effectuer demandent un effort tel qu'elles peuvent plonger l'enseignant dans un état de surcharge cognitive (Roberge, 2006).

Cet effet de fatigue se retrouve également chez d'autres auteurs, comme Chaumont et Leroux (2018), qui, au travers de leur étude, indiquent que les biais de fatigue ou de l'humeur du correcteur peuvent entraîner une diminution de l'attention et de la précision du correcteur. Guyomar et al. (2015), quant à eux, soulignent l'importance de limiter le nombre de critères dans la grille d'évaluation afin d'éviter **la lassitude** de l'évaluation et ce, tant pour l'évaluateur que pour l'évalué.

2.3.3.3 Les biais liés à l'évalué

A côté de ces biais liés à l'évaluateur, des particularités intrinsèques à l'élève peuvent également influencer subjectivement le jugement évaluatif des enseignants (Leclerq et al., 2004).

Pour illustrer cela, prenons tout d'abord l'étude réalisée par Caverni et al. (1975) au sein de laquelle ils proposent à des enseignants d'évaluer 4 mêmes copies accompagnées de 5 notes censées avoir été obtenues précédemment par l'auteur de la copie. Il en ressort que **les résultats antérieurs** d'un élève, même inconnu, exercent une influence sur le jugement de l'évaluateur (De Landsheere, 1992). Leclerq et ses collègues (2004) expliquent cet effet par la tendance avec laquelle un correcteur attribuerait à un étudiant une note comparable à celle qu'il a obtenue antérieurement. Toutefois, ce n'est pas le seul élément en lien avec l'évalué qui peut venir biaiser le jugement de l'évaluateur. Nous identifions également trois autres éléments que voici :

Il s'avère que **l'origine sociale** des élèves exerce également une influence sur le jugement évaluatif des évaluateurs (Pourtois, 1978 ; Sprietsma, 2013).

Pour illustrer cet effet, prenons l'étude réalisée par Sprietsma (2013) qui s'est penchée sur l'influence que pouvait avoir l'origine ethnique de l'élève sur l'évaluation. Dans le cadre de cette étude, elle attribue de manière aléatoire des noms allemands ou turcs à 10 dissertations d'élèves de 4^e année secondaire. Il en ressort que les enseignants ont tendance à évaluer plus négativement une copie lorsque celle-ci a été réalisée par un enfant ayant un prénom qui ne correspond pas à la nationalité de l'enseignant. Cette attitude moins positive des enseignants à l'égard des personnes d'origine étrangère n'est pas liée à la notation ou au biais de recommandation mais à des attentes de performance plus faibles à leur égard. Ainsi, en corrigeant la copie de l'élève d'origine étrangère, l'enseignant va chercher davantage pour trouver des erreurs supplémentaires qui confirment ses attentes (Sprietsma, 2013).

Toutefois, selon Sprietsma (2013), étant donné que les élèves migrants sont déjà confrontés à des compétences linguistiques et à un capital culturel moindres, ainsi qu'à une discrimination statistique sur le marché du travail, la suppression d'une sanction supplémentaire telle que celle liée à la discrimination dans l'évaluation serait la bienvenue.

Le même type de conclusion peut également être tirée concernant le **genre de l'évalué**. Dans le processus d'évaluation, Dardenne (2019) met en évidence l'existence d'une reproduction des inégalités de genre. Ainsi, les attentes différenciées des enseignants en fonction du genre de leurs élèves se reflètent dans les évaluations des copies des élèves. En effet, plusieurs auteurs ont montré que, en mathématiques et en physique, les garçons ont tendance à être surévalués lorsqu'ils sont bons et à être sous-évalués lorsqu'ils sont plus faibles. A l'inverse, les évaluateurs ont tendance à être plus laxistes envers les filles faibles et à sous-évaluer leurs bonnes performances (Lafontaine & Monseur, 2009 ; Mosconi, 1999).

Plus particulièrement, dans l'étude expérimentale de Lafontaine et Monseur (2009), les auteurs ont comparé les évaluations réalisées par des enseignants de mathématiques sur des copies de différentes qualités (mauvaises, insuffisantes, bonnes ou très bonnes), réalisées par des élèves en fin d'enseignement secondaire, en fonction du genre de l'évalué et de l'évaluateur. Pour ce faire, chaque copie de chaque niveau a été soit attribuée à un garçon soit attribuée à une fille. Face à des copies de mathématiques identiques, mais attribuées dans un cas à des garçons, dans l'autre à des filles, il apparaît que les enseignants ont tendance à récompenser ce qui correspond à leurs stéréotypes, par exemple « les filles sont plus faibles en mathématiques que les garçons ». De sorte qu'en moyenne, les copies attribuées aux garçons sont évaluées légèrement plus positivement que celles attribuées aux filles.

De plus, les évaluateurs ont tendance à surestimer les copies moyennes et bonnes des garçons et à sous-estimer les copies équivalentes attribuées à des filles. L'inverse est observé pour les copies plus faibles et très faibles : lorsqu'elles sont attribuées à une fille, elles sont évaluées de manière plus indulgente. Toutefois, il s'avère que lorsque l'on soumet à ces enseignants un questionnaire destiné à appréhender l'existence éventuelle d'attentes stéréotypées, ils soutiennent massivement les propositions égalitaires (non sexistes) et s'opposent à une possible différence entre les garçons et les filles par rapport aux mathématiques. Les résultats obtenus aux questionnaires sur les attentes laissent donc à penser que les enseignants ne sont pas conscients de ces inégalités de traitement.

Enfin, le **statut socio-économique** (SES) d'un élève s'avère également être un biais lié à l'évalué, et l'étude d'Autin et al. (2019), qui suit le même design de recherche que l'étude de Sprietsma (2013) mais en étudiant les biais de l'évaluation relatifs au statut socio-économique des élèves évalués, permet de l'illustrer. À nouveau, seul le prénom de l'élève diffère entre les copies totalement identiques par ailleurs (ex : Kevin pour SES faible vs Pierre-Henry pour SES élevé). Il ressort de cette étude que, dans des situations d'évaluation normatives, les enseignants ont tendance à sous-évaluer les compétences de l'élève à faible SES comparativement à celles de l'élève à SES élevé. Ainsi, les connaissances des évaluateurs quant à la classe sociale de leurs élèves peuvent influencer leur évaluation en faveur des élèves privilégiés et ainsi biaiser l'évaluation. Toutefois, les recherches actuelles suggèrent que ce n'est pas l'évaluation en elle-même qui déclenche un tel biais et que ce ne sont pas non plus les enseignants eux-mêmes qui sont biaisés mais que ce serait plutôt la fonction (ou le contexte) donnée à l'évaluation qui influence les biais de l'évaluation chez l'enseignant. En effet, c'est ce contexte/cette fonction de l'évaluation qui peut déclencher ou empêcher une évaluation discriminatoire. Ces recherches ont donc permis de mettre en évidence qu'un écart entre les classes sociales en matière d'évaluation apparaît lorsque l'évaluation est utilisée à des fins sélectives (trier les élèves) plutôt que lorsqu'elle est utilisée à des fins éducatives (favoriser l'apprentissage). Afin de garantir l'égalité dans les établissements d'enseignement, il convient donc d'accorder une plus grande attention au rôle et à la signification de l'évaluation. Par conséquent, les résultats de ces diverses études concernant les biais liés à l'évalué attirent l'attention sur un phénomène d'inégalité de traitement persistant qui, par ailleurs, est susceptible d'influencer le sentiment d'efficacité, la perception de soi et l'attribution de la réussite à des causes internes ou externes (Lafontaine & Monseur, 2009).

Compte tenu de l'existence de tous ces biais d'évaluation, il convient de développer des outils adaptés à chacune des méthodes d'évaluation (Chirart et al., 2014), ce qui nous amène à la section suivante.

2.4 Les grilles d'évaluation

Pour réduire cette subjectivité présente dans le phénomène de l'évaluation, plusieurs auteurs (Berthiaume et al., 2011 ; Reddy & Andrade, 2009 ; Timmerma et al., 2011) préconisent l'utilisation de grilles d'évaluation, et plus spécifiquement des grilles critériées. Ces grilles permettent de définir les critères d'évaluation, les indicateurs ou les niveaux de performance correspondant à chacun des critères (Berthiaume et al., 2011). Elles permettent donc de centrer l'évaluation sur des points précis en lien avec les objectifs d'apprentissage et de limiter les variations d'interprétation qu'il pourrait y avoir entre différents correcteurs à propos d'une même copie (Chirart et al., 2014). Brown et Glasner (1999 cités par Mémo pédagogique, 2016) rejoignent cette idée en mettant en évidence que l'usage des grilles critériées est un bon moyen pour atténuer les biais d'évaluation.

Selon Berthiaume, et al. (2011, p.8) :

Une grille critériée est un tableau qui détaille à la fois les critères utilisés pour interpréter la preuve d'apprentissage fournie par l'étudiant dans un travail écrit ou encore lors d'une présentation orale, et les indicateurs ou niveaux de performance possibles pour chaque critère.

Ces guides de notation (Mertler, 2001) sont rarement utilisés pour évaluer les réponses des élèves dans des tests qui impliquent des réponses courtes, mais sont généralement utilisés afin de juger l'adéquation des réponses des élèves à un test de performance. Les tests de performance, quant à eux, sont habituellement utilisés pour identifier le degré de maîtrise d'un élève en regard d'une compétence importante (Popham, 1997).

Par ailleurs, au niveau de la composition des grilles critériées, on retrouve habituellement (Chirart et al., 2014) : les dimensions qui sont directement en lien avec les objectifs d'apprentissage et qui reprennent les éléments sur lesquels le regard de l'évaluateur va porter ; les critères qui correspondent à une décomposition d'une dimension et qui permettent d'apporter davantage de précision dans la dimension évaluée ; une échelle d'appréciation qui permet de distinguer les différents niveaux de performance (ex : « Validé/non validé », « 1,2,3,4, », etc) ; des descripteurs de performance (indicateurs) qui permettent de définir la qualité observable de la performance attendue, et enfin une pondération et un barème qui renseignent sur l'importance d'une dimension ou d'un critère en regard de la note finale.

Outre le fait qu'elles permettent de réduire la subjectivité interprétative apparaissant lors de l'évaluation des apprentissages et donc d'augmenter le niveau de fidélité (ou fiabilité) (Berthiaume et al, 2011), les grilles d'évaluation critériées comportent également de nombreux autres avantages dont en voici quelques-uns.

Tout d'abord, selon Berthiaume et ses collègues (2011), un des principaux intérêts de l'utilisation des grilles critériées est qu'elles peuvent être utilisées à différents moments de l'apprentissage.

Les grilles critériées peuvent aussi bien être utiles durant les périodes d'enseignement que durant l'évaluation (Herawati, 2011).



Figure 2. *l'usage de la grille critériée inscrite dans le temps (Berthiaume et al., 2011).*

En effet, en amont du cours, ces grilles d'évaluation critériées vont d'une part, permettre à l'enseignant de faciliter la structuration de ses séances d'enseignement et d'autre part, l'obliger à définir les objectifs à l'avance (Berthiaume et al., 2011) et donc à clarifier les apprentissages visés (Chirart et al., 2014).

Ensuite, communiquées aux élèves au début du cours, elles leur permettent de mieux percevoir les exigences/attentes de l'enseignant, et donc d'avoir une idée claire des attendus de l'évaluation (Andrade, 2000 ; Berthiaume et al., 2011 ; Chirart et al., 2014 ; *Mémo pédagogique*, 2016).

Durant le cours, en connaissant les différents critères et en les intériorisant, les élèves sont davantage capables de s'auto-évaluer (Andrade, 2000). Berthiaume et ses collègues (2011) rejoignent cette idée en soulignant que la connaissance et la compréhension des critères permettraient le développement de compétences réflexives de sorte que les élèves apprendraient à s'auto-évaluer et à adapter leur manière d'apprendre en fonction de leurs résultats aux auto-évaluations. Herawati (2011) rejoint cette idée en déclarant que ces grilles d'évaluation critériées aident « [...] les élèves à développer un processus de réflexion sur la manière de réussir [...] » (p.445).

De plus, les grilles critériées donnent l'opportunité aux élèves d'avoir un retour systématique sur leurs prestations, permettant ainsi d'améliorer leurs performances (Berthiaume et al., 2011 ; Herawati, 2011). Elles permettent également d'aider les élèves dans la planification de leur travail (Herawati, 2011) et de favoriser leurs apprentissages (Andrade, 2000).

Enfin, après le cours, l'usage des grilles critériées permet de fournir des informations sur les points forts et les points faibles des élèves (Andrade, 2000 ; Herawati, 2011).

Les grilles critériées semblent présenter des avantages tant pour l'évaluateur que pour les évalués. Autrement dit, comme le mettent en évidence les propos tenus par d'Amours (s.d), les grilles critériées renforcent la perception de l'équité du jugement grâce à l'usage de critères explicites et à l'utilisation d'un instrument uniforme. Elles permettent également de renforcer la validité de l'évaluation en s'assurant que l'évaluation soit en lien avec les objectifs visés par la séquence d'enseignement. Elles rendent aussi les évaluation plus objectives et augmentent la fidélité de l'évaluation, c'est-à-dire la probabilité que différents correcteurs attribuent un résultat similaire à une même copie. De plus, elles renforcent la perception de la transparence en communiquant de manière explicite les attentes de l'enseignant aux élèves. Enfin, lorsque les grilles critériées sont utilisées dans le cadre d'évaluations formatives et sommatives, elles permettent de fournir une rétroaction aux élèves (d'Amours, s.d).

3. Objet de recherche et hypothèses

3.1 Objet de recherche

Suite à la mise en évidence de toutes ces inégalités et compte tenu de la pratique récurrente du redoublement en Belgique, il nous semble légitime de nous poser les questions suivantes : les élèves redoublants subissent-ils des inégalités de traitement dans les évaluations au même titre que les élèves migrants dans une épreuve de français, que les filles dans une épreuve de mathématiques et que les élèves de statut socio-économique faible dans une épreuve à visée sélective ? Un enseignant va-t-il, inconsciemment, évaluer de manière plus négative un élève sous prétexte qu'il a connu le redoublement ? Le statut de redoublant peut-il être considéré comme un facteur biaisant l'évaluation ?

En d'autres termes, l'**objectif** du présent mémoire est de répondre à la question de recherche suivante :

Dans quelle mesure, l'ancrage du stéréotype du redoublement dans la copie d'un élève de 3^e année secondaire, accompagnée ou non d'une grille d'évaluation, va-t-il influencer l'évaluation donnée par des enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles sur cette copie ?

3.2 Hypothèses

En réponse à cette question, et suite aux diverses lectures menées sur le sujet, voici **les trois hypothèses** de recherche que nous pouvons formuler :

3.2.1 Hypothèse n°1 :

A l'heure actuelle, plusieurs études se sont intéressées à l'effet que pouvait avoir l'ancrage d'un stéréotype sur le jugement évaluatif d'évaluateurs : Autin et al. (2019) se sont intéressés à l'effet du statut économique de l'évalué tandis que Lafontaine et Monseur (2009) se sont centrés sur l'effet du genre de l'évalué. Sprietsma, quant à elle, (2013) s'est penchée sur l'influence de l'origine ethnique de l'élève, Caverni et al. (1975) sur l'influence des résultats antérieurs de l'élève, etc. Malgré toutes ces études, aucune n'a traité spécifiquement de l'effet de l'ancrage du stéréotype relatif au statut de redoublant sur le jugement évaluatif. Par conséquent, pour formuler les propos que nous avançons ci-dessous, nous nous sommes principalement inspirée des études précédemment citées qui traitent de l'impact d'autres stéréotypes sur le jugement évaluatif d'évaluateurs.

Tout d'abord, mentionner sur la copie de l'élève si celui-ci a connu ou non le redoublement va avoir pour incidence d'activer chez l'enseignant les stéréotypes associés au redoublement. Ces stéréotypes étant négatifs (Baye & Dachet, 2020), cette activation pourrait, parallèlement, impacter négativement le jugement évaluatif de l'enseignant de sorte que l'élève ayant connu le redoublement serait sous-évalué comparativement à celui qui ne l'a, quant à lui, jamais connu (Autin et al., 2019 ; Boraita, 2015 ; Crahay, 2019 ; Crisafulli et al., 2002, cités par Crahay 2019 ; Detroz, 2021 ; Lafontaine et al., 2019 ; Sprietsma, 2013.). Autrement dit, à performance égale, l'ensemble des connaissances des évaluateurs relatives au stéréotype, peuvent venir biaiser inconsciemment leur évaluation en défaveur des élèves portant le stéréotype – dans ce cas-ci les élèves redoublants –, contribuant ainsi à la reproduction d'inégalités sociales, et donc à une forme de discrimination dans la manière d'évaluer des enseignants (Lafontaine et al., 2019).

Cela étant, voici l'hypothèse que nous pouvons formuler : lorsque l'élève est présenté comme étant redoublant (inscrit sur la copie de l'élève), les enseignants évaluent plus négativement sa copie, comparativement à la même copie mais sur laquelle l'élève n'est pas signalé comme étant redoublant.

3.2.2 Hypothèse n°2 :

Concernant la présence ou non d'une grille d'évaluation, comme le mettent en évidence plusieurs auteurs, le recours aux grilles critériées permettrait de réduire la subjectivité inhérente à tout acte évaluatif et, partant, d'augmenter la fidélité et l'équité dans la correction (Berthiaume et al., 2011 ; Blown & Glasner, 1999, cités par Mémo pédagogique, 2016 ; Chirart et al., 2014 ; Reddy & Andrade, 2009 ; Timmerman et al., 2011). En d'autres termes, l'usage des grilles d'évaluation pourrait participer à la diminution de l'impact des biais en évaluation (Blown & Glasner, 1999, cités par Mémo pédagogique, 2016 ; Detroz, 2021).

En effet, à l'aide des grilles critériées, un certain nombre de malentendus peuvent être écartés et des suggestions précises d'amélioration peuvent être formulées par les évaluateurs à l'égard des élèves. En leur absence, on donne la possibilité aux évaluateurs d'avoir une variation de jugement au cas par cas, ce qui risque d'amener les évaluateurs à juger des travaux similaires d'élèves sur base de critères différents (Berthiaume et al., 2011).

Sur base de ces données, nous formulons l'hypothèse selon laquelle la variance observée entre les notes attribuées aux copies corrigées à l'aide d'une grille d'évaluation, sera moins importante que celle observée entre les notes des copies corrigées sans grille d'évaluation.

3.2.3 Hypothèse n°3 :

Les grilles d'évaluation étant des outils efficaces pour diminuer les biais d'évaluation (Brown & Glasner, 1999, cités par Mémo pédagogique, 2016 ; Detroz, 2021), il nous a semblé intéressant d'observer la manière avec laquelle ces grilles pouvaient influencer, si effet il y a, l'effet de l'ancrage stéréotype sur le jugement évaluatif des participants.

Par conséquent, cette troisième hypothèse relève de l'étude de l'effet d'interaction entre l'ancrage du stéréotype « redoublant » et la présence d'une grille d'évaluation. Face à cet effet d'interaction, nous formulerons l'hypothèse suivante : puisque l'usage des grilles d'évaluation est un bon moyen pour réduire la subjectivité présente dans le jugement évaluatif des évaluateurs (Berthiaume et al., 2011 ; Brown & Glasner, 1999, cités par Mémo pédagogique, 2016 ; Chirart et al., 2014 ; Reddy & Andrade, 2009 ; Timmerman et al., 2011), la grille d'évaluation devrait avoir une incidence positive sur l'effet négatif qu'engendrerait la présence du stéréotype « redoublant ».

En d'autres termes, dans le cadre de cette étude expérimentale, nous nous attendons à ce que les enseignants aient tendance à noter de manière plus favorable une copie redoublant accompagnée d'une grille d'évaluation qu'une même copie mais non accompagnée d'une grille d'évaluation.

Quant aux copies présentées comme étant réalisées par un élève promu, nous pouvons nous attendre à ce que l'effet aille également dans ce sens mais soit moindre que celui observé avec les « copies redoublants ».

3.2.4 Hypothèses relatives aux quatre variables sociodémographiques

Au-delà de ces trois hypothèses principales, nous allons tenter de nous positionner sur l'effet éventuel que pourrait avoir chacune des quatre **variables indépendantes relatives aux caractéristiques des participants** (le genre, la discipline d'enseignement, le nombre d'année(s) d'ancienneté dans l'enseignement, et le fait d'avoir vécu ou non le redoublement durant son parcours scolaire) sur la note totale attribuée à la rédaction, voici les hypothèses que nous pouvons émettre.

Commençons par nous positionner sur la caractéristique relative au genre (homme/femme) de nos participants. Il ressort de notre revue de la littérature que les évalués de sexe féminin ont tendance à recevoir des notes plus élevées de la part des évaluateurs féminins et des notes plus faibles de la part des évaluateurs de sexe masculin (Basow, 1995 ; Basow & Silberg, 1987). Par ailleurs, si on suit les résultats de l'étude de Basow (1995) il semblerait que les notes attribuées aux évalués de sexe masculin ne soient pas affectées par le sexe des évaluateurs. Ainsi, compte tenu de tous ces éléments et étant donné que, dans le cadre de notre étude, la rédaction à évaluer a été réalisée par un élève de sexe masculin, nous pouvons nous attendre à n'observer aucune différence significative entre les professeurs masculins et féminins dans leur manière d'évaluer cette dernière.

Intéressons-nous à présent à la variable relative à la discipline d'enseignement ainsi qu'à celle portant sur l'ancienneté des participants. La littérature met en évidence que les compétences des évaluateurs expliqueraient, en partie, les différences de résultats entre élèves (Nye et al., 2004). Or, les compétences d'une personne se fondent notamment sur sa formation et sur son expérience (Johansson et al., 2013).

Au niveau de la formation des évaluateurs, il ressort de notre revue de la littérature qu'il existe une différence entre les évaluateurs novices (peu ou pas formés dans la matière qu'ils doivent évaluer) et les évaluateurs experts (formés dans la matière qu'ils doivent évaluer) en ce qui concerne la sévérité générale de la notation et leurs stratégies de notation, de sorte que les évaluateurs novices ont tendance à être partiellement plus sévères que les évaluateurs experts, tendance qui semble disparaître après que ceux-ci aient suivi une formation (Shohamy et al., 1992, cités par Lim, 2011 ; Weigle 1999).

Face à cet écart entre les évaluateurs novices et ceux expérimentés, **la formation** des évaluateurs peut donc s'avérer être d'une certaine utilité pour réduire ces différences (Calfee, 1983, cité par Weigle, 1999 ; Weigle, 1999). Par ailleurs, les évaluateurs formés auraient tendance à être plus fiables dans leur manière d'évaluer que les évaluateurs non formés (Shohamy et al., 1992, cités par Lim, 2011).

Nous pouvons, dès lors, émettre l'hypothèse selon laquelle les participants « non professeurs » de français attribueront une note plus faible à la rédaction, comparativement participants professeurs de français pour lesquels corriger une rédaction relève de leur domaine de compétence.

Dans une même logique, nous pouvons nous attendre à ce que les enseignants ayant un certain nombre d'années d'ancienneté dans le monde de l'enseignement, soient plus laxistes dans leur manière d'évaluer que les enseignants débutants, qui à l'inverse, se montreraient plus sévères. Plus précisément, nous nous attendons à ce que les enseignants débutants attribuent une note plus faible à la rédaction et ce, comparativement aux enseignants ayant un certain nombre d'années d'ancienneté.

Enfin, positionnons-nous sur la caractéristique relative au fait d'avoir connu ou non le redoublement durant son parcours scolaire.

Premièrement, soulignons que les effets négatifs du redoublement pour celui/elle qui le subit sont nombreux : on y retrouve une incidence négative sur son adaptation socio-émotionnelle (Jimerson, 2001), un effet négatif sur sa performance future avec une diminution de ses performances qui peut s'expliquer par la crainte qu'il a d'échouer (Crahay, 2019 ; Selimbegovic et al., 2011, cités par Crahay, 2019), une augmentation du risque de décrochage scolaire qui s'avère d'autant plus importante que l'élève a été retenu plusieurs fois (Jimerson et al., 2002 ; Witte et al., 2013, cités par Galand et al., 2019), etc. Par ailleurs, les stéréotypes dont sont fréquemment victimes les élèves redoublants (ex : ce sont des « fainéants, paresseux, malpolis, etc. »), entraînent à leur tour des effets délétères tels que : le phénomène de confirmation des attentes (Snyder & Stukas, 1999), l'effet solo (Crahay, 2019), de la stigmatisation qui, à son tour, peut entraîner une baisse d'estime de soi, un désinvestissement dans les domaines concernés, du désespoir ou encore de la colère (Crahay 2019). On retrouve également dans les effets délétères des stéréotypes, la menace du stéréotype (Crahay, 2019) ; dès qu'ils seront confrontés à une situation où leurs compétences scolaires seront amenées à être évaluées, les élèves redoublants craindront de confirmer le stéréotype qui porte sur eux.

Cette crainte, à son tour, entraînera une diminution de leurs résultats scolaires. La confrontation aux stéréotypes négatifs peut entraîner, à long terme, des conséquences au niveau des émotions, de la cognition et de la réalisation des individus stigmatisés (Crocker & Major, 1989 ; Croizet & Martinot, 2003). En effet, à force d'être confrontés à ces sentiments de menace, les individus vont, progressivement, se désengager et se désinvestir du domaine dans lequel ils sont impliqués (Crahay, 2019).

Deuxièmement, puisque le vécu exerce une influence sur les attentes, sur les stéréotypes que l'on peut avoir et puisque les attentes peuvent, à leur tour, affecter la façon dont les enseignants perçoivent et évaluent les performances des élèves (Sprietsma, 2013), nous pouvons nous attendre à obtenir, dans notre étude, le résultat suivant: comparativement aux enseignants n'ayant pas vécu le redoublement durant leur parcours scolaire, les enseignants l'ayant vécu font preuve de davantage de compréhension et d'empathie envers les copies au sein desquelles l'élève est présenté comme ayant également vécu le redoublement. Autrement dit, nous nous attendons à ce que les enseignants ayant vécu le redoublement attribuent une meilleure note aux « copies redoublants » comparativement aux enseignants n'ayant pas vécu le redoublement durant leur parcours scolaire.

3.2.5 Étude confirmatoire et exploratoire

Compte tenu de ces hypothèses, notre recherche peut être à la fois qualifiée de confirmatoire et d'exploratoire. En effet, tout d'abord au niveau de la première et de la troisième hypothèse, la littérature met en évidence que nombreuses sont les études qui se sont intéressées à l'influence que pouvaient avoir les caractéristiques des évalués sur le jugement évaluatif des évaluateurs. Toutefois, aucune étude, à l'heure actuelle, ne s'est intéressée à l'effet que pouvait avoir l'ancrage du stéréotype relatif au statut de redoublant sur le jugement évaluatif d'un évaluateur et encore moins à l'effet d'interaction que cela pouvait avoir avec la présence d'une grille d'évaluation. Ainsi, ces deux hypothèses permettent d'attester du caractère exploratoire de notre recherche puisqu'elles s'intéressent à un phénomène qui n'a jamais été mis à l'épreuve dans aucune étude. Ensuite, au niveau de la deuxième hypothèse, celle-ci fait référence à un phénomène qui a déjà été largement étudié, et dont les effets positifs ont été attestés, à savoir l'influence que peut avoir l'utilisation d'une grille d'évaluation sur le jugement évaluatif d'un évaluateur. Ainsi, dans le cadre de cette hypothèse, il s'agira de voir si les résultats des études antérieures menées sur le sujet se voient également confirmés par notre étude.

Enfin, concernant les hypothèses relatives aux variables sociodémographiques, bien que les résultats des études s’y référant ne soient pas aussi éclairants que ceux obtenus en regard de la deuxième hypothèse, nous pouvons malgré tout identifier une certaine tendance dans ceux-ci. Par conséquent, ces dernières hypothèses ont été formulées de manière à confirmer des suppositions, des résultats existants sur le sujet, témoignant ainsi à nouveau du caractère confirmatoire de notre étude.

Passons maintenant à la partie suivante de ce travail, qui concerne la méthodologie de la recherche.

4. Méthodologie

Cette partie rend compte de la méthodologie que nous avons utilisée pour mener à bien ce mémoire. Pour ce faire, nous commencerons par présenter le design de la recherche, puis nous expliquerons précisément les instruments et les données qui ont été récoltées. Ensuite, nous parlerons de l’échantillon et du public cible, pour enchaîner ensuite avec la procédure de recrutement et le laboratoire cognitif. Nous expliquerons comment nous avons réalisé ce dernier avant de présenter ses résultats. Nous nous centrerons ensuite sur les aspects éthiques de ce mémoire ainsi que sur le codage des données, pour terminer avec l’explication de la méthode utilisée pour analyser les données récoltées.

4.1 Design de la recherche

La réalisation du présent mémoire se fonde sur un design **expérimental** au sein duquel l’effet de 2 variables nominales dichotomiques sur le jugement évaluatif d’enseignants du secondaire (variable métrique) a été étudié, à savoir : redoublant vs promu et, avec grille d’évaluation vs sans grille d’évaluation.

Pour ce faire, nous avons choisi de soumettre aux participants deux « tâches » à réaliser selon un ordre bien précis.

Pour commencer, nous leur avons fourni une rédaction réalisée par un élève de 3^e année secondaire à évaluer (annexes n°4 et n°5, p.96-97). Sur cette copie, nous retrouvons la consigne générale de la rédaction portant sur le mariage forcé, ainsi qu’une courte description de la cible reprenant : le prénom de l’élève, son année scolaire et sa date de naissance. En plus de ces informations, dans certaines évaluations, nous avons ajouté, dans le profil fictif de l’élève, une information supplémentaire : le fait d’avoir redoublé en 2^e année secondaire et ce, en vue d’analyser l’effet de cette variable sur le jugement évaluatif des enseignants.

Nous avons donc choisi, comme c'est le cas dans l'étude d'Autin et ses collègues (2019), de présenter l'information relative au statut de redoublant de l'élève parmi une série d'autres informations, quant à elles identiques pour chaque évaluateur (prénom, année scolaire et date de naissance de l'élève) afin que l'activation du stéréotype se fasse de manière subtile.

De plus, dans un souci de clarté, nous avons fait le choix de nommer dans ce mémoire les copies dans lesquelles nous avons activé de manière subtile le stéréotype « copies redoublants », tandis que les autres ont été nommées « copies promus ».

Par ailleurs, pour évaluer cette rédaction, nous avons attribué à certains enseignants une grille d'évaluation (annexe n°6, p.98-99) accompagnée d'un guide de correction tandis que les autres enseignants devaient se contenter de la description de la cible et de la consigne générale de rédaction pour émettre leur jugement évaluatif concernant cette copie.

De cette manière, nous nous attachons à comparer le jugement évaluatif des quatre groupes suivants :

1. Un groupe de participants auquel il sera attribué une « copie redoublant » avec une grille d'évaluation.
2. Un groupe de participants auquel il sera attribué une « copie promu » avec une grille d'évaluation.
3. Un groupe de participants auquel il sera attribué une « copie redoublant » sans grille d'évaluation.
4. Un groupe de participants auquel il sera attribué une « copie promu » sans grille d'évaluation.

La visée de cette étude est donc **comparative**, et l'affectation des sujets à chacun de ces groupes se fera de manière strictement aléatoire. Plus particulièrement, il s'agira de générer sur Excel une séquence aléatoire de 140 chiffres variant de 1 à 4, et d'attribuer à chacun des participants la copie qui convient en suivant l'ordre de cette séquence aléatoire générée. De plus, dans cette séquence de 140 chiffres, pour nous assurer d'avoir des sous-échantillons de taille égale, nous devons veiller à obtenir 35 fois chacun des nombres.

Enfin, pour la deuxième et dernière épreuve, nous avons fourni aux participants un questionnaire contextuel à compléter, comprenant 4 questions (annexe n°8, p.102) dont l'intérêt sera expliqué dans la section suivante.

4.2 Instruments de mesure et données récoltées

4.2.1 Instruments de mesure

Pour mesurer l'effet de la variable relative au stéréotype du redoublant sur le jugement évaluatif d'un évaluateur, nous nous sommes procuré la **rédaction** d'un élève de 3^e année secondaire. Nous avons ensuite construit une grille **d'évaluation** ainsi qu'un **guide de correction** pour soutenir les enseignants dans son utilisation. Enfin, nous avons conçu un **questionnaire contextuel**. C'est sur ces instruments de mesure que portera le développement de cette section.

4.2.1.1 La rédaction

La rédaction utilisée dans le cadre de cette recherche nous a été fournie par un professeur de français de l'enseignement secondaire général et a été réalisée par un de ses élèves de 3^e année secondaire, dans le cadre de l'examen de français de fin de 3^e année secondaire.

L'objet de cette recherche étant de mesurer l'impact de l'ancrage du stéréotype « redoublant » sur le jugement évaluatif d'un enseignant, nous nous devons de nous assurer que l'effet que nous observons soit uniquement dû à l'ancrage du stéréotype « redoublant » dans la copie de l'élève. C'est pourquoi, sachant que la calligraphie de l'élève peut constituer un biais d'évaluation, nous avons décidé de retranscrire la rédaction à l'ordinateur en veillant à bien reproduire les erreurs commises par l'élève (Craig, 2001 ; Hughes et al., 1983 ; Markham, 1976 ; Powers et al., 1994). Dès lors, la copie ainsi dactylographiée, la qualité de la calligraphie de l'élève n'était plus en mesure d'interférer avec le jugement évaluatif de l'évaluateur. De plus, si on se réfère aux résultats de l'étude de Bridgeman et Cooper (1998) il semblerait que les copies dactylographiées, comparativement aux copies manuscrites, seraient associées à une plus grande fiabilité dans la notation. Dans le cadre de notre recherche, nous avons donc veillé à minimiser les biais liés à la copie. Par ailleurs, nous avons fait le choix de proposer une évaluation réalisée par un élève inconnu des participants car nous savons que les attentes des enseignants peuvent être influencées par des facteurs peu fiables concernant l'élève tels que : la performance des frères et sœurs, l'attractivité physique, le sexe, l'origine ethnique et sociale (Dusek & Josphe, 1983, cités par Trouilloud & Sarrazin, 2003 ; Jussim et al., 1944, cités par Trouilloud & Sarrazin, 2003).

En outre, nous avons choisi de présenter cette rédaction comme étant réalisée par un élève de genre masculin pour que cela soit en concordance avec les connaissances que l'on a, en FW-B, concernant le taux de redoublement selon le sexe des élèves.

En effet, les résultats de PISA 2018 mettent en évidence que le taux de redoublement en FW-B, dans l'enseignement secondaire, est systématiquement plus élevé pour les garçons que pour les filles (Lafontaine et al., 2019). De plus, comme le met en évidence Crahay (2019), le sexe masculin est davantage utilisé pour décrire l'enfant redoublant que le sexe féminin. Ainsi, présenter cette copie redoublant comme étant réalisée par un élève de genre masculin, nous paraissait plus en concordance avec la réalité du terrain.

Par ailleurs, nous sommes partie du principe que si nous souhaitons observer des variations dans le jugement évaluatif des enseignants, partir d'une copie d'un niveau moyen (satisfaisant ou suffisant) rendrait ces variations plus nombreuses, comparativement à une copie très bonne et à une copie très faible. Or, si on se réfère au système de notation en Belgique, une copie à laquelle on attribue une note allant entre 5,00 et 5,90 est jugée comme étant d'un niveau suffisant, c'est-à-dire dont le résultat satisfait les critères minimaux. En revanche, une copie oscillant entre 6,00 et 6,90 est considérée comme étant d'un niveau satisfaisant (Bureau du Secrétaire général des Écoles européennes. s.d ; Circulaire 7351 : Modalités d'utilisation des bulletins destinés aux élèves de l'enseignement secondaire ordinaire et spécialisé de forme 4. Année scolaire 2019-2020, 2019). Par conséquent, la rédaction que nous avons choisie comme instrument de mesure était une copie à laquelle était initialement attribuée, à l'aide d'une grille d'évaluation, une note de 6,67/10, ce qui correspond à une copie d'un niveau satisfaisant (annexe n°9, p.103-106).

Toutefois, voulant nous assurer que cette copie soit bien d'un niveau satisfaisant et souhaitant améliorer notre grille d'évaluation, nous avons réalisé un laboratoire cognitif au sein duquel nous avons demandé à 8 enseignants d'évaluer cette rédaction. Compte tenu des résultats de ce laboratoire cognitif, nous avons été amenée à modifier quelque peu la rédaction et, dans une plus large mesure, la grille d'évaluation et, en conséquence, la notation attribuée à cette copie. Ce laboratoire cognitif fera l'objet de plus amples explications dans une section ultérieure de ce mémoire.

4.2.1.2 La grille d'évaluation et le guide de correction

Pour réaliser cette grille d'évaluation, nous nous sommes principalement inspirée des outils d'évaluation disponibles sur le portail de l'enseignement en FW-B. Ces outils nous ont semblé être de bons instruments de référence puisqu'ils ont fait l'objet d'expérimentations auprès d'élèves de classes appartenant aux différents réseaux de la Communauté Française, et qu'ils ont ensuite été ajustés sur base des résultats obtenus mais également suite à de multiples remarques d'enseignants ayant participé à ces expérimentations (Fédération Wallonie-Bruxelles, 2010).

Par ailleurs, ces outils ont été développés afin d'aider les professeurs à adapter leur évaluation à un apprentissage centré sur l'acquisition de compétences (Fédération Wallonie-Bruxelles, 2010). Plus spécifiquement, pour construire notre grille, nous nous sommes inspirée de l'outil d'évaluation de la famille de tâches « convaincre » (annexe n°10, p.107-110) destiné aux élèves du 2^e degré. Les tâches reprises dans cette catégorie ont pour objectif de communiquer par écrit à autrui un jugement personnel argumenté (Fédération Wallonie-Bruxelles, 2010), ce qui correspond à l'objectif poursuivi de notre évaluation.

Toutefois, nous n'avons pas repris tous les critères tels qu'ils étaient présentés dans cette grille d'évaluation de référence : nous en avons modifié certains pour les rendre plus explicites et supprimé d'autres que nous ne jugions pas nécessaires et/ou adaptés à la recherche menée. En effet, notre échantillon étant composé d'enseignants du secondaire exerçant dans des disciplines diverses, rendre les critères explicites nous paraissait être un élément clé pour garantir la bonne compréhension de cette grille d'évaluation, et donc la validité de nos résultats.

Toujours dans cette idée d'explicitation des critères, suite aux résultats du laboratoire cognitif, nous avons décidé de joindre à cette grille d'évaluation un guide de correction (annexe n°7, p100-101).

Globalement, dans le cadre de cette rédaction, la compétence scripturale de l'élève est évaluée selon trois blocs de critères (annexe n°6, p.98-99) : 1. Respect de l'intention de convaincre; 2. Cohérence textuelle ; 3. Respect des normes linguistiques, dans lesquels chaque critère comprend un ou des indicateur(s) qualitatif(s) (Gérard, 2002) auxquels sont associés trois niveau(x) de maîtrise en vis-à-vis desquels on retrouve une pondération. De plus, en référence à l'étude de Guyomar et ses collègues (2015), afin d'éviter les biais liés à la fatigue de l'évaluateur, nous avons essayé de réduire au maximum le nombre de critères présents dans la grille d'évaluation tout en veillant à conserver une validité satisfaisante.

4.2.1.3 Le questionnaire contextuel

Enfin, un questionnaire contextuel a été soumis aux participants. Ce dernier a pour objectif de récolter des informations sur les participants relatives à leur genre, leur nombre d'année(s) d'ancienneté dans l'enseignement, la discipline dans laquelle ils/elles enseignent, ainsi que des informations relatives à leur vécu face au redoublement (l'ont-ils déjà vécu ou non ?).

En effet, et comme nous l'avons précédemment mis en évidence dans la revue de la littérature, ces variables relatives à l'évaluateur, que nous nommerons variables sociodémographiques, sont susceptibles d'exercer une influence sur le jugement évaluatif de ce dernier. Ainsi, nous accorderons également une attention particulière à ces variables lors de l'analyse des réponses des participants.

4.2.2 Données récoltées

Les données récoltées au travers de cette étude expérimentale sont essentiellement **quantitatives**, c'est-à-dire qu'il s'agit de données numériques qui nécessitent un traitement statistique (Dachet, 2020). Effectivement, dans le cadre de ce mémoire, il sera question de récolter et d'analyser la note totale de la rédaction donnée par chacun des évaluateurs, ainsi que la note attribuée à chacun des trois blocs de critères.

4.3 Public cible et échantillon

Dans le cadre d'une étude quantitative, la taille de l'échantillon relève d'une grande importance pour établir des statistiques valides. Il est donc important d'avoir un nombre suffisant d'observations pour obtenir des données fiables et être en mesure de tirer des conclusions pertinentes. Toutefois, compte tenu du temps que nécessite la participation à notre étude (15-30 minutes) nous ne pouvions nous permettre de viser un échantillon trop important (ex : 300) au risque de ne pas obtenir le nombre de participants souhaité. Ainsi, constituer un échantillon de convenance de minimum 120 et de maximum 200 enseignants de l'enseignement secondaire ordinaire en FW-B nous paraissait faisable compte tenu du délai à respecter. Autrement dit, le choix de la taille de notre effectif relève d'un choix arbitraire.

Par ailleurs, la seule contrainte de notre échantillon étant d'être composé d'enseignants du secondaire en FW-B, nous sommes partie du principe que les participants avaient pour point commun d'être tous des enseignants de l'enseignement secondaire ordinaire en FW-B et avaient pour différences leur genre, leur nombre d'année(s) d'ancienneté dans l'enseignement, leur vécu face au redoublement et la discipline dans laquelle ils/elles enseignent.

Ces diverses informations à leur sujet ont été récoltées au moyen d'un questionnaire contextuel. Le choix de ce « large » échantillon résulte du fait que, ultérieurement, lors de l'analyse de nos résultats, nous étudierons ceux-ci en regard de ces variables sociodémographiques afin d'observer dans quelle mesure ces dernières peuvent également exercer une influence sur le jugement évaluatif des enseignants.

Dans les faits, notre étude expérimentale a été menée auprès d'un échantillon de 140 enseignants dont voici le détail de la composition.

<i>Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes</i>																													
SEXE (a)		ANCIENNETÉ (b)					VÉCU PAR RAPPORT REDOUBLEMENT (c)		DISCIPLINE (d)																				
	F	H	A	B	C	D	E	A	B	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Groupe 1 (N= 35)	26	9	15	7	3	2	8	16	19	9	9	2	1	6		1		1	3	1					1	1			
Groupe 2 (N=35)	24	11	21	5	4	1	4	15	20	8	10	1	4	4	1		1		2		1	1		1				1	
Groupe 3 (N=35)	25	10	13	5	6	3	8	18	17	7	8	2	2	6		2	1	1	2				1					2	1
Groupe 4 (N=35)	29	6	16	4	2	3	10	15	20	14	6		1	7	1	1						1			1	1	1		

Tableau 2. *Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes*

Note a : F= Femme, H= Homme

Note b : A = 0an < x ≤ 5ans, B = 5ans < x ≤ 10ans, C = 10ans < x ≤ 15ans, D = 15ans < x ≤ 20ans, E = x > 20ans

Note c : A = connu le redoublement, B= pas connu le redoublement

Note d : A = français, B = mathématiques, C = histoire, D = sciences (chimie, biologie, physique), E = langues modernes (anglais, néerlandais, allemand, espagnol, français langue étrangère), F = langues anciennes (latin), G = religion (catholique, protestante etc.), H = philosophie et citoyenneté, I = cours artistiques (art plastique, éducation musicale), J = éducation physique, K = agriculture, L = mécanique, M = hôtellerie, N = horticulture, O = informatique, P = économie, Q = sciences économiques, R = sciences humaines, S = sciences sociales, T= agencement intérieur.

Tout d'abord, concernant le genre de nos participants, le tableau ci-dessus nous montre que notre échantillon se compose majoritairement d'enseignants de sexe féminin (104) et ce, à raison de 74,29%. Les enseignants de sexe masculin, quant à eux, sont au nombre de 36 et représentent donc 25,71% de notre échantillon.

Ensuite, au niveau de la discipline d'enseignement, de manière générale sur 140 participants, nous avons 38 professeurs de français (27,14%) et 102 professeurs (72,86%) exerçant dans d'autres disciplines. Plus précisément, dans ces 102 professeurs, nous comptons 33 professeurs de mathématiques (23,57%), 23 professeurs de langues modernes (16,43%), 8 professeurs de sciences (5,71%), 7 professeurs d'éducation physique (5%), 5 professeurs d'histoire (3,57%), 4 professeurs de religion (2,86%), 3 professeurs de sciences sociales (2,14%), 2 professeurs de latin (1,43%), 2 professeurs de philosophie et citoyenneté (1,43%), 2 professeurs de cours artistiques (1,43%), 2 professeurs d'hôtellerie (1,43%), 2 professeurs d'informatique (1,43%), 2 professeurs d'économie (1,43%), 1 professeur d'horticulture (0,71%), 1 professeur de sciences économiques (0,71%), 1 professeur de sciences humaines (0,71%), 1 professeur d'agencement intérieur (0,71%), 1 professeur d'agriculture (0,71%), et 1 professeur de mécanique (0,71%).

Au niveau de l'ancienneté des participants dans l'enseignement, nous obtenons les résultats suivants : 65 participants (46,43%) ont entre 0 et 5 ans d'ancienneté, 21 participants (15%) ont entre 5 et 10 ans d'ancienneté, 15 participants (10,71%) ont entre 10 et 15 ans d'ancienneté, 9 participants (6,43%) ont entre 15 et 20 ans d'ancienneté, et enfin 30 participants (21,43%) ont plus de 20 ans d'ancienneté. La majorité de notre échantillon est donc composée d'enseignants que l'on pourrait qualifier de novices.

Pour terminer, concernant le fait d'avoir connu ou non le redoublement durant leur parcours scolaire, 64 participants sur 140 ont connu le redoublement, soit 45,71% de notre échantillon, tandis que 76 participants (54,29%) n'ont pas connu le redoublement durant leur parcours scolaire. Ainsi, nous comptabilisons dans notre échantillon un peu plus d'enseignants qui n'ont pas connu le redoublement durant leur parcours que d'enseignants l'ayant connu.

4.4 Procédure de recrutement

De début janvier à fin mars, nous avons procédé au recrutement de nos participants. Pour cela, nous avons posté une annonce sur les réseaux sociaux Facebook et Instagram (annexe n°11, p.111) dans laquelle étaient mentionnés l'objectif de la recherche, sa durée ainsi que les conditions pour y participer.

Par ailleurs, dans cette annonce, nous invitons les participants volontaires et répondant aux différents critères à commenter celle-ci afin que nous puissions par la suite les contacter en privé et leur communiquer, via la messagerie Zimbra, les documents qu'ils avaient pour consigne de lire et de signer (le formulaire de consentement en annexe n°15, p.140-142 et le formulaire d'information en annexe n°16, p.143-146). Nous leur avons également transmis par mail les deux « tâches » qu'ils ont dû réaliser selon un ordre bien précis : d'abord la rédaction à corriger (avec ou non le stéréotype et avec ou non une grille d'évaluation et le guide de correction) ensuite le questionnaire contextuel à compléter.

Les deux « tâches » ont donc été réalisées par les participants selon leur meilleure convenance, puisque le tout s'est déroulé à distance, sans contrôle de notre part.

De plus, pour optimiser le recrutement de nos participants, nous avons envoyé un mail à l'intention de la direction de 32 écoles de l'enseignement secondaire, pour lesquelles l'adresse mail était disponible sur internet, afin que celle-ci relaie notre annonce auprès des enseignants de leur établissement (annexe n°12, p.111).

4.5 Laboratoire cognitif

Dans ce laboratoire cognitif, nous avons fourni à 8 enseignants (1 professeur de français de l'enseignement secondaire et 7 enseignants du primaire) la rédaction que nous programmions d'utiliser dans le cadre de ce mémoire (annexes n°1 et n°2, p.92-93) ainsi que la grille d'évaluation créée en regard de celle-ci (annexe n°3, p.94-95). Sur base de cette rédaction, nous leur avons dans un premier temps demandé d'attribuer une note en se référant à la grille d'évaluation, et cela dans le but de vérifier le niveau de notre rédaction, à savoir : le fait d'être une rédaction d'un niveau satisfaisant (6,67/10).

Dans un deuxième temps, nous leur avons demandé d'analyser en détails la grille d'évaluation en se positionnant sur chacun de ses critères, ainsi que sur le système de notation (barèmes de points). Pour les guider dans leur analyse, nous leur avons fourni différentes questions auxquelles ils devaient répondre :

- 1) Les critères utilisés, sont-ils compréhensibles ou comportent-ils une (des) ambiguïté(s) ?
- 2) Les mots utilisés pour formuler chacun des critères, sont-ils opportuns ou faudrait-il en remplacer certains par d'autres plus appropriés ?
- 3) Le système de notation attribué à chaque critère, est-il approprié ? (Si pas, comment changeriez-vous cela ?)
- 4) Le système de notation, est-il compréhensible ?
- 5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

Ce laboratoire cognitif avait donc pour intérêt, d'une part, d'apporter d'éventuelles améliorations à la grille d'évaluation et, d'autre part, de s'assurer que les 8 enseignants interrogés attribuent bien une note proche d'un niveau satisfaisant à cette production écrite.

4.5.1 Résultats du laboratoire cognitif

Dans cette partie nous passerons brièvement en revue les résultats de ce laboratoire cognitif qui nous ont amenée à réaliser des modifications dans notre grille d'évaluation et dans notre rédaction. Un compte rendu plus détaillé est disponible en annexe (annexe n°13, p.113-117).

Tout d'abord, au niveau du premier **critère relatif à l'impression globale**, ce dernier a fait l'objet de nombreux commentaires. La plupart des participants ont mis en évidence le caractère subjectif de ce critère et donc la difficulté qu'ils ont éprouvée pour attribuer une note à ce dernier et, certains ont suggéré de déplacer ce critère en fin de grille afin qu'ils puissent l'évaluer de manière plus objective.

Néanmoins, partant initialement du principe que, pour observer des variations entre les jugements évaluatifs des participants évaluant à l'aide d'une grille d'évaluation, il est important qu'une forme de liberté soit laissée dans la notation, le critère relatif à *l'impression globale* nous semblait donc être idéal. De plus, concernant sa position en début de grille, nous avons fait ce choix car notre objectif au travers de ce critère est d'avoir l'impression globale et spontanée de l'évaluateur à la suite de la lecture de la rédaction. Si nous le positionnons à la fin de la grille d'évaluation, l'impression globale risque de subir l'influence des autres critères et donc de ne plus être spontanée. Toutefois, sur base des résultats obtenus lors du laboratoire cognitif, nous constatons la présence de variations inter-codeurs au niveau de plusieurs critères : seuls les niveaux de maîtrise « présence de la thèse », « présence d'une conclusion » de l'indicateur 2.1 et l'indicateur 4.3 ont obtenu une note identique de la part des évaluateurs. Dès lors, étant donné les raisons qui nous ont motivée à intégrer le critère *impression globale* et compte tenu des résultats précédemment cités, il nous semble que ce critère, problématique pour la plupart des enseignants interrogés, peut être supprimé.

Ensuite, dans un souci de **clarté**, sur base des commentaires des participants, nous avons décidé de remplacer certains mots par d'autres : dans les niveaux de maîtrise de l'indicateur 2.3 le mot « avis » a été remplacé par le mot « argument », dans les indicateurs 4.1, 4.2 et 4.3 le mot « correction » a été remplacé par le mot « maîtrise » tandis que dans la consigne générale de rédaction le mot « introduction » a été remplacé par le mot « thèse ».

Par ailleurs, la plupart des participants ont soulevé le caractère imprécis de certains éléments dans plusieurs critères de cette grille, et voici ce que certains ont spécifiquement mis en évidence : préciser le mot thèse (3/8), préciser ce qu'on entend par « arguments recevables » (2/8) (à partir de quand un argument peut-il être considéré comme recevable ? Sur base de quels critères ?), préciser si les contre-arguments sont acceptés ou non (1/8), préciser ce qu'on entend par « correction du vocabulaire » (1/8). Par conséquent, puisque cette grille d'évaluation est amenée à être utilisée par des enseignants de toutes disciplines, il nous a semblé important de tenir compte de ces diverses remarques et de l'aménager en fonction. C'est pourquoi, pour nous assurer de la bonne compréhension et utilisation de cette grille d'évaluation, nous avons joint à cette dernière un guide de correction qui complète et précise chacun des critères qui y sont présents (annexe n°7, p.100-101).

En outre, tout en réalisant notre guide de correction, pour plus de précision, nous avons décidé de scinder l'indicateur 2.2 en deux parties : a) et b) (annexe n°7, p.101).comme cela est suggéré dans la grille de correction disponible sur le portail de l'enseignement en FW-B dans la famille de tâches « convaincre » destinée aux élèves du 2^e degré (annexe n°10, p.107-110).

Au niveau de système de notation, de manière générale, il a été mis en évidence que la pondération était trop catégorique et ne laissait pas place à un continuum plus large des points. Ainsi, voici les modifications qui ont été appliquées :

Premièrement, concernant les niveaux de maîtrise de l'indicateur 2.2, la plupart des participants jugeant que deux arguments sur trois étaient recevables dans cette rédaction, nous avons décidé de supprimer le niveau de maîtrise « arguments non recevables » et avons gardé uniquement celui « arguments recevables » auquel nous avons attribué la pondération suivante : 0-1-2-3. Ainsi, les participants sont en mesure de choisir si aucun, un, deux ou tous les arguments sont recevables (annexe n°13, p.115).

Deuxièmement, plusieurs participants (3/8) nous ont exposé la difficulté qu'ils avaient ressentie pour attribuer une note à l'indicateur 4.1. Dès lors, sur base des recommandations de ces derniers, nous avons choisi d'évaluer la maîtrise orthographique en s'inspirant de la méthode utilisée pour les indicateurs 4.2 et 4.3, à savoir parler en termes de nombre d'erreurs.

Troisièmement, au niveau de l'indicateur 4.2, certains participants jugeant que beaucoup de points étaient accordés à cet indicateur et que la transition entre le 5 et le 0 était trop radicale, nous avons fait le choix d'attribuer 3 points à l'élève qui commet entre 0 et 2 erreurs de ponctuation et de syntaxe, et 0 point à l'étudiant qui commet plus de 2 erreurs.

Quatrièmement, si nous tenons compte de toutes ces modifications apportées, nous arrivons à une évaluation notée sur 38. Or, certains évaluateurs n'ayant pas de grille d'évaluation pour appuyer leur jugement évaluatif, il nous semble important de ramener l'évaluation de cette copie sur 40. C'est pourquoi nous avons décidé de changer la pondération du niveau de maîtrise associé à l'indicateur 2.3 en passant d'une pondération de 2 à une pondération de 4. En effet, nous avons estimé que ce critère revêtait une grande importance compte tenu de l'objectif de l'évaluation et qu'il méritait donc qu'on y accorde une plus grande pondération dans la note finale.

Cinquièmement, au niveau de la note attribuée par chacun des participants, nous observons de grands écarts entre évaluateurs (tableau n°1, p.147) : les notes varient de 26/50 à 39/50 alors qu'il s'agissait pourtant de la même copie et que tous ont reçu une grille d'évaluation pour émettre leur jugement évaluatif.

De plus, nous constatons qu'en moyenne, les évaluateurs ont attribué une note de 34,75/40 à cette rédaction ce qui correspond à une note de 6,95/10, et donc à une copie d'un niveau satisfaisant (Bureau du Secrétaire général des Écoles européennes. s.d), ce qui rejoint la note que nous avons initialement attribuée, à l'aide de la première version de la grille d'évaluation (annexe n°9, p.103-106), à savoir celle de 6,67/10.

Enfin, certains participants (3/8) ne comprenant pas toujours l'utilité des cases à cocher, nous les avons laissées uniquement lorsque cela était pertinent, c'est-à-dire lorsqu'il fallait choisir entre différents niveaux de maîtrise, comme c'est le cas pour les niveaux de maîtrise relatifs aux indicateurs 2.2, 2.3, 3.3, 4.2, et 4.3.

Pour conclure, en mettant en application toutes les modifications précédemment citées, nous obtenons une 2^e version de la rédaction (annexes n°4 et n°5, p96-97), ainsi qu'une 2^e version de la grille d'évaluation (annexe n°6, p.98-99) qui nous amène à revoir la note que nous avons initialement attribuée à cette copie. En effet, en mettant en application cette nouvelle grille, nous attribuons à la rédaction une note de 5,75/10. Nous passons donc d'une copie initialement évaluée comme étant d'un niveau satisfaisant (6,67/10) à une copie d'un niveau suffisant (5,75) (Bureau du Secrétaire général des Écoles européennes. s.d) ⁵.

4.6 Aspects éthiques et codage

Début novembre, la méthodologie de la présente recherche étant définie, nous avons pu soumettre au comité éthique l'ensemble du dossier (N°4721) pour lequel nous avons directement reçu un avis favorable. Dans ce dernier, nous avons été particulièrement attentive au codage des données. En effet, étant amenée à récolter des informations permettant d'identifier les participants (nom et prénom), dans un souci de confidentialité, nous avons attribué à chaque participant un numéro d'identification (ex : 01= nom du premier participant).

⁵ Les instruments finaux, à savoir la 2^e version de la rédaction, la 2^e version de la grille d'évaluation accompagnée d'un guide de correction et le questionnaire contextuel sont disponibles dans leur intégralité dans les annexes de ce travail.

Concernant le codage des autres variables indépendantes, celui-ci s'est opéré de la manière suivante : 0 = copie promu, 1 = copie redoublant, 0 = pas de grille d'évaluation, 1 = grille d'évaluation, 0 = femme et 1 = homme, 0 = jamais connu le redoublement durant son parcours scolaire et 1 = déjà connu le redoublement durant son parcours scolaire, 0 si enseignant dans une autre discipline que celle du français et 1 si enseignant dans la discipline du français. De plus, afin de garder les variables dans leur modalité la plus précise et d'éviter des effets de regroupements inadaptés, nous avons codé les réponses relatives à l'ancienneté des participants en termes d'années. Dans la même optique, pour ce qui est de l'encodage des points octroyés à chacune des copies, nous avons encodé la note totale, attribuée aux copies sans grille d'évaluation, telle qu'elle nous a été donnée par chacun des participants. Concernant la note totale attribuée aux copies à l'aide d'une grille d'évaluation, avant de la coder, nous avons vérifié que la somme des points octroyés par chacun des participants, par copie, soit correcte. Enfin, quant aux points attribués au sein même de la grille d'évaluation, nous avons effectué la somme des points au niveau de chacun des blocs de critères de manière à n'obtenir qu'une seule note pour chacun des blocs de critères (/15 pour le B1, /13 pour le B2, /12 pour le B3). L'ensemble de ces informations sont stockées, en version informatisée, dans un fichier Excel.

Quant à l'ensemble des documents qui nous ont été renvoyés par les participants (rédaction, grille d'évaluation, questionnaire contextuel, formulaire d'information et formulaire de consentement) ils sont conservés en format électronique sur un disque dur indépendant et protégés par un mot de passe pour une durée de 2 ans. Après cette période, les données seront détruites. Seuls le chercheur responsable, la promotrice et le doctorant associés à la recherche auront accès à toutes ces données qui permettront par la suite la réalisation des analyses statistiques.

Ensuite, ayant reçu fin novembre un avis favorable de la part du comité d'éthique, nous avons pu mettre à l'essai notre grille d'évaluation au travers d'un laboratoire cognitif mené auprès de 8 enseignants début décembre. C'est seulement début janvier, après avoir finalisé notre méthodologie et avoir conçu nos instruments de mesure définitifs, que nous avons pu débiter la procédure de recrutement de nos participants.

Afin d'éviter de biaiser les résultats de cette recherche, on ne peut signaler de manière explicite l'objectif de cette dernière auprès des participants, à savoir mesurer l'impact de l'ancrage du stéréotype relatif au redoublement sur le jugement évaluatif d'un enseignant. Dès lors, nous avons décidé de formuler cet objectif de manière plus large, en indiquant le thème principal de la recherche. Plus précisément, voici comment l'objectif a été formulé auprès des participants : l'objectif de cette étude sera de mesurer les facteurs qui interviennent dans l'évaluation d'une production écrite en vue d'améliorer les pratiques évaluatives.

Ainsi, cet objectif étant véritable, les participants n'ont à aucun moment été induits en erreur sur les buts et procédures de ce mémoire. Par conséquent, aucun consentement supplémentaire n'a été requis à la fin de la participation à l'étude.

Après avoir présenté aux participants la recherche au travers du formulaire d'information (annexe n°16, p.143-146) ainsi que le caractère non-obligatoire et purement volontaire de leur participation, nous leur avons demandé de signer un formulaire de consentement (annexe n°15, p.140-142) avant de leur soumettre les deux « tâches » à réaliser. Ce dernier formulaire consiste en un consentement écrit destiné à des personnes majeures stipulant que les données individuelles issues de la participation à l'étude ne seront pas transmises à des tiers.

4.7 Les analyses statistiques

Le traitement et l'analyse des données ont été opérés en plusieurs étapes. Mais avant de nous attarder sur chacune d'entre elles, il nous paraît important de rappeler que toutes les opérations décrites ci-après ont été réalisées sur un nombre égal de sujets dans chacun des groupes (groupe 1 : 35 sujets, groupe 2 : 35 sujets, groupe 3 : 35 sujets, groupe 4 : 35 sujets). Nous avons fait le choix d'avoir des échantillons de taille égale afin de minimiser le risque que l'effet principal d'un facteur soit contaminé par l'effet principal d'un autre facteur. En effet, plus les différences d'effectifs sont importantes, plus le risque de mauvaise(s) interprétation(s) avec une analyse de variance est grand (Quertemont, 2021).

4.7.1 Hypothèse n°1 et hypothèse n°3

Pour tester les première et la troisième hypothèses, nous avons décidé de réaliser trois types d'analyses statistiques : une analyse des moyennes et des écarts-types, une analyse des ampleurs de l'effet, ainsi qu'une analyse de la variance simple (ANOVA).

Tout d'abord, avant de nous lancer dans la réalisation de ces trois analyses nous avons vérifié la significativité du modèle (ANOVA) calculé au niveau de la note totale ainsi qu'au niveau de la note attribuée à chacun des blocs de critères.

Ensuite, pour tester les première et la troisième hypothèses nous avons procédé à une analyse des **moyennes** et des **écarts-types**⁶. Plus particulièrement, pour la première hypothèse, les moyennes et les écarts-types ont été calculés sur base des notes (notes totales et notes par blocs de critères) attribuées aux « copies promus » ainsi qu'aux « copies redoublants ».

⁶ Étant donné que nos données ne représentent qu'un échantillon de la population nous avons décidé de travailler avec un écart-type standard

Au niveau de la troisième hypothèse, le calcul des moyennes et des écarts-types ont été réalisés sur base des notes totales attribuées à chacun des quatre groupes de copies (groupe 1 = « copies redoublants avec grille » ; groupe 2 = « copies promus avec grille » ; groupe 3 = « copies redoublants sans grille » ; groupe 4 = « copies promus sans grille »).

S'en est suivie une analyse des **ampleurs de l'effet**. De manière générale, l'ampleur de l'effet permet de mesurer la force de l'effet observé d'une variable sur une autre et ce, indépendamment de la taille de l'échantillon (Goudeseune, 2018). Dans le cadre de notre étude, il s'agit de mesurer la taille de l'effet observé de la variable indépendante : « redoublant » pour la première hypothèse et « groupe de copies » pour la troisième hypothèse, sur la variable dépendante « note ». Plus particulièrement, nous avons utilisé l'ampleur de l'effet du *d* de Cohen (Borenstein et al., 2009) qui est une grandeur statistique descriptive qui permet de caractériser la grandeur d'un effet associé dans une population en regard d'une hypothèse nulle. En d'autres termes, le *d* de Cohen permet de caractériser la force de cet effet hypothétique (Goudeseune, 2018). Par ailleurs, dans le cadre de notre étude, étant face à des groupes indépendants, nous avons calculé les ampleurs de l'effet sur base des différences de moyennes standardisées entre les groupes, pris deux par deux, en mettant en application la formule suivante :

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}} \quad \text{et} \quad S_{within} = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}}$$

Au niveau de la première hypothèse, nous avons d'abord réalisé cette opération sur base de la note totale obtenue au niveau de chacune des copies selon qu'il s'agit d'une « copie redoublant » ou non. Ensuite, nous nous sommes centrée uniquement sur la moitié de l'effectif, à savoir les copies accompagnées d'une grille d'évaluation et avons calculé les ampleurs de l'effet pour chacun des trois blocs de critères. Que ce soit pour le calcul de l'ampleur de l'effet sur base de la note totale ou sur base des trois blocs de critères, le groupe expérimental correspond aux « copies redoublants » et le groupe contrôle aux « copies promus ».

Pour la troisième hypothèse, nous avons centré le calcul des ampleurs de l'effet uniquement sur la note totale. En effet, dans le cadre de cette hypothèse nous nous intéressons au fait de savoir si, de manière générale, l'influence qu'exerce la présence de la grille d'évaluation au niveau de la correction des « copies redoublants » est de même ordre que celle constatée au niveau des « copies promus ». Au total six ampleurs de l'effet ont été calculées, mais seules celles qui nous permettent de répondre à cette hypothèse ont été analysées.

Il s'agit de l'ampleur de l'effet qui concerne la comparaison du groupe de « copies redoublants avec grille » (GC) par rapport au groupe de « copies redoublants sans grille » (GE), et de l'ampleur de l'effet portant sur la comparaison du groupe de « copies promus avec grille » (GC) par rapport au groupe de « copies promus sans grille » (GE).

Toutefois, l'ampleur de l'effet du d de Cohen à lui seul, n'est pas en mesure de porter un jugement sur le caractère significatif de l'écart entre les moyennes. Pour démontrer l'existence de cet effet au niveau de la population de référence, un test statistique doit être réalisé (Goudeseune, 2018). C'est ainsi, qu'en vue d'identifier si l'effet étudié est significatif ou non pour les variables indépendantes considérées, nous avons clôturé nos analyses statistiques quant à ces deux hypothèses par deux **analyses de la variance simple** (ANOVA) réalisées à l'aide du logiciel statistique SAS. Plus précisément, dans le cas que nous traitons, la variable dépendante (ou quantitative) correspond au jugement évaluatif des évaluateurs qui est représenté par les notes (les notes totales et les notes attribuées à chacun des blocs de critères) attribuées par les participants à la rédaction tandis que les variables indépendantes sont au nombre de six. Nous retrouvons dans ces six variables indépendantes, deux variables catégorielles dichotomiques (redoublants/promus, avec grille/sans grille), ainsi que quatre variables sociodémographiques dont 3 variables catégorielles dichotomiques (homme/femme, connu le redoublement/pas connu le redoublement, professeur de français/pas professeur de français) et une variable métrique (ancienneté des participants).

Cependant, notons qu'un effet statistiquement significatif ne témoigne pas forcément de l'existence d'un effet important du point de vue théorique ou pratique. En effet, il est possible qu'un test se révèle être statistiquement significatif et pourtant négligeable d'un point de vue pratique. C'est pourquoi, il s'avère être nécessaire que les tests statistiques soient accompagnés d'une mesure de la taille de l'effet (Quertemont, 2021).

4.7.2 Hypothèse n°2

Pour vérifier cette deuxième hypothèse, nous avons d'une part calculé et comparé, de manière descriptive, la **variance** des notes attribuées aux copies évaluées à l'aide d'une grille d'évaluation à celles des copies évaluées sans grille d'évaluation, et d'autre part, nous nous sommes intéressée à calculer le rapport de ces deux variances.

4.7.3 Hypothèses relatives aux quatre variables sociodémographiques

Nous terminerons cette section relative aux analyses statistiques avec les analyses réalisées en regard des 4 hypothèses relatives aux variables sociodémographiques . Plus précisément, pour mettre à l'épreuve ces variables nous avons réalisé une **régression linéaire** de la note totale selon la présence du stéréotype « redoublant », de la grille d'évaluation, de l'ancienneté des évaluateurs, du fait qu'ils aient connu ou non le redoublement, de leur genre et de s'il s'agit de professeurs de français ou non. Les valeurs obtenues, suite à la réalisation de cette régression linéaire, ont été calculées à l'aide du logiciel statistique SAS. Par ailleurs, portant un intérêt particulier pour la variable « connu le redoublement » nous avons décidé de poursuivre l'analyse de cette dernière avec une analyse des **moyennes**, des **écarts-types**, ainsi que de l'**ampleur de l'effet**.

5. Résultats et interprétations

Dans cette partie relative à la présentation des résultats, nous mettons à l'épreuve, les unes à la suite des autres, nos trois hypothèses principales ainsi que nos hypothèses formulées en regard des quatre variables sociodémographiques. En d'autres termes, au travers de cette partie, nous cherchons tout d'abord à déterminer si l'activation subtile du stéréotype « redoublant » au sein de copies a une incidence sur le jugement évaluatif des enseignants et, si cette incidence varie ou non en fonction de la présence d'une grille d'évaluation. Ensuite, nous cherchons à déterminer l'influence que peuvent avoir une série de variables indépendantes relatives à l'ancienneté professionnelle des participants, leur genre, leur discipline d'enseignement, et s'ils ont déjà vécu ou non le redoublement durant leur parcours scolaire.

Pour analyser nos données, nous avons utilisé le logiciel SAS. Cet outil permet d'analyser les données récoltées sous forme de traitements statistiques.

Avant de tester nos hypothèses, nous avons vérifié la normalité des données obtenues au niveau de la note totale attribuée à la rédaction pour chacun des 4 groupes de copies, ainsi qu'au niveau des notes attribuées à chacun des blocs de critères pour les copies « promus » et les copies « redoublants ».

Il existe plusieurs tests qui permettent de vérifier la normalité d'une distribution, dans le cas de notre étude, nous avons utilisé le test de Shapiro-Wilk dont les résultats sont présentés à la page suivante.

Conditions	Variables	Test	Statistique		Valeur de P	
			W		Pr<W	
	Copies promus sans grille (groupe 4)	Shapiro-Wilk	W	0,966872	Pr<W	0.3632
	Copies promus avec grille (groupe 2)	Shapiro-Wilk	W	0,960966	Pr<W	0.2444
	Copies redoublants sans grille (groupe 3)	Shapiro-Wilk	W	0.972903	Pr<W	0.5278
	Copies redoublants avec grille (groupe 1)	Shapiro-Wilk	W	0.976834	Pr<W	0.6543

Tableau 3. *Tableau récapitulatif reprenant les résultats au test de normalité pour les notes totales attribuées à chacun des groupes de copies*

Tout d'abord, au niveau de la note totale attribuée à la rédaction, pour chacun des groupes de copies, nous constatons que la probabilité de dépassement est supérieure au seuil d'erreur de 0,05, ce qui nous amène à tolérer la normalité des données au niveau de ces 4 groupes de copies. Nous pouvons donc considérer que les différents résultats obtenus auprès des participants concernant la note totale attribuée à chacun des groupes de copies sont distribués normalement.

Variables	B1			B2			B3		
	Test	Statistique (W)	Valeur de P (Pr<W)	Test	Statistique (W)	Valeur de P (Pr<W)	Test	Statistique (W)	Valeur de P (Pr<W)
Promu	Shapiro-Wilk	0,930476	0,0289	Shapiro-Wilk	0,952769	0,1378	Shapiro-Wilk	0,939504	0,0540
Redoublant	Shapiro-Wilk	0,948197	0,0997	Shapiro-Wilk	0,937956	0.0485	Shapiro-Wilk	0,942722	0,0678

Tableau 4. *Tableau récapitulatif reprenant les résultats au test de normalité pour les notes attribuées à chacun des blocs de critères selon qu'il s'agit d'une « copie redoublant » ou d'une « copie promu »*

Ensuite, concernant les notes attribuées à chacun des blocs de critères, nous pouvons constater que les probabilités de dépassement sont toutes supérieures au seuil alpha de 0,05 sauf pour le premier bloc de critères des « copies promus » (0,0289) ainsi que pour le deuxième bloc de critères des « copies redoublants » (0,0485). Ce qui nous amène à tolérer la normalité des données pour le premier bloc de critères des « copies redoublants », pour le deuxième bloc de critères des « copies promus » ainsi que pour le troisième bloc de critères, aussi bien pour les « copies promus » que pour les « copies redoublants ». Concernant le premier bloc de critères des « copies promus » et le deuxième bloc de critères des « copies redoublants », la petite taille de notre effectif pourrait être à l'origine de cette non normalité des données observées.

Ainsi, étant donné que seulement deux cas de non normalité ont été observés au niveau des sous-groupes, nous décidons de rester dans des analyses paramétriques (ANOVA) en vertu de leur robustesse et des nombreux résultats attestant de la normalité des distributions d'une majorité des sous-groupes.

Cela étant, passons maintenant au développement et à l'analyse de nos résultats en regard de nos différentes hypothèses.

5.1 Hypothèse n°1

Hypothèse n°1 : les copies présentées comme étant réalisées par un élève redoublant sont évaluées plus négativement comparativement aux copies réalisées par un élève promu.

Pour rappel, afin de se positionner de la manière la plus complète possible face à cette première hypothèse, nous allons réaliser trois types d'analyses dont nous présenterons et mettrons en relation les résultats. Le tout sera calculé, dans un premier temps, sur base de la note totale attribuée à chacune des copies, selon qu'il s'agisse d'une copie au sein de laquelle le stéréotype relatif au statut du redoublant est présent ou non. Dans un deuxième temps, il s'agira de réaliser ces analyses sur base de la note attribuée à chacun des blocs de critères de la grille d'évaluation selon qu'il s'agisse de « copies redoublants » ou « de copies promus ». Avant de nous lancer dans ces analyses, nous nous devons de vérifier la **significativité du modèle** (ANOVA) que nous allons notamment utiliser pour mettre à l'épreuve cette hypothèse.

L'ANOVA réalisée sur base de la moyenne des notes totales attribuées aux copies, atteste d'une valeur de F de 3,79 avec une probabilité de dépassement de 0,0009, inférieure au seuil alpha de 0,05, nous amenant à rejeter l'hypothèse nulle (tableau n°5, p.149). Nous pouvons donc envisager l'hypothèse alternative selon laquelle la note totale attribuée à la rédaction serait influencée par certaines des variables intégrées dans cette analyse de la variance. Le modèle a donc du sens pour expliquer la variable dépendante qui est dans ce cas-ci la note totale. Par ailleurs, le coefficient de détermination du modèle s'élevant à 0,16 il met en évidence que les variables indépendantes intégrées dans le modèle n'expliquent que 16% de la variance de la variable dépendante, à savoir la variable « note totale » (tableau n°5, p.149). Signifiant ainsi qu'environ 84% de la variance de la note totale peuvent être expliqués par d'autres variables qui ne sont pas prises en compte dans le modèle présenté ici.

Cela étant, nous pourrions être encline à nous demander si cette significativité du modèle se retrouve également au niveau de l'analyse de la variance (ANOVA) réalisée pour chacun des trois blocs de critères. Malheureusement, la probabilité de dépassement observée pour chacun des blocs de critères étant supérieure au seuil alpha 0,05 (B1p=0,1434 ; B2p=0,1829, B3p=1702) cela nous amène à tolérer l'hypothèse nulle selon laquelle les notes attribuées à chacun des blocs de critères ne seraient pas influencées par les variables intégrées dans le modèle.

Par ailleurs, le coefficient de détermination pour les trois blocs de critères s'élevant à environ 11% ($B_1=0,12$; $B_2=0,11$; $B_3=0,11$), la part de variance expliquée par le modèle est relativement faible (tableau n°6, p.150). Cela signifie que les variables indépendantes du modèle n'expliquent qu'environ 11% de la variance observée au niveau des notes attribuées à chacun des blocs de critères et que plus ou moins 88% de cette variance seraient expliqués par d'autres variables non intégrées dans le modèle.

Passons à présent à l'application de ces trois analyses sur base de la note totale attribuée à chacune des copies.

Redoublant	N	Note totale attribuée à la rédaction	
		Moyenne (écart-type)	Ampleur de l'effet (1) vs (2)
Oui (1)	70	24,42 (5,47)	-0,33
Non (2)	70	26,33 (6,11)	

Tableau 7. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée à la rédaction selon la présence ou non du stéréotype relatif au statut redoublant

	DL	Valeur de F	PR>F
Redoublant	1	3,45	0,0654
Grille	1	2,76	0,0990
Redoublant * grille	1	0,05	0,8320
Ancienneté	1	5,96	0,0160
Connu le redoublement	1	7,73	0,0062
Genre	1	0,51	0,4781
Français	1	4,25	0,0411

Tableau 8. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale

En nous penchant tout d'abord sur l'analyse des **moyennes**, nous pouvons constater que, dans notre échantillon, les copies accompagnées du stéréotype présentent, en moyenne, une note totale moins élevée (24,42/40) comparativement aux mêmes copies, mais au sein desquelles le stéréotype n'est pas présent (26,33/40). Cette différence s'élève à 1,91 point en faveur des « copies promus ». Le calcul de **l'ampleur de l'effet** va également dans ce sens avec pour résultat une ampleur de l'effet de -0,33 attestant que, dans notre échantillon, les copies redoublants (GE) sont, en moyenne, au niveau de la note totale, moins bien évaluées que les « copies promus » (GC).

Ensuite, **l'analyse de la variance** nous indique un effet non significatif de la variable « redoublant », nous obtenons une valeur de F de 3,45 avec une probabilité de dépassement de <0,0654, légèrement supérieure au seuil alpha de 5% nous amenant à conserver l'hypothèse nulle. Autrement dit, nous n'avons pas suffisamment de preuves pour affirmer que l'effet du stéréotype sur le note totale traduit un phénomène réel dans la population de référence.

L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons affirmer que la note totale attribuée à la rédaction soit influencée de manière significative par la présence du stéréotype relatif au redoublement.

Nous poursuivrons plus amplement l'analyse de la variance des autres variables mises en évidence dans ce tableau dans la section relative à la troisième hypothèse, ainsi que dans la section relative aux variables sociodémographiques.

Dans un deuxième temps, et toujours en réponse à cette première hypothèse, nous allons nous pencher, à nouveau, sur une analyse des moyennes, des ampleurs de l'effet ainsi que de la variance simple mais cette fois-ci calculés sur base de la note attribuée à chacun des blocs de critères de la grille d'évaluation selon qu'il s'agisse d'une copie au sein de laquelle le stéréotype relatif au statut du redoublant est présent ou non. En d'autres termes, dans cette analyse nous nous intéresserons uniquement aux copies évaluées à l'aide d'une grille d'évaluation.

Variables Conditions	B1			B2			B3		
	N	μ (σ)	Ampleur de l'effet (1) vs (2)	N	μ (σ)	Ampleur de l'effet (1) vs (2)	N	μ (σ)	Ampleur de l'effet (1) vs (2)
Redoublant (1)	35	10,34 (2,85)	-0,11	35	8,86 (2,29)	0,31	35	6,69 (2,73)	-0,66
Promu (2)	35	10,66 (3,03)		35	8,26 (1,50)		35	8,29 (2,11)	

Tableau 9. Moyennes, écarts-types et ampleurs de l'effet des notes attribuées à chacun des blocs de critères de la grille d'évaluation selon la présence ou non du stéréotype relatif au statut redoublant

Variables Conditions	B1			B2			B3		
	DL	Valeur de F	PR > F	DL	Valeur de F	PR > F	DL	Valeur de F	PR > F
Redoublant	1	0,39	0,5361	1	1,31	0,2564	1	7,36	0,0086
Ancienneté	1	0,62	0,4344	1	0,16	0,6870	1	0,35	0,5537
Connu le redoublement	1	0,92	0,3418	1	2,76	0,1014	1	0,49	0,4871
Genre	1	0,04	0,8516	1	2,55	0,1153	1	0,00	0,9748
Français	1	6,40	0,0139	1	0,93	0,3395	1	0,02	0,8752

Tableau 10. Tableau récapitulatif reprenant les résultats des sommes de carrés de type III de l'analyse de variances simples pour chacun des blocs de critères

Tout d'abord, concernant le premier bloc de critères, les analyses attestent de résultats non significatifs. En effet, les **moyennes** mettent en évidence que les « copies redoublants » (10,34/15) sont légèrement moins bien notées que les « copies promus » (10,66/15) et ce, avec une différence de points s'élevant à 0,32 point en faveur des « copies promus ». Toutefois, le résultat obtenu suite au calcul de **l'ampleur de l'effet** démontre un effet de petite taille (-0,11) ne permettant pas d'attester d'une influence substantielle de l'ancrage du stéréotype « redoublant », au niveau du premier bloc de critères, sur le jugement évaluatif des enseignants du secondaire. De plus, lorsque nous poursuivons nos analyses avec celle de la **variance simple**, nous obtenons au niveau de la variable « redoublant » une valeur de F de 0,39 avec une probabilité de dépassement de $<0,5361$ largement supérieure au seuil alpha de 5% nous amenant à conserver l'hypothèse nulle. Ainsi, n'ayant pas suffisamment de preuves pour affirmer de l'effet réel du stéréotype sur la note du premier bloc de critères au niveau de la population de référence. L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons affirmer que la note attribuée au premier bloc de critères soit influencée de manière significative par la présence du stéréotype relatif au statut de redoublant sur la copie.

Ensuite, dans l'analyse des **moyennes** relatives au troisième bloc de critères, nous retrouvons un constat similaire à celui mis en évidence au niveau du premier bloc de critères avec une différence de 1,6 point en faveur des « copies promus »; soit une différence moyenne de points entre les deux groupes de copies plus importante que celle observée au niveau du premier bloc de critères. Le résultat du calcul de **l'ampleur de l'effet** va également dans ce sens avec une valeur de -0,66 signifiant qu'au niveau du troisième bloc de critères, les « copies redoublants » (GE) ont, en moyenne, une note moins élevée que les « copies promus » (GC). De plus, les résultats obtenus suite à **l'analyse de la variance simple** attestent d'une valeur de F de 7,36 avec une probabilité de dépassement de $<0,0086$ largement inférieure au seuil alpha de 5% nous amenant à rejeter l'hypothèse nulle. Ainsi, nous pouvons affirmer que les résultats que nous obtenons ne sont pas dues au hasard de l'échantillonnage, mais qu'il existe bien un effet réel, dans la population de référence, de la présence stéréotype sur la note attribuée au troisième bloc de critères. Avec un risque de 5%, l'hypothèse alternative, selon laquelle la note attribuée au troisième bloc de critères est influencée de manière significative par l'ancrage du stéréotype « redoublant » sur la copie, peut donc être envisagée. Plus précisément, dans ce cas-ci, il semblerait que les « copies redoublants » soient moins bien notées que les « copies promus ».

Enfin, en nous penchant sur le deuxième bloc de critères, contre toute attente nous constatons, qu'en **moyenne**, les « copies redoublants » ont une note légèrement plus élevée (8,86/13) que les « copies promus » (8,26/13) et ce, avec une différence de 0,6 point en faveur des « copies redoublants ». Le calcul de **l'ampleur de l'effet** rejoint également ce constat avec une taille de l'effet de 0,31 attestant que les « copies redoublants » (GE) sont mieux notées que les « copies promus » (GC). Toutefois, l'analyse de la variance révèle un résultat non significatif. Elle atteste d'une valeur de F de 1,31 avec une probabilité de dépassement de <0,2564 supérieure au seuil alpha de 5% nous amenant à conserver l'hypothèse nulle. En d'autres termes, nous n'avons pas suffisamment de preuves pour affirmer que l'effet du stéréotype sur la note attribuée au deuxième bloc de critères, traduit d'un phénomène réel dans la population de référence. L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons certifier que la présence du stéréotype « redoublant » exerce une influence significative sur la note attribuée au niveau du deuxième bloc de critères.

5.2 Hypothèse n°2

Hypothèse n°2 : la variance observée entre les notes attribuées aux copies corrigées à l'aide d'une grille d'évaluation sera moins importante que celle observée entre les notes des copies corrigées sans grille d'évaluation.

Conditions Variable	Pas grille			Grille		
	N	μ (σ)	σ^2	N	μ (σ)	σ^2
Note totale	70	24,76 (6,18)	38,23	70	25,99 (5,49)	30,16

Tableau 11. Calcul de la variance des notes totales selon la présence ou non d'une grille d'évaluation

En nous penchant sur le tableau ci-dessus, nous pouvons constater que, dans notre échantillon, la **variance** de la note totale attribuée aux copies non accompagnées d'une grille d'évaluation est plus élevée (38,23) que celle des copies accompagnées d'une grille d'évaluation (30,16). Plus précisément, lorsque nous effectuons le ratio de ces deux variances, nous obtenons la valeur de 1,27, signifiant qu'évaluer la rédaction sans grille d'évaluation amène à un variance de notes 1,27 fois plus élevée que lorsque la rédaction est évaluée à l'aide d'une grille d'évaluation. Ainsi, lorsque la correction de la rédaction est effectuée avec une grille d'évaluation, les notes totales attribuées par les correcteurs ont tendance à être moins dispersées les unes par rapport aux autres, plus proches de la moyenne et ce, comparativement aux notes totales attribuées à la rédaction sans grille d'évaluation.

5.3 Hypothèse n°3

Hypothèse 3 : Les enseignants évaluent de manière plus favorable une « copie redoublant » accompagnée d'une grille d'évaluation qu'une même copie mais non accompagnée d'une grille d'évaluation. Cette tendance s'observe également auprès des « copies promus », mais dans une moindre mesure.

Pour vérifier cette hypothèse, nous allons, dans un premier temps, comparer la moyenne des notes totales attribuées à chacun des groupes de copies.

Conditions Variable	Grille et redoublant (1)		Grille et promu (2)		Pas de grille et redoublant (3)		Pas de grille et promu (4)	
	N	μ (σ)	N	μ (σ)	N	μ (σ)	N	μ (σ)
Note totale	35	25,40 (6,17)	35	26,57 (4,74)	35	23,43 (4,56)	35	26,09 (7,29)

Tableau 12. Moyenne et écarts-types de la note totale pour chacune des 4 copies

Lorsque l'on met en relation la **moyenne** des notes attribuées aux « copies redoublants » accompagnées d'une grille d'évaluation (25,40/40) avec celle des notes attribuées aux « copies redoublants » non accompagnées de grille d'évaluation (23,43/40) nous pouvons constater une différence de points de 1,97 point en faveur des « copies redoublants » accompagnées d'une grille d'évaluation. Concernant les « copies promus », nous observons la même tendance mais dans une moindre mesure avec une différence s'élevant à 0,48 point en faveur des « copies promus » accompagnées d'une grille d'évaluation. Ainsi, de manière descriptive, nous relevons un effet d'interaction entre la variable relative à la grille d'évaluation et celle relative à la présence du stéréotype. En effet, dans le cadre de notre échantillon, il semblerait, qu'en moyenne, la présence de la grille d'évaluation a davantage bénéficié aux élèves redoublants et ce, comparativement aux élèves promus.

Dans un deuxième temps, nous allons procéder à une analyse des **ampleurs de l'effet**. Plus précisément, nous allons nous focaliser sur celles qui nous intéressent, à savoir les deux ampleurs des effets suivants :

- 1) L'ampleur de l'effet utilisée pour quantifier la relation (ou la différence) entre le groupe de « copies redoublants » accompagnées d'une grille d'évaluation (1) et le groupe de « copies redoublants » mais non accompagnées d'une grille d'évaluation (3)
- 2) La taille de l'effet utilisée pour quantifier la relation (ou la différence) entre le groupe de « copies promus » accompagnées d'une grille d'évaluation (2), et le groupe de copies au sein desquelles ni le stéréotype ni la grille d'évaluation ne sont présents (4).

Conditions	Grille et redoublant (1)		Grille et promu (2)		Pas de grille et redoublant (3)		Pas de grille et promu (4)		(1)vs(2)	(1)vs(3)	(1)vs(4)	(2)vs(3)	(2)vs(4)	(3)vs(4)
	N	μ (σ)	N	μ (σ)	N	μ (σ)	N	μ (σ)	Ampleur de l'effet					
Note totale	35	25,40 (6,17)	35	26,57 (4,74)	35	23,43 (4,56)	35	26,09 (7,29)	-0,21	-0,36	-0,10	-0,68	-0,08	-0,44

Tableau 13. Calcul de l'ampleur de l'effet entre les 4 groupes de copies en fonction de la note totale

Les résultats repris dans le tableau ci-dessus nous indiquent une ampleur de l'effet de -0,36 pour le groupe de copies (1) vs (3) signifiant que, dans le cadre de notre échantillon, les « copies redoublants » non accompagnées d'une grille d'évaluation (GE) sont, en moyenne, moins bien évaluées que les « copies redoublants » accompagnées d'une grille d'évaluation (GC). Quant à l'ampleur de l'effet (2) vs (4), il est de -0,08 attestant d'une différence ténue entre le groupe de « copies promus » accompagnées d'une grille d'évaluation (GC) et le groupe de « copies promus » non accompagnées d'une grille d'évaluation (GE).

Dans un troisième et dernier temps, reprenons le tableau relatif à **l'analyse de la variance** (ANOVA) présenté précédemment dans la section relative à la première hypothèse et tentons de poursuivre l'analyse de ce dernier.

	DL	Valeur de F	PR>F
Redoublant	1	3,45	0,0654
Grille	1	2,76	0,0990
Redoublant * grille	1	0,05	0,8320
Ancienneté	1	5,96	0,0160
Connu le redoublement	1	7,73	0,0062
Genre	1	0,51	0,4781
Français	1	4,25	0,0411

Tableau 8. Résultats des sommes de carrés de type III de l'analyse de variance simple. *su*
la note totale

Tout d'abord, comme nous l'avons mis en évidence dans la section relative à la première hypothèse, les résultats de l'analyse de la variance de la variable « redoublant » nous amènent à conserver l'hypothèse nulle ($0,05 < p = 0,0654$). Nous ne pouvons donc pas considérer que la note totale attribuée à la rédaction soit influencée de manière significative par la présence du stéréotype relatif au redoublement.

Ensuite, concernant l'influence de la présence ou non d'une grille d'évaluation sur la note totale attribuée à la rédaction, nous observons une valeur de F de 2,76 avec une probabilité de dépassement de $< 0,0990$ supérieure au seuil alpha (5%) nous amenant également à conserver l'hypothèse nulle.

Ainsi, nous n'avons pas suffisamment de preuves pour affirmer, au niveau de la population de référence, de l'effet réel de la présence de la grille d'évaluation sur la note totale attribuée aux copies. L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons affirmer que la présence de la grille d'évaluation exerce une influence significative sur la note totale attribuée à la rédaction.

Enfin, concernant l'effet d'interaction entre les variables dichotomiques « redoublant » et « grille » nous obtenons une valeur de F de 0,05 avec une probabilité de dépassement de <0,8320 largement supérieure au seuil alpha (5%) nous amenant à tolérer l'hypothèse nulle. L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons affirmer que l'effet de la présence du stéréotype « redoublant » sur la copie varie, de manière significative, si la copie est accompagnée ou non d'une grille d'évaluation.

5.4 Les 4 variables sociodémographiques

Les résultats quant aux trois hypothèses principales étant développés, nous allons nous intéresser à présent à l'effet que pourrait avoir chacune des variables indépendantes, relatives aux caractéristiques des participants, sur leur manière d'évaluer la rédaction. Plus précisément, dans cette section nous nous centrerons uniquement sur les variables sur lesquelles nous ne nous sommes pas encore positionnée, à savoir : le genre, la discipline d'enseignement, le nombre d'année(s) d'ancienneté dans l'enseignement et le fait d'avoir connu ou non le redoublement. Pour ce faire, nous réaliserons une analyse de l'influence de chacune de ces quatre variables indépendantes sur la note totale attribuée par les évaluateurs.

Effet	Estimation	Erreur standard	Valeur t	Pr > t
Intercept	25,20725	1,20516	20,916	< 2e-16
Redoublant	-1,95425	1,34635	-1,452	0,14901
Grille	1,36474	1,35115	1,010	0,31432
Ancienneté	0,10346	0,04238	2,442	0,01595
Connu le redoublement	-2,61730	0,94146	-2,780	0,00623
Genre	-0,76703	1,07818	-0,711	0,47808
Français	2,20474	1,06895	2,063	0,04112
Redoublant*grille	0,40319	1,89689	0,213	0,83200

Tableau 14. Régression linéaire de la note totale selon la présence du stéréotype « redoublant », de la grille d'évaluation, de l'ancienneté des évaluateurs, du fait qu'ils aient connu ou non le redoublement, de leur genre et de s'il s'agit ou non de professeurs de français

Commençons par nous centrer sur la variable relative à l'ancienneté des enseignants : pour chaque année supplémentaire d'ancienneté, on observe, en moyenne, une augmentation de 0,10 point sur la note totale attribuée à la rédaction. A première vue, cette différence de points par année d'ancienneté pourrait paraître minime, néanmoins au terme de 10 années d'expérience, nous pouvons nous attendre à une différence de l'ordre de 1 point. De plus, lorsque l'on s'intéresse aux résultats obtenus au travers de la **régression linéaire** de la note totale en fonction d'une série de variables indépendantes dont l'ancienneté des participants, on constate que cette différence de points s'avère statistiquement significative. En effet, pour la variable « ancienneté », la statistique t vaut 2,44 avec une probabilité de dépassement de $<0,01595$, inférieure au seuil alpha de 5% , ce qui nous amène à rejeter l'hypothèse nulle. Ainsi, nous pouvons affirmer qu'il existe bien un effet réel, dans la population de référence, de la présence stéréotype sur la note totale. L'inférence de cet effet étant prouvée au niveau de la population de référence, nous pouvons donc considérer que la note totale attribuée à la rédaction se voit influencée, de manière significative, par l'ancienneté des évaluateurs de sorte que, les enseignants plus expérimentés ont, en moyenne tendance, comparativement aux enseignants débutants, à attribuer plus de points à la rédaction, différence qui semble s'accroître au fur et à mesure que les années d'expérience s'accumulent.

Ensuite, si nous nous centrons sur la variable « connu le redoublement », nous constatons que pour tout changement de 1 unité sur la variable « connu le redoublement » on observe en moyenne, une diminution de 2,62 points sur la note totale. Autrement dit, il semblerait que, dans notre échantillon, lorsque les enseignants ont connu le redoublement durant leur parcours scolaire, ils ont tendance à moins bien évaluer les copies de 2,62 points et ce, comparativement aux enseignants qui, n'ont pas connu le redoublement durant leur parcours scolaire. Par ailleurs, nous constatons que pour la variable « connu le redoublement », la statistique t vaut $-2,78$ avec une probabilité de dépassement de $<0,00623$ inférieure au seuil alpha de 5%, ce qui nous amène à rejeter l'hypothèse nulle. L'inférence de cet effet étant prouvée au niveau de la population de référence, nous pouvons considérer que, dans notre échantillon, la note totale attribuée à la rédaction est influencée, de manière significative, par le fait que les enseignants ont connu ou non le redoublement durant leur parcours scolaire.

Cela étant, afin de nous positionner de manière plus précise face à l'effet de la variable « connu le redoublement » sur le jugement évaluatif, nous allons poursuivre nos analyses statistiques avec une **analyse des moyennes, des écarts-types et de l'ampleur de l'effet** de la note totale attribuée aux « copies redoublants » selon que l'évaluateur a connu ou non le redoublement, dont les résultats sont présentés dans le tableau suivant :

Variables	Condition	Redoublant		
		N	μ (σ)	Ampleur de l'effet (1) vs (2)
Note totale attribuée par des enseignants ayant connu le redoublement (1)		34	23,26 (4,51)	-0,41
Note totale attribuée par des enseignants n'ayant pas connu le redoublement (2)		36	25,50 (6,12)	

Tableau 15. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée aux « copies redoublants » selon que l'évaluateur a connu ou non le redoublement

Les résultats mis en évidence au sein de ce tableau rejoignent ceux précédemment développés. En effet, ils démontrent que les enseignants ayant déjà vécu le redoublement durant leur parcours scolaire attribuent, en moyenne, une note totale moins élevée (23,26/40) aux « copies redoublants » et ce, comparativement aux enseignants qui n'ont jamais connu le redoublement durant leur parcours scolaire (25,50/40). Par ailleurs, le résultat obtenu suite au calcul de l'ampleur de l'effet va également dans ce sens avec une taille de l'effet de -0,41 signifiant que les enseignants ayant vécu le redoublement durant leur parcours scolaire ont tendance à attribuer, en moyenne, aux « copies redoublants » une note totale inférieure, comparativement aux enseignants n'ayant pas connu le redoublement.

Concernant la variable relative au genre des participants, nous constatons que lorsque nous passons d'un participant de sexe féminin (=0) à un participant de sexe masculin (=1), on observe, en moyenne, une diminution de 0,77 point sur la note totale attribuée à la rédaction, soit une différence de points que l'on peut considérer comme minime. Autrement dit, étant donné ce faible effet, nous ne pouvons considérer que, dans la cadre de cet échantillon, le genre des enseignants exerce une influence substantielle sur leur manière d'évaluer la rédaction. Par ailleurs, il apparaît que, pour la variable « genre », la statistique t vaut -0,71 avec une probabilité de dépassement de <0,47808, supérieure au seuil alpha de 5%, ce qui nous amène à conserver l'hypothèse nulle.

Dès lors, nous n'avons pas suffisamment de preuves pour affirmer, au niveau de la population de référence, de l'effet réel du genre des enseignants sur la note totale attribuée aux copies. L'inférence de cet effet n'étant prouvée au niveau de la population de référence, nous ne pouvons affirmer que la note totale attribuée à la rédaction soit influencée par le « genre » des enseignants.

Enfin, la variable « français » témoigne que pour tout changement de 1 unité sur cette variable, on constate une augmentation de 2,20 points sur la note totale attribuée à la rédaction. En d'autres termes, lorsque les copies sont corrigées par des professeurs de français elles ont tendance à recevoir, en moyenne, 2,20 points de plus que les copies qui, quant à elles, sont évaluées par des enseignants qui exercent dans une(des) autre(s) discipline(s) que celle du français. Par ailleurs, nous constatons que, pour la variable « français », la statistique t vaut 2,06 avec une probabilité de dépassement de $<0,04112$, inférieure au seuil alpha de 5%, ce qui nous amène à rejeter l'hypothèse nulle. Ainsi, l'inférence de cet effet étant prouvée au niveau de la population de référence, nous pouvons considérer que la note attribuée à la rédaction est influencée, de manière significative, par le fait que les évaluateurs soient ou « non professeurs » de français.

6. Discussion

Passons à présent à la discussion de ce mémoire. A travers celle-ci, nous allons aborder notre question de recherche ainsi qu'examiner les différentes hypothèses que nous avons formulées et les discuter au regard des résultats que nous avons obtenus. Afin de comparer et nuancer nos résultats, nous établirons des liens entre ceux-ci, nos observations et la littérature préexistante concernant les différentes thématiques abordées dans ce mémoire.

6.1 Données sociodémographiques

Pour rappel, 140 enseignants ont participé à notre recherche expérimentale. Notons tout d'abord que les hommes ne représentent qu'un quart des participants de notre étude (25,71%) et ce, malgré le fait que l'annonce que nous avons postée sur les réseaux sociaux concernait les enseignants hommes et femmes de l'enseignement secondaire ordinaire en FW-B. Cette sous-représentation de participants de genre masculin dans notre échantillon est représentative de la réalité de terrain. En effet, les chiffres attestent qu'en 2021 dans l'enseignement secondaire ordinaire en FW-B, nous comptons 63,5% d'enseignants de sexe féminin et seulement 36,5% d'enseignants de sexe masculin (Fédération Wallonie-Bruxelles, s.d).

Ensuite, notre étude recense une majorité de participants ayant entre 0 et 5 années d'ancienneté dans l'enseignement (46,43%). Cette observation pourrait s'expliquer par le fait que notre annonce a été principalement partagée sur des groupes d'étudiants (« Master en Sciences de l'Éducation », « Étudiants de Liège », etc.) et ces groupes sont composés principalement d'une population que l'on pourrait qualifier de « jeune ». Il n'est donc pas surprenant d'observer qu'une majorité de nos participants est composée d'enseignants novices.

Notons également qu'au niveau de la discipline d'enseignant, notre échantillon est composé d'une majorité de professeurs de français et, dans une proportion plus faible, de professeurs de mathématiques (23,57%), de langues modernes (16,43%), de sciences (5,71%), d'éducation physique (5%), d'histoire (3,57%), de sciences sociales (2,14%), de religion (2,86%), de latin (1,43%), de philosophie et citoyenneté (1,43%), de professeurs de cours artistiques (1,43%), d'agriculture (0,71%), de mécanique (0,71%), d'hôtellerie (1,43%), d'informatique (1,43%), d'économie (1,43%), d'horticulture (0,71%), de sciences économiques (0,71%), de sciences humaines (0,71%), d'agencement intérieur (0,71%). Constat qui n'est pas étonnant étant donné que la tâche qui a été demandée aux participants relève de la discipline du français. Ainsi, nous pouvons supposer que les professeurs de français, étant plus à l'aise avec la tâche demandée ont participé plus volontiers à cette dernière, à l'inverse des enseignants exerçant dans d'autres disciplines, qui ont déclaré avoir une certaine appréhension concernant leur capacité à réaliser la tâche requise.

Enfin, concernant le fait d'avoir connu ou non le redoublement, notre échantillon démontrant une faible différence entre le nombre de participants ayant connu le redoublement (45,71%) et ceux ne l'ayant pas connu (54,29%) nous n'allons pas nous attarder davantage sur la différence de proportion entre ces deux groupes.

6.2 Discussion autour de nos différentes hypothèses

Hypothèse n°1 : les copies présentées comme étant réalisées par un élève redoublant sont évaluées plus négativement comparativement aux copies réalisées par un élève promu.

Pour commencer, en ce qui concerne notre première hypothèse, la comparaison des moyennes des deux groupes de copies et le calcul de l'ampleur de l'effet témoignent, dans notre échantillon, d'une influence de l'ancrage du stéréotype sur la **note totale** attribuée aux copies de sorte que les « copies redoublants » seraient, en moyenne, moins bien notées que les « copies promus ».

Si on s'en tient aux résultats obtenus suite à ces premières analyses, il semblerait que les erreurs deviennent plus des faiblesses lorsqu'elles sont commises par un élève présenté comme redoublant que lorsqu'elles sont présentées comme étant commises par un élève qui, à l'inverse, n'a pas connu le redoublement. Ces premiers résultats semblent rejoindre, en partie, ceux mis en évidence dans plusieurs études (Autin et al., 2019 ; Caverni et al., 1975, Dardenne, 2019 ; Lafontaine & Monseur, 2009 ; Mosconi, 1999 ; Sprietsma, 2013) qui attirent l'attention sur un phénomène d'inégalités de traitement en démontrant les attentes différenciées que peuvent avoir les enseignants à l'égard de leurs élèves selon certaines de leurs caractéristiques, comme leur genre, leur origine sociale, leurs résultats antérieurs, etc. Attentes qui, à leur tour, peuvent exercer une influence sur les actions des enseignants, de sorte qu'ils vont avoir tendance à, inconsciemment, agir de manière à confirmer ces dernières (Sprietsma, 2013). Mettant ainsi en lumière le phénomène selon lequel les attentes des enseignants ont tendance à se refléter dans les évaluations des copies des élèves et à limiter la réussite de ceux-ci (Dardenne, 2019).

Cependant, l'ancrage du stéréotype « redoublant » n'affecte pas de manière significative la note totale attribuée à la rédaction. La non-significativité d'un résultat peut trouver sa cause dans un des trois points suivants (Quertemont, 2021) : 1) Il n'y a réellement pas d'effet dans la population 2) La puissance statistique de l'étude était insuffisante pour mettre en évidence cet effet 3) Des erreurs méthodologiques ont été commises lors de la planification ou exécution de l'étude. Dans notre cas, la non-significativité des résultats pourrait trouver sa cause dans la petite taille de notre effectif (N=70). En effet, il se pourrait qu'en réalité il y ait bel et bien un effet, au sein de la population de référence, de l'ancrage du stéréotype « redoublant » sur la note totale attribuée aux copies mais qui n'a pu être démontré en raison d'un effectif insuffisant et donc d'une puissance statistique trop faible.

Ainsi, cette première hypothèse se voit confirmée uniquement au niveau de notre échantillon et ne peut être généralisée au niveau de la population de référence.

Ensuite, lorsque l'on s'intéresse aux copies corrigées à l'aide d'une grille d'évaluation, et plus précisément à la décomposition de la **note par bloc de critères**, nous constatons une certaine variabilité dans les résultats. En effet, tant pour le premier que pour le troisième bloc de critères, les résultats de la comparaison des moyennes des deux groupes de copies vont tous deux dans le même sens, en mettant en évidence que les « copies redoublants » sont moins bien notées que les « copies promus ».

Contre toute attente, le deuxième bloc de critères, quant à lui, relève d'un résultat opposé attestant que les « copies redoublants » seraient, en moyenne, mieux notées que les « copies promus ». Ainsi, il semblerait qu'au niveau de l'évaluation donnée à ce deuxième bloc de critères, les enseignants de notre échantillon se présentent, en moyenne, comme moins exigeants à l'égard des « copies redoublants » comparativement aux « copies promus ». Toutefois, malgré de nombreuses réflexions à ce sujet, nous reconnaissons ne trouver aucune hypothèses plausibles qui pourraient justifier une telle discordance dans nos résultats.

Par ailleurs, bien que les résultats obtenus au niveau des moyennes semblent attester d'une certaine influence de la présence du stéréotype « redoublant » sur les notes attribuées au niveau de chacun des **blocs de critères**, les résultats de l'analyse de la variance (ANOVA) attestent, au niveau de la variable « redoublant », d'un effet significatif uniquement au niveau du troisième bloc de critères. Les résultats des ampleurs de l'effet, quant à eux, démontrent un effet substantiel uniquement pour les deux derniers blocs de critères. Nous ne pouvons expliquer la non-significativité de ces résultats par la taille de nos effectifs puisqu'elle est identique dans chacun des blocs de critères (N=35) et pourtant, dans certains cas, nous obtenons des résultats significatifs alors que dans d'autres cas pas. Il semblerait donc que les critères eux-mêmes soient à l'origine de cette influence différentielle. Tout d'abord, si nous analysons plus précisément le contenu de ceux-ci, nous constatons que les deux derniers blocs de critères font référence à des éléments qui ont déjà été vus et soumis à une certaine évaluation durant les années de scolarité antérieures (Ministre de la Communauté Française, 2000b). En effet, ceux-ci concernent les maîtrises de l'orthographe, de la ponctuation, de la segmentation, etc. Ainsi, les évaluateurs seraient sujets à avoir des attentes différentielles pour ces critères, selon qu'il s'agisse d'une « copie redoublant » ou d'une « copie promu », et donc auraient tendance à adopter un jugement différentiel. Cette hypothèse interprétative permettrait d'expliquer les ampleurs de l'effet que nous constatons au niveau de ces deux blocs de critères mais n'est pas en mesure d'expliquer pourquoi le troisième bloc de critères rapporte des résultats significatifs alors que ce n'est pas le cas pour le deuxième bloc de critères. Ainsi, la présence d'un jugement différentiel concernant le deuxième bloc de critères ne peut être affirmé qu'au niveau de notre échantillon mais l'inférence au niveau de la population de référence ne peut être prouvée. Par contre, concernant le troisième bloc de critères, l'effet de la présence du stéréotype sur le jugement évaluatif est constaté dans notre échantillon et peut également être généralisé au niveau de la population de référence.

Ensuite, concernant l'absence de résultats substantiels et significatifs constatés au niveau du premier bloc de critères, il pourrait s'expliquer par le fait qu'être capable de rédiger un texte avec l'intention de convaincre est un objectif attendu en fin de troisième année secondaire (Ministre de la Communauté Française, 2000a), or l'élève redoublant, étant présenté comme répétant pour sa deuxième année secondaire, ne peut être considéré comme « avantagé » quant à la maîtrise de cette compétence. Les évaluateurs n'ont donc peu/pas d'attentes différentielles vis-à-vis des « copies redoublants » concernant ce premier bloc de critères, d'où la faiblesse de l'ampleur observé. Quant au caractère non significatif de ce résultat, il pourrait s'expliquer simplement par le fait qu'il n'y a pas réellement, dans la population de référence, d'effets liés à la présence stéréotype « redoublant » sur ce premier bloc de critères. Toutefois, les enseignants de notre échantillon ne sont pas tous professeurs de français et par conséquent, ne connaissent pas tous le contenu des programmes de français. Ainsi, il est peu probable que cette hypothèse interprétative soit également valable pour les enseignants de notre échantillon « non professeurs » de français.

Hypothèse n°2 : la variance observée entre les notes attribuées aux copies corrigées à l'aide d'une grille d'évaluation sera moins importante que celle observée entre les notes des copies corrigées sans grille d'évaluation.

Les résultats que nous avons obtenus en regard de cette deuxième hypothèse corroborent notamment les propos tenus par Chirart et ses collègues (2014) qui mettent en évidence que corriger une copie avec une grille d'évaluation permet aux évaluateurs de se centrer sur des points précis à évaluer en lien avec les objectifs d'apprentissage et ainsi de minimiser les variations d'interprétation qu'il pourrait y avoir entre différents évaluateurs à propos d'une même copie (Chirart et al., 2014). Par conséquent, grâce aux grilles d'évaluation, la subjectivité interprétative, survenant lors de l'évaluation des apprentissages, se voit réduite et le niveau de fidélité (ou fiabilité) augmenté (Berthiaume et al., 2011). En effet, les variances obtenues indiquent que l'évaluation des copies non accompagnées d'une grille d'évaluation amène à plus de variation inter-évaluateurs et ce, comparativement aux copies qui, quant à elles, sont évaluées à l'aide d'une grille d'évaluation. Notre deuxième hypothèse se voit ainsi confirmée par ce résultat.

Hypothèse n°3 : les enseignants évaluent de manière plus favorable une « copie redoublant » accompagnée d'une grille d'évaluation qu'une même copie mais non accompagnée d'une grille d'évaluation. Cette tendance s'observe également avec les « copies promus » mais dans une moindre mesure.

Les résultats obtenus suite à la comparaison des moyennes, des écarts-type entre les 4 groupes de copies ainsi qu'au calcul des ampleurs de l'effet nous permettent de constater que, au niveau de notre échantillon, la présence du stéréotype relatif au statut de redoublant sur la copie n'a pas le même effet selon que cette dernière soit accompagnée ou non d'une grille d'évaluation.

En effet, les résultats de ces analyses statistiques montrent que la présence de la grille d'évaluation amène à accorder plus de points au niveau de la **note totale** aux « copies redoublants » corrigées à l'aide d'une grille d'évaluation, qu'aux « copies redoublants » corrigées sans grille d'évaluation. Cette tendance s'observe également au niveau des « copies promus » mais dans une moindre mesure. Ainsi, cela **permet d'attester, de manière descriptive, de la présence d'un effet d'interaction entre la variable « redoublant » et la variable « grille d'évaluation »**. Ces constats rejoignent les propos tenus par Brown et Glasner (1999 cités par Mémo pédagogique, 2016) dans lesquels ils mettent en évidence que l'usage des grilles critériées est un bon moyen pour atténuer les biais d'évaluation.

Toutefois, lorsque nous poursuivons les analyses statistiques avec l'analyse de la variance simple, nous pouvons constater que cette **inférence n'est pas prouvée au niveau de la population de référence, et donc cette troisième hypothèse se voit confirmée dans notre échantillon mais invalidée dans la population de référence**. En effet, les résultats montrent qu'il n'y a pas d'interaction significative entre la présence du stéréotype « redoublant » et la présence de la grille d'évaluation. Dans notre étude, nous pourrions expliquer cette absence de résultats significatifs par la taille de nos effectifs (N=35 par groupe de copies). En effet, ces effectifs étant de très petite taille, il est possible qu'il y ait un effet significatif de la présence de la grille d'évaluation sur la note totale attribuée aux copies et que cet effet varie selon qu'il s'agit d'une « copie redoublant » ou non, mais que la significativité de cet effet d'interaction n'ait pu être démontré en raison d'un effectif de taille insuffisante et donc d'une puissance statistique trop faible.

Les variables sociodémographiques

Lorsque nous prêtons une attention particulière aux caractéristiques des participants qui constituent notre échantillon, nous constatons que certaines d'entre elles, comme celles relatives à leur(s) année(s) d'ancienneté, la discipline dans laquelle ils/elles enseignent, ou encore le fait qu'ils/elles aient connu ou non le redoublement durant leur parcours scolaire, peuvent exercer une influence substantielle sur leur jugement évaluatif (note totale) face à une rédaction à évaluer.

Premièrement, au niveau de **l'ancienneté** des participants, les résultats obtenus témoignent d'une influence significative de l'ancienneté des évaluateurs sur la note totale attribuée à la rédaction. Plus précisément, dans notre échantillon, les résultats démontrent que les enseignants expérimentés, en opposition aux enseignants débutants, attribuent, en moyenne une note totale plus élevée à la rédaction. En d'autres termes, il semblerait que les enseignants expérimentés se montrent plus laxistes dans leur manière d'évaluer cette rédaction que les enseignants débutants qui, à l'inverse, se montrent plus exigeants. Par ailleurs, cette différence entre ces deux groupes d'évaluateurs semble s'accroître au fur et à mesure que les années d'expérience s'accumulent. Ainsi, nos attentes initiales, quant à l'influence de l'ancienneté des enseignants sur leur jugement évaluatif, se voient confirmées dans notre échantillon et les résultats étant significatifs, l'inférence de ces résultats au niveau de la population de référence se voit également prouvée.

Deuxièmement, les résultats obtenus quant à **la discipline d'enseignement** des participants, attestent qu'au sein de notre échantillon, les professeurs de français ont tendance à attribuer, en moyenne et de manière significative, une note totale plus élevée à la rédaction comparativement aux professeurs exerçant dans une autre discipline que celle du français. Il semblerait donc que, dans notre échantillon, les professeurs de français fassent preuve de davantage d'indulgence dans leur manière d'évaluer la rédaction que les enseignants qui exercent dans une autre discipline que celle du français qui, eux, se montreraient plus sévères dans leur notation. Ces résultats rejoignent les propos tenus par Shohamy et ses collègues (1992, cités par Lim, 2011) lorsqu'ils soutiennent l'existence d'une différence entre les évaluateurs novices (peu ou pas formés dans la matière qu'ils doivent évaluer) et les évaluateurs experts (formés dans la matière qu'ils doivent évaluer) au niveau de leur sévérité générale de notation. Les résultats de l'étude de Weigle (1999) permettent également de soutenir les résultats de notre étude en attestant notamment que les évaluateurs novices ont tendance à être partiellement plus sévères que les évaluateurs experts tendance qui semble disparaître après que ceux-ci aient suivi une formation.

Ainsi, compte tenu de ces constats ; nos attentes initiales quant à l'influence de la discipline d'enseignement, sur le jugement évaluatif des enseignants, se voient confirmées au sein de notre échantillon. De plus, l'existence de cette influence peut également être généralisée au niveau de notre population de référence, étant donné la significativité de nos résultats.

Troisièmement, notons qu'étant donné les effets négatifs qu'entraîne le redoublement, parmi lesquels nous pouvons citer : la baisse d'estime de soi (Crahay, 2019), la stigmatisation (Crahay, 2019), la baisse de la performance scolaire (Crahay, 2019 ; Selimbegovic et al., 2011, cités par Crahay, 2019), l'augmentation du risque de décrochage scolaire (Jimerson et al., 2002), etc., nous pourrions être encline à penser que les enseignants l'ayant vécu, comparativement à ceux qui ne l'ont jamais vécu, feraient preuve de plus d'empathie envers les élèves qui l'ont également connu en évaluant leur copie avec davantage d'indulgence. Néanmoins, contrairement à nos attentes initiales, les résultats des analyses (régression linéaire et ampleur de l'effet) réalisées en regard de la variable « **connu ou non le redoublement** » attestent que le fait d'avoir connu le redoublement en tant qu'enseignant du secondaire influence négativement et, de manière significative, le jugement évaluatif de ce dernier, comparativement à l'enseignant ne l'ayant jamais vécu. De sorte que les enseignants l'ayant connu ont tendance à attribuer, en moyenne, une note totale moins élevée à la rédaction et ce, comparativement aux enseignants n'ayant jamais vécu le redoublement durant leur parcours scolaire. Ainsi, il semblerait que, les enseignants l'ayant connu aient tendance à se montrer plus sévères dans leur cotation que les enseignants qui, à l'inverse, ne l'ont jamais expérimenté durant leur parcours scolaire.

Nous ne pouvons affirmer connaître les raisons de l'existence de ce lien négatif entre le fait d'avoir connu le redoublement en tant qu'évaluateur et la note totale que l'on attribue à une « copie redoublant ». Cependant nous pouvons supposer que cela a un lien avec le phénomène d'internalisation du stéréotype. En effet, Crahay (2019) démontre qu'à force d'être confronté à un stéréotype, l'individu stigmatisé pourrait finir par l'internaliser et adhérer au stéréotype qui porte sur lui, orientant ainsi ses attitudes et ses comportements. Si nous transférons les propos tenus par Crahay (2019) au cas de notre étude, il semblerait que, dans notre échantillon, les enseignants ayant vécu le redoublement, comparativement à ceux ne l'ayant jamais vécu, aient fini par internaliser les stéréotypes négatifs qui leur sont associés (« mauvais », « paresseux », etc.), orientant ainsi leurs attentes vis-à-vis des élèves redoublants et donc leur jugement évaluatif à leur égard. En effet, comme le met en évidence Sprietsma (2013) les attentes des enseignants peuvent affecter la façon dont ils perçoivent et évaluent les performances des élèves. Ceci nous amène à formuler l'hypothèse interprétative suivante : dans notre échantillon, les enseignants ayant connu le redoublement durant leur parcours scolaire ont plus de risques d'être auteurs de la perpétuation de stéréotypes négatifs à l'égard des élèves redoublants que les enseignants ne l'ayant jamais connu. Autrement dit, l'hypothèse que nous avons initialement formulée en regard de la variable sociodémographique « connu le redoublement » se voit infirmée par les résultats notre étude.

Quatrièmement, par rapport à la caractéristique relative au **genre** des participants, les résultats de notre étude viennent confirmer nos attentes en démontrant l'absence de résultats significatifs quant à l'influence du genre des participants sur la note totale attribuée à la rédaction. Nous ne pouvons donc pas considérer que, dans le cadre de notre échantillon, la note totale attribuée à la rédaction soit influencée de manière substantielle par le genre des enseignants. De plus, au niveau de la population de référence, l'inférence ne peut être prouvée compte tenu de la non significativité des résultats. Par ailleurs, étant donné que notre rédaction est renseignée comme étant réalisée par un garçon, nos résultats sont en concordance avec ceux d'études antérieures qui se sont intéressées aux effets liés au genre de l'évaluateur (Basow, 1995 ; Basow & Silberg, 1987 ; Martin, 1984). Plus précisément, dans ces études au sein desquelles le rôle de l'évaluateur est incarné par les élèves et le rôle d'évalué par les enseignants, les résultats indiquent que globalement, les évalués de sexe féminin ont tendance à recevoir des notes plus élevées de la part des évaluateurs féminins et des notes plus faibles de la part des évaluateurs de sexe masculin (Basow, 1995 ; Basow & Silberg, 1987). En ce qui concerne les notes attribuées aux professeurs de sexe masculin, si on suit les résultats de l'étude de Basow (1995), il semblerait qu'elles ne soient pas affectées par le sexe des étudiants.

Nous clôturerons cette partie de discussion en disant ceci : dans notre échantillon, un élève a plus de risques d'obtenir une note plus faible à sa rédaction s'il est connu comme étant un élève redoublant, si sa copie est évaluée, sans l'aide d'une grille d'évaluation, par un enseignant débutant qui a connu le redoublement et qui n'est pas professeur de Français et ce, comparativement à une copie promu, évaluée avec une grille d'évaluation, par un enseignant expérimenté qui n'a pas connu le redoublement et qui est professeur de Français.

7. Limites et perspectives

Avant de terminer ce présent travail par une conclusion générale, nous allons présenter les limites et perspectives éventuelles de notre recherche afin de l'améliorer et de l'approfondir. Tout d'abord, la préoccupation majeure de cette étude réside dans la validité d'un échantillon de petite taille. En effet, notre échantillon compte un total de 140 participants répartis en 4 groupes, soit 35 participants par groupe. Or dans le cadre d'une étude quantitative, la taille d'un échantillon est très importante pour établir des statistiques valides, et « un échantillon trop petit diminue la pertinence des résultats d'une recherche [...] » (Ajar et al., 1983, p.10).

De plus, la petite taille de nos sous-échantillons, amène à avoir une certaine prudence dans l'interprétation des ampleurs de l'effet. En effet, comme le démontrent Slavin et Smith dans leur étude (2009), il existe une relation négative importante entre la taille d'un échantillon et le calcul de l'ampleur de l'effet. Dans leur méta-analyse, Cheung et Slavin (2016) soulignent une différence statistiquement significative entre les études à grande échelle ($n =$ ou > 250 sujets) et celles à petite échelle ($n < 250$ sujets). Plus précisément, ils démontrent que les études avec des échantillons de petite taille sont associées à des ampleurs d'effets presque deux fois plus importantes comparativement aux études avec des échantillons de grande taille. Ils expliquent notamment cette relation négative par le fait que les études à petite échelle, étant fréquemment plus étroitement contrôlées que les études à grande échelle, sont plus enclines à produire des résultats positifs (Cheung & Slavin, 2016). Cronbach et ses collègues (1980, cités par Cheung & Slavin, 2016) rejoignent ces propos en mettant en évidence que les résultats positifs des études à petite échelle pourraient être expliqués par l'effet de « **super-présentation** » : dans les études à petite échelle, comparativement aux études à large échelle, les chercheurs sont davantage en mesure de maintenir une grande fidélité de mise en œuvre ou de fournir un soutien supplémentaire. Par conséquent, répliquer cette recherche auprès d'un échantillon de plus grande taille permettrait notamment d'augmenter la puissance statistique de nos résultats ainsi que leur validité. Par ailleurs, ayant travaillé avec un échantillon de convenance, notre recherche a été menée auprès d'un échantillon restreint de professeurs qui n'est pas représentatif de la population de base étudiée, ce qui peut entraîner des biais d'échantillonnage.

Ensuite, dans la conception de l'analyse de cette étude, plusieurs autres facteurs limitent, dans une certaine mesure, la généralisation de ces résultats à la population de référence.

Premièrement, le fait que les enseignants aient été influencés par l'ancrage du stéréotype « redoublant » dans une situation expérimentale ne signifie pas nécessairement qu'ils seraient influencés de la même manière au sein de leur propre classe. En effet, en situation expérimentale, certains éléments de contexte, que l'on retrouve en situation naturelle, peuvent ne pas être présents étant donné le contrôle de certaines variables. Ainsi, il n'est pas inenvisageable que les effets observés résultent, en partie, de la situation expérimentale.

Deuxièmement, la situation environnementale dans laquelle la rédaction a été évaluée varie d'une participant à un autre. En effet, la participation à l'étude se réalisant à distance, il nous était difficile de contrôler, pour chacun des participants, la situation environnementale dans laquelle les deux tâches allaient être réalisées.

Ainsi, la participation à notre étude expérimentale a été réalisée par chacun des participants selon sa meilleure convenance, ce qui peut entraîner une influence sur nos résultats et biaiser ces derniers.

De plus, nous ne pouvons être certaine que chacun des participants ait lu la description de la cible inscrite sur la copie de l'élève. Autrement dit, il est possible que, pour certains évaluateurs, l'ancrage du stéréotype dans la description de la cible n'ait pas eu l'opportunité d'exercer une influence sur le jugement évaluatif de ces derniers. Par ailleurs, la description de la cible comportant plusieurs informations (prénom, année scolaire, date de naissance, redoublement), il est possible qu'une information, autre que la présence du stéréotype, ait exercé une influence sur le jugement évaluatif des enseignants, comme cela pourrait être éventuellement le cas pour le prénom de l'élève.

Ensuite, nous tenons à mettre en évidence que nous sommes consciente qu'un certain nombre de conditions d'applicabilité doivent être vérifiées avant de réaliser une analyse de la variance simple, comme la normalité des données ou encore l'homogénéité des variances (Quertemont, 2021). Toutefois, étant donné que dans notre recherche seule la première condition est rencontrée (voir tableau n°16 p.154), nous pourrions envisager de répliquer notre étude en appliquant un test alternatif qui permettrait de comparer des moyennes dont les variances sont inégales. Il existe plusieurs tests alternatifs et la procédure de Welch en est un exemple (Quertemont, 2021).

Pour terminer, afin de mieux comprendre la manière d'évaluer de chacun des participants, nous pensons qu'il serait intéressant de voir comment chacun d'entre eux se positionne quant aux stéréotypes qui pèsent sur les élèves redoublants. En effet, dans le cadre de notre étude, nous ne savons pas si les enseignants qui ont attribué moins de points aux copies redoublants l'ont fait parce qu'ils avaient de faibles attentes à leur égard et donc se sont comportés de manière à confirmer ces attentes, ou à l'inverse, s'ils avaient des attentes élevées envers les copies redoublants les amenant ainsi à être plus exigeants dans leur notation.

Pour ce faire, à l'instar de l'étude menée par Lafontaine et Monsieur (2009), nous pourrions envisager de soumettre aux participants, après qu'ils aient complété la rédaction et le questionnaire contextuel, un questionnaire composé d'une série d'items relatifs à leurs attentes envers les élèves redoublants (par exemple : les élèves redoublants sont moins rigoureux dans leur travail scolaire que les élèves promus). Sur base de ces items, les participants devraient marquer leur accord selon l'échelle de Likert à quatre échelons.

Ainsi, ce questionnaire permettrait d'appréhender l'existence éventuelle d'attentes stéréotypées que pourraient avoir les enseignants du secondaire envers les élèves redoublants.

8. Conclusion

Dans le cadre de ce mémoire, nous nous sommes intéressée à mesurer l'effet du statut redoublant, supposé d'un élève, ainsi qu'à l'effet d'interaction que pourrait avoir l'ancrage du stéréotype avec la présence d'une grille d'évaluation, sur le jugement évaluatif d'enseignants de l'enseignement secondaire ordinaire en FW-B.

Pour mener à bien notre étude, nous avons attribué de manière strictement aléatoire un type de copie à 140 enseignants. Les copies étaient toutes identiques en termes de contenu rédactionnel, mais leur différence résidaient d'une part, dans l'accompagnement ou non d'une grille d'évaluation et, d'autre part dans la présence ou non de l'activation subtile du stéréotype « redoublant ».

Grâce à cette étude expérimentale, nous avons été en mesure de formuler une réponse à notre question de recherche « Dans quelle mesure, l'ancrage du stéréotype du redoublement dans la copie d'un élève de 3^e année secondaire, accompagnée ou non d'une grille d'évaluation, va-t-il influencer l'évaluation donnée par des enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles sur cette copie ? ». En effet, les résultats obtenus nous permettent de mettre en évidence que, dans notre échantillon, face à des copies strictement identiques, mais attribuées dans certains cas à un élève redoublant et dans d'autres cas à un élève promu, les enseignants de l'enseignement secondaire ordinaire en FW-B, ayant participé à notre étude, ont tendance, en moyenne, à sous-évaluer les « copies redoublants » comparativement aux « copies promus ». Il semblerait donc que, au niveau de notre échantillon, les enseignants se comportent de manière plus exigeante avec les copies lorsqu'elles sont renseignées comme étant réalisées par un élève redoublant que lorsqu'elles sont signalées comme étant réalisées par un élève promu.

Ce jugement différentiel se retrouve également au niveau des copies lorsqu'elles sont accompagnées d'une grille d'évaluation, mais dans une moindre mesure.. En effet, nous constatons qu'en moyenne, la différence de points entre les « copies redoublants » et les « copies promus » est plus marquée lorsque les copies ne sont pas accompagnées d'une grille évaluation que lorsqu'elles le sont.

De plus, la comparaison des « copies redoublants avec grille » et des « copies redoublants sans grille » nous montre que les évaluateurs ont tendance à accorder plus de points aux « copies redoublants » lorsque celles-ci sont accompagnées d'une grille d'évaluation que lorsqu'elles ne le sont pas. Il semblerait donc qu'il y ait, au niveau de l'échantillon, un certain effet d'interaction entre l'ancrage du stéréotype « redoublant » et la présence de la grille d'évaluation de sorte que la présence de la grille d'évaluation minimise l'effet négatif que pourrait avoir la présence du stéréotype sur le jugement évaluatif des enseignants. Par ailleurs, de manière générale les résultats montrent que la présence de la grille d'évaluation permet de réduire la variabilité des résultats entre évaluateurs et ainsi de rendre le jugement évaluatif de ceux-ci plus objectif, plus fiable. L'ensemble de ces constats formulés en regard de notre question de recherche sont confirmés pour notre échantillon mais, en raison de la non significativité des résultats obtenus, leur inférence au niveau de la population de référence ne peut être prouvée.

Concernant les caractéristiques sociodémographiques des enseignants, obtenues grâce au questionnaire contextuel qui leur a été fourni, les caractéristiques relatives à l'ancienneté, la discipline d'enseignement ainsi que le fait qu'ils aient connu ou non le redoublement durant leur parcours scolaire, témoignent de la présence, dans notre échantillon, de certains biais de notation. Constats qui peuvent également être inférés à notre population de référence. De sorte que, les évaluateurs débutants ont tendance à se montrer, tant pour les « copies redoublants » que pour les « copies promus », plus exigeants dans leur évaluation comparativement aux évaluateurs expérimentés. Concernant la discipline d'enseignement, les professeurs de français se présentent comme étant plus indulgents dans leur notation, comparativement aux enseignants pour lesquels le contenu de la rédaction ne relève pas de leur discipline d'enseignement. Quant au fait d'avoir connu ou non le redoublement, contre toute attente, les enseignants ayant connu le redoublement se montrent, en moyenne, plus exigeants dans leur évaluation et ce comparativement aux évaluateurs n'ayant pas connu le redoublement durant leur parcours scolaire. Ce dernier constat se retrouve non seulement dans notre échantillon mais l'inférence dans la population de référence peut également être prouvée. Enfin, le genre des évaluateurs, quant à lui, n'exerce pas d'influence substantielle au niveau de la manière d'évaluer des enseignants de l'enseignement secondaire ordinaire en FW-B.

La présente recherche a donc permis, dans une certaine mesure, de mettre en lumière l'influence de certains facteurs, relatifs à l'évalué mais aussi relatifs aux évaluateurs, sur le jugement évaluatif des enseignants. Cependant, les limites méthodologiques de notre recherche réduisent l'application de certains de nos résultats, uniquement à notre échantillon et ne permettent donc pas la généralisation de ces derniers au niveau de la population de référence, c'est-à-dire à l'ensemble des enseignants de l'enseignement secondaire ordinaire en FW-B. Ces limites étant ainsi évoquées, nous pourrions envisager une réplication de notre étude auprès d'un effectif plus important. De plus, peut-être serait-il également intéressant de fournir à chacun des participants, dans une deuxième phase, une série d'items à connotation stéréotypée envers les élèves redoublants sur lesquels ils devraient se positionner selon une échelle de Likert à quatre échelons, permettant ainsi de connaître leur positionnement quant aux stéréotypes qui pèsent sur les élèves redoublants. Ce mémoire est donc une porte ouverte à d'autres recherches qui pourraient être envisagées sur le sujet.

9. Bibliographie

Ajar, D., Dassa, C., & Gougeon, H. (1983). L'échantillonnage et le problème de la validité externe de la recherche en éducation. *Revue des sciences de l'éducation*, 9(1), 3-21. <https://doi.org/10.7202/900396ar>

Allen, C.S., Chen, Q., Wilson, V.L., Hughes, J.N. (2009). Quality of research design moderates effects of grade retention on achievement : A meta-analytic, multilevel analysis, *Educational Evaluation and Policy Analysis*, 31(4), 480-499. <https://doi.org/10.3102/0162373709352239>

Andrade, H.-G. (2000). Using Rubrics to Promote Thinking and Learning, *Educational Leadership*, 57(5), 13-18.

Arnold, V., Legas, J., Obler, S., Pacheco, M. A., Russell, C., & Umbdenstock, L. (1990). *Do students get higher scores on their word processed papers? A study of bias in scoring handwritten versus word processed papers*. Rio Hondo College, Whittier, CA. <https://files.eric.ed.gov/fulltext/ED345818.pdf>

Autin, F., Batruch, A., & Butera, F. (2019). The Function of Selection of Assessment Leads Evaluators to Artificially Create the Social Class Achievement Gap. *American Psychological Association*, 111(4), 717-735. <http://dx.doi.org/10.1037/edu0000307>

Barthélémy, V. (2020). Le dispositif de remédiation et de lutte contre le décrochage scolaire : Ou comment repérer et accompagner le lycéen dans son parcours scolaire? *Nouveaux cahiers de la recherche en éducation*, 22(2), 45-68. <https://doi.org/10.7202/1074422ar>

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently ? *Journal of Educational Psychology*, 79(3), 308-314. <https://doi.org/10.1037/0022-0663.79.3.308>

Basow, S.A. (1995). Student evaluations of college professors: When gender matters. *Journal of educational Psychology*, 87(4), 656-665. <https://doi.org/10.1037/0022-0663.87.4.656>

Baye, A., & Dachet, D. (2020). Introduction aux Sciences de l'Éducation et de la Formation : notes de cours. Unpublished document, Université de Liège, Liège.

Baye, A. (2019). Introduction aux Sciences de l'Éducation et de la Formation : notes de cours. Unpublished document, Université de Liège, Liège.

Berthiaume, D., David, J., & David, T. (2011). Réduire la subjectivité lors de l'évaluation des apprentissages à l'aide d'une grille critériée : repères théoriques et applications à un enseignement interdisciplinaire. *Revue internationale de pédagogie de l'enseignement supérieur*, 27(2). <https://doi.org/10.4000/ripes.524>

Boraitai, F. (2016). Croyances à propos du redoublement de futurs enseignants entrant en formation dans différents contextes éducatifs. *Éducation comparée/nouvelle série*, 16, 91-115.

Boraitai, F. (2013). Effet d'un module de formation sur les croyances de futurs enseignants eu égard au redoublement. *Revue suisse des sciences de l'éducation*, 35(2), 347-369. <https://doi.org/10.25656/01:10276>

Boraitai, F. (2015). Les croyances de futurs enseignants sur le redoublement au regard de leurs connaissances sur ses effets et de leurs conceptions psychopédagogiques. *Revue des sciences de l'éducation*, 41(3), 483-508. <https://doi.org/10.7202/1035314ar>

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester : Wiley.
- Bourdieu, P., & Champagne, P.(1992). Les exclus de l'intérieur. *Actes de la recherche en sciences sociales*, 91/92, 71-75. <https://doi.org/10.3406/arss.1992.3008>
- Bridgeman, B., & Cooper, P. (1998). *Comparability of Scores on Word-Processed and Handwritten Essays on the Graduate Management Admissions Test*. <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=eric3&NEWS=N&AN=ED421528>
- Brown, A. (in press). Legibility and the rating of second language writing: an investigation of the rating of handwritten and word-processed IELTS Task Two essays. (4). https://www.ielts.org/-/media/research-reports/ielts_rr_volume04_report4.ashx
- Bureau du Secrétaire général des Écoles européennes. (s.d). Nouvelle échelle de notation pour le cycle secondaire : *Une évaluation améliorée et normalisée pour favoriser la réussite des apprenants*. Schola Europaea. <https://www.eursec.eu/fr/European-Schools/studies/marking-scale>
- Caverni, J-P., Fabre, J-M., & Noizet, G. (1975). Dépendance des évaluations scolaires par rapport à des évaluations. *Le Travail Humain*, 38(2), 213-222.
- Charney, D. (1984). The validity of using holistic scoring to evaluate writing. *Research in the teaching of English*, 18, 65-81. <https://doi.org/141417473>
- Chaumont, M., & Leroux, J. (2018). Le jugement évaluatif : Subjectivité , biais cognitifs et posture du professeur. *Pédagogie collégiale*, 31(3), 27-33.
- Cheung, A., & Slavin, R.E. (2016). How methodological features affect effect sizes in education. *Society for Research on Educational Effectiveness*. <https://doi.org/10.3102/0013189X16656615>
- Chirart, F., Dupre, A., & Leuleu, B. (2014). *Développer des grilles d'évaluation critériées : Mémo de l'atelier* <https://apprendre.auf.org/wp-content/opera/13-BF-References-et-biblio-RPT-2014/D%C3%A9velopper%20des%20grilles%20d%E2%80%99%C3%A9valuation%20crit%C3%A9ri%C3%A9es%20M%C3%A9mo%20de%20l%E2%80%99atelier.pdf>
- Circulaire n°7351 de la communauté française du 15 octobre 2019 : Circulaire 7351 - Modalités d'utilisation des bulletins destinés aux élèves de l'enseignement secondaire ordinaire et spécialisé de forme 4. Année scolaire 2019-2020. (2019). *Gallilex*, 15 octobre.
- Crahay, M. (2004). Peut-on conclure à propos des effets du redoublement ? *Revue Française de Pédagogie*, 148, 11-23. <https://doi.org/10.3406/rfp.2004.3246>
- Crahay, M. (2019). *Peut-on lutter contre l'échec scolaire ?* (4^e édition) Bruxelles : De Boeck Supérieur.
- Graig, D. A. (2001). *Handwriting legibility and word-processing in assessing rater reliability*. <http://www.danielcraig.com/illinois/mathesis/>
- Crespo, M. (1988). Analyse longitudinale de l'effet Pygmalion. *Revue des sciences de l'éducation*. 14(1), 3-23. <https://doi.org/10.7202/900581>
- Crocker, J., & Major, B. (1989). Social Stigma and Self-Esteem: The Self-Protective Properties of Stigma. *Psychological Review* , 96(4), 608-630. <https://doi.org/10.1037/0033-295x.96.4.608>

Croizet, J-C., & Martinot, D. (2003). Stigmatisation et estime de soi. In J-C. Croizet & J-P. Leyens (Eds). *Mauvaise réputation : Réalités et enjeux de la stigmatisation sociale* (pp.25-59). Paris : Armand Colin.

Cvencek, D. Meltzoff, A.N & Greenwald, A.G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, 82(3), 766-779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>

Dachet, D. (2020). Mémoire partim 1 : notes de cours. Unpublished document, Université de Liège, Liège.

Dachet, D., & Baye, A. (2020). Evidence-Based Education: The (Not So Simple) Case of French-Speaking Belgium. *ECNU Review of Education*, 4(1), 1-26. <https://doi.org/10.1177/2096531120928086>

Dardenne, B. (2019). Psychologie sociale : notes de cours. Unpublished document, Université de Liège, Liège.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1-44. <https://doi.org/10.14507/epaa.v8n1.2000>

D'Amours, G (s.d). *Les grilles d'évaluations*. HEC Montréal, Direction de l'apprentissage et de l'innovation pédagogique.

Décret du 24 juillet 1997 définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre (1997). *Moniteur belge*, 24 juillet 1997, p. 21557.

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

Draelants, H. (2006). Le redoublement est moins un problème qu'une solution : Comprendre l'attachement social au redoublement en Belgique francophone. *Les Cahiers de Recherche en Education et Formation*, 52, 3-21.

Fédération Wallonie-Bruxelles. (s.d). *Enseignants et personnel de l'enseignement - Taux de féminisation*. Retrieved May 1, 2022, from <https://statistiques.cfwb.be/transversal-et-intersectoriel/genre/enseignants-et-personnel-de-lenseignement-taux-de-feminisation/>

Fédération Wallonie-Bruxelles. (2010b, février 18). *Outils d'évaluation - 2e et 3e degrés des humanités générales et technologiques - Français*. Retrieved October 1, 2021, from <http://www.enseignement.be/index.php?page=24187&navi=1718>

Fédération Wallonie-Bruxelles. (2020, novembre). *Les indicateurs de l'enseignement en Fédération Wallonie-Bruxelles* (15ème édition). Namur, Belgique : Administration générale de l'Enseignement.

Galand, B., Lafontaine, D., Baye, A., Dachet, D., & Monseur, C. (2019). Le redoublement est inefficace, socialement injuste, et favorise le décrochage scolaire. *Service d'analyse des Systèmes et des Pratiques d'enseignement (aSPe)*, 38.

Gary-Bobo, R., & Robin (2012). Le redoublement est-il efficace et nuisible ? *Regards croisés sur l'économie*, 2(12), 98-113. <https://doi.org/10.3917/rce.012.0098>

Gérard, F.-M. (2002). L'indispensable subjectivité de l'évaluation. *Antipodes*, 156, 26-34.

Goffin,C., Monseur,C., & Fagnant,A.(2013). Les croyances relatives au redoublement, les conceptions de l'intelligence et le sentiment d'auto-efficacité du secondaire supérieur : quelle articulation ? Communication présentée au congrès d'Actualité de la recherche en éducation et formation (AREF), Montpellier, France.

Goos, M., Van Damme, J., Onghena, P., Petry, K., de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*(3), 323–347. <https://doi.org/10.1016/j.jsp.2013.03.002>

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 24, 2022, from <https://par-temps-clair.blogspot.com/2018/07/ampleur-deffet-un-indicateur.html?fbclid=IwAR3xGtme7OeIcu-15S58iTCyMdGtb17s31PMCWQhc5BXaj9wO4UgbQZjxpo>

Grisay A. (1993). *Le fonctionnement des collèges et ses effets sur les élèves de sixième et de cinquième*, France, Paris.

Guyomar, A., Karnfelt, C., & Picouet, P. (2015). Comment développer l'utilisation de grilles critériées (rubrics) dans les apprentissages ? *Questions de pédagogies dans l'enseignement supérieur : Innover : pourquoi et comment ? 1*, 343 – 348.

Herawati., A. (2011). The use of rubric: an opportunity to promote thinking and learning. *HUMANIORA. 2*(1), 445-451. <https://doi.org/10.21512/humaniora.v2i1.3053>

Holmes,T., & Mattews,K.(1984). The effets of Nonpromotion on Elementary and Junior Hugh School Pupils : A Meta-Analysis. *Review of Educational Research Summer, 54*(2), 225-236. <https://doi.org/10.2307/1170303>

Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement, 20*(1), 65-70. <https://doi.org/10.1111/j.1745-3984.1983.tb00190.x>

Hutmacher, W. (1993). Quand la réalité résiste à la lutte contre l'échec scolaire : Analyse du redoublement dans l'enseignement primaire genevois. Genève : *Cahier du service de la recherche sociologique, 36*.

Jimerson, S.R., Anderson, G.E., & Whipple, A.D. (2002). Winning the battle and losing the war : Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools, 39*(4),441-457. <https://doi.org/10.1002/pits.10046>

Jimerson, S.R. (2001). Meta-analysis of grade retention research: Implications fort practice in the 21st century. *School Psychology Review, 30*, 420-437. <https://doi.org/10.1080/02796015.2001.12086124>

Johansson, S., Strietholt, R., Rosén, M., & Myrberg, E. (2013). Valid inferences of teachers' judgements of pupils' reading literacy : does formal teacher competence matter? *School Effectiveness and School Improvement : an International Journal of Research, Policy and Pratices. https://doi.org/10.1080/09243453.2013.809774*

Jussim,L., Crawford,J.T., & Rubinstein, R.S.(2015). Stereotype (in) accuracy in perceptions of groups and individuals. *Current Directions in Psychological Sciences, 24*(6), 490-497. <https://doi.org/10.1177/0963721415605257>

Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing, 10*(2), 134-148. <http://dx.doi.org/10.1016/j.asw.2005.05.002>

Lafontaine, D., Bricteux, S., Hindryckx, G., Matoul, A., & Quittre, V. (2019). Performances des jeunes de 15 ans en lecture, mathématiques et science. Service d'analyse des Systèmes et des Pratiques d'enseignement (*aSPE*). Fédération Wallonie-Bruxelles. Retrieved March 5, 2021, from <http://www.enseignement.be/index.php?page=26997>

Lafontaine, D., & Monseur Ch. (2009). 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>

Leclerq, 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)

Lecompte, J. (1997). Robert Rosenthal l'effet Pygmalion. *Les grands dossiers des sciences humaine*, 54.

Leveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue suisse des sciences de l'éducation*, 30(30), 483-500. <https://doi.org/10.25656/01:4230>

Leyens, J. P., & Scaillet, N. (2012). *Sommes-nous tous des psychologues?* Mardaga.

Lim, G.S. (2011). The development and maintenance of rating quality in performance writing assessment : A longitudinal study of new and experienced raters. *Language testing*, 28(4), 543-560. <https://doi.org/10.1177/0265532211406422>

Marcroux, G., & Crahay, M. (2008). Mais pourquoi continuent-ils à faire redoubler ? Essai de compréhension du jugement des enseignants concernant le redoublement. *Revue suisse des sciences de l'éducation*, 30(3), 501-518. <https://doi.org/10.25656/01:4231>

Markham, L. R. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal*, 13(4), 277-283. <https://doi.org/10.3102/00028312013004277>

Martin, E. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Journal of Women in Culture and Society*, 9, 482-492.

Mémo pédagogique (2016). *Mémo pédagogique : Utiliser une grille critériée pour réduire la subjectivité de l'évaluation*. Université de Bourgogne, Centre d'innovation pédagogique et d'évaluation. <https://inspe.u-bourgogne.fr/cipe-ressources-pedagogiques>

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research Evaluation*, 7(25). <https://doi.org/10.7275/gcy8-0w24>

Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review*, 8(2), 193-210.

Ministre de la Communauté Française (2000a). *Enseignement secondaire ordinaire de plein exercice : Humanités générales et technologiques - Enseignement secondaire général et technique de transition : Deuxième et troisième degrés* (60/2000/240). Administration Générale de l'Enseignement et de la Recherche Scientifique.

Ministre de la Communauté Française (2000b). *Enseignement secondaire ordinaire de plein exercice : Premier degré commun : 1^{ère} année A - 2^e année commune* (43/2000/240) : Programme d'études du cours de français Administration Générale de l'Enseignement et de la Recherche Scientifique.

Mosconi, N. (1999). Les recherches sur la socialisation différentielles des sexes à l'école. *Débats Jeunes*, 4, 85-116.

Mottier L.L., & Allal, L. (2008). Le jugement professionnel en évaluation: un acte cognitif et une pratique sociale située. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30 (3), 465-482. <https://doi.org/10.25656/01:4229>

Mottier,L.L., & Tessaro,W. (2016). Introduction : pourquoi interroger le jugement professionnel dans l'évaluation des apprentissages des élèves. In Peter Lang, *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages*. (pp.1-4).

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257. <https://doi.org/10.3102/01623737026003237>

Odell, L., & Cooper, C. R. (1980). Procedures for Evaluating Writing: Assumptions and Needed Research. *College English*, 42, 35-43. <https://doi.org/10.2307/376031>

Perrenoud, P. (1989). La triple fabrication de l'échec scolaire. *Psychologie française*, 34(4), 237-245.

Popham, W. J. (1997). What's Wrong--and What's Right--with Rubrics, *Schools as Safe Havens*, 55(2), 72-75.

Pourtois, J-P. (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. <https://doi.org/10.3406/rfp.1978.1665>

Powers, D. E., Fowles, M., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs ? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31(3), 220-233. <https://doi.org/10.1111/j.1745-3984.1994.tb00444.x>

Quertemont, E (2021). Psychostatistique descriptive et inférentielle partim 2 : notes de cours. Unpublished document, Université de Liège, Liège.

Reddy, M. Y., & Andrade, H. (2009). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>

Roberge, J. (2006). *Corriger les textes de vos élèves : Précisions et stratégies*. Montréal : Chenelière éducation.

Romainville, M. (2011). Objectivité versus subjectivité dans l'évaluation des acquis des étudiants: *Revue internationale de pédagogie de l'enseignement supérieur*, 27(2). <https://doi.org/10.4000/ripes.499>

Rudman, L., Ashmore, R., & Gary, M. (2001). Unlearning Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes. *Journal of Personality Social Psychology*, 81(5), 856-868. <https://doi.org/10.1037//0022-3514.81.5.856>

Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University Press.

Shmader, T., Johns, M., & Forbes,C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336-356. <https://doi.org/10.1037/0033-295X.115.2.336>

Slavin, R. E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic review in education. *Educational Evaluation and Policy Analysis*, 31(4), 500-506. <http://dx.doi.org/10.3102/0162373709352369>

Snyder, M., & Stukas Jr, , A.A. (1999). Interpersonal processes : The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology*, 50, 273-303. [https://doi.org/0084-6570/99/0201-0273\\$08.00](https://doi.org/0084-6570/99/0201-0273$08.00)

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empir Econ*, 45, 523-538. <https://doi.org/10.1007/s00181-012-0609-x>

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a « universal » rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment Evaluation in Higher Education*, 36(5), 509-547. <https://doi.org/10.1080/02602930903540991>

Trouilloud,D., & Sarrazin,P.(2003). Les connaissances actuelles sur l'effet Pygmalion : processus, poids et modulateurs. *Revue Française de Pédagogie*, 145, 89-119. <https://doi.org/ffhal00388839f>

Weigle, S.C.(1999).Investigation des interactions évaluateur/protocole dans l'évaluation de l'écriture : Approches quantitatives et qualitatives. *Assessing Writing*, 6(2), 154-178. [https://doi.org/2935\(00\)00010-6](https://doi.org/2935(00)00010-6)

Wolfe, E.W., Kao, C-W., & Ranney, M. (1998). Cognitive Differences in Proficient and Nonproficient Essay Scorers.. *Written Communication*, 15(4), 465. <https://doi.org/10.1177/0741088398015004002>

Yzerbyt, V.(2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90-95. <https://doi.org/10.1016/J.copsy.2016>

10. Résumé

Notre revue de la littérature nous a permis de mettre en exergue les inégalités de traitement dont peuvent être victimes les élèves dans le contexte de l'évaluation. Que ce soit en raison de leur origine ethnique, de leur genre, de leur statut socioéconomique ou encore de leur faculté calligraphique, tous ces facteurs sont susceptibles d'influencer, inconsciemment, le jugement évaluatif des enseignants, si bien que leur jugement évaluatif se verrait biaisé par ces différents facteurs. Nos recherches menées sur les grille d'évaluation, nous ont permis, quant à elles, de constater que leur utilisation permettait de minimiser la subjectivité interprétative et donc de rendre l'évaluation plus fiable, fidèle. Ainsi, compte tenu de ces constats et de la prédominance de la pratique du redoublement dans notre système éducatif en FW-B, nous nous sommes intéressée à mesurer l'effet du statut redoublant, supposé d'un élève, ainsi qu'à l'effet d'interaction que pourrait avoir l'ancrage du stéréotype avec la présence d'une grille d'évaluation sur le jugement évaluatif d'enseignants de l'enseignement secondaire ordinaire en FW-B. Pour ce faire, nous avons fourni à des enseignants du secondaire une rédaction sur laquelle nous avons activé de manière subtile, dans certains cas le stéréotype « redoublant », tandis que dans d'autres cas pas. En plus de cela, dans certains cas la rédaction s'est vue accompagnée d'une grille d'évaluation et dans d'autres cas pas. En plus de cette rédaction, nous avons fourni à chacun des participants un questionnaire contextuel afin de recueillir des informations à leur sujet telles que leur genre, leur nombre d'année(s) d'ancienneté dans l'enseignement, s'ils ont connu ou non le redoublement durant leur parcours scolaire, et leur discipline d'enseignement.

Grâce à la participation de 140 enseignants de l'enseignement secondaire ordinaire en FW-B, nous avons pu observer, dans notre échantillon, que les « copies redoublants » avaient tendance à être moins bien évaluées que les « copies promus ». Il semblerait donc que les enseignants aient tendance à se comporter de manière plus exigeante avec les copies lorsqu'elles sont renseignées comme étant réalisées par un élève redoublant que lorsqu'elles sont signalées comme étant réalisées par un élève promu. Toutefois, la présence du stéréotype n'exerce pas la même influence sur le jugement évaluatif des copies selon qu'elles soient accompagnées ou non d'une grille d'évaluation. De sorte que les évaluateurs ont tendance à accorder plus de points aux « copies redoublants » lorsqu'elles sont accompagnées d'une grille d'évaluation que lorsqu'elles ne le sont pas. Quant à l'influence des caractéristiques sociodémographiques des enseignants telles que l'ancienneté, le fait d'avoir connu ou non le redoublement et la discipline d'enseignement, les résultats témoignent d'une influence de ces facteurs sur le jugement évaluatif des enseignants. Seule la variable relative au genre des enseignants ne relève d'aucune influence sur leur jugement évaluatif de ceux-ci.

11. Annexes

Table des annexes

Annexe 1. 1 ^e version de la rédaction avec ancrage du stéréotype « redoublant ».....	92
Annexe 2. 1 ^e version de la rédaction sans ancrage du stéréotype « redoublant ».....	93
Annexe 3. 1 ^e version de la grille d'évaluation	94
Annexe 4. 2 ^e version de la rédaction avec ancrage du stéréotype « redoublant ».....	96
Annexe 5. 2 ^e version de la rédaction sans ancrage du stéréotype « redoublant ».....	97
Annexe 6. 2 ^e version de la grille d'évaluation	98
Annexe 7. Guide de correction	100
Annexe 8. Questionnaire contextuel.....	102
Annexe 9. Évaluation initiale de la rédaction.....	103
Annexe 10. Outil d'évaluation du portail de l'enseignement en FW-B.....	107
Annexe 11. Annonce de recrutement sur les réseaux sociaux (Facebook et Instagram)	111
Annexe 12. Mail à l'intention de la direction	111
Annexe 13. Compte rendu des modifications apportées suite au laboratoire cognitif	113
Annexe 14. Résultats du laboratoire cognitif.....	118
Annexe 15. Formulaire de consentement	140
Annexe 16. Formulaire d'information	143

Table des figures

Figure 1. Les étapes du processus d'évaluation des apprentissages.....	147
Figure 2. L'usage de la grille critériée inscrite dans le temps (Berthiaume et al., 2011).....	147

Table des tableaux

Tableau 1. Résultats du laboratoire cognitif.....	147
Tableau 2. Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes.	148
Tableau 3. Tableau récapitulatif reprenant les résultats au test de normalité pour les notes totales attribuées à chacun des groupes de copies	148
Tableau 4. Tableau récapitulatif reprenant les résultats au test de normalité pour les notes attribuées à chacun des blocs de critères selon qu'il s'agit d'une « copie redoublant » ou d'une « copie promu »	149
Tableau 5. Analyse de la variance simple pour la note totale.....	149
Tableau 6. Analyse de la variance simple pour la note attribuée à chacun des blocs de critères	150
Tableau 7. Moyennes, écarts-types et ampleurs de l'effet de la note totale attribuée à la rédaction selon la présence ou non du stéréotype relatif au statut redoublant	151
Tableau 8. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale.....	151
Tableau 9. Moyennes, écarts-types et ampleurs de l'effet des notes attribuées à chacun des blocs de critères de la grille d'évaluation selon la présence ou non du stéréotype relatif au statut redoublant.....	151

Tableau 10 : Tableau récapitulatif reprenant les résultats des sommes de carrés de type III de l'analyse de variances simples pour chacun des blocs de critères	152
Tableau 11. Calcul de la variance des notes totales selon la présence ou non d'une grille d'évaluation	152
Tableau 12. Moyenne et écarts-types de la note totale pour chacune des 4 copies	152
Tableau 13. Calcul de l'ampleur de l'effet entre les 4 groupes de copies en fonction de la note totale.....	153
Tableau 14. Régression linéaire de la note totale selon la présence du stéréotype « redoublant », de la grille d'évaluation, de l'ancienneté des évaluateurs, du fait qu'ils aient connu ou non le redoublement, de leur genre et de s'il s'agit de professeurs de français ou non.	153
Tableau 15. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée aux « copies redoublants » selon que l'évaluateur a connu ou non le redoublement	153
Tableau 16. Résultats au test d'homogénéité des variances pour la variable note totale (test de Levene).....	154

Texte argumentatif sur le mariage forcé

Consigne générale de rédaction : rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une introduction, trois arguments à développer et une conclusion.

Description de la cible : Texte argumentatif réalisé par Thomas, élève de 3^e année secondaire, né le 20 avril 2006, ayant redoublé en 2^e année secondaire.

Selon moi, le mariage forcé n'est pas quelque chose de bon et n'est donc pas à prôner. Ce sont souvent les femmes de pays occidentaux qui en sont victimes. Je vais donc expliquer mon point de vue vis-à-vis du mariage forcé.

Tout d'abord, je ne trouve pas que le mariage forcé soit une chose juste et digne de l'être humain. Le fait qu'un père oblige sa fille à se marier avec un inconnu souvent beaucoup plus âgé qu'elle est cruel. Sa progéniture devrait avoir le droit de se marier avec la personne de son choix et avec qui elle désire passer le restant de sa vie. Tout simplement, avec celui qu'elle aime.

Ensuite la jeune épouse ne sera plus libre, libre de ses choix, d'elle-même. En effet, elle devra vivre la maison de son mari à qui elle devra une obéissance absolue. Elle vivra loin de sa famille sans jamais la revoir.

Par contre, le mariage peut être bénéfique pour la jeune femme.

Par conséquent, elle peut épouser un homme riche et doux qui ne la battra jamais et qui l'aimera pour ce qu'elle est.

En conclusion, le mariage forcé selon moi est un acte cruel envers sa progéniture. Les femmes devraient avoir le choix de se marier avec la personne qu'elle souhaite et ne pas avoir une obéissance absolue envers son conjoint. Cependant, il peut être bénéfique pour la personne qui se marie si elle est entre de bonnes mains.

Texte argumentatif sur le mariage forcé

Consigne générale de rédaction : rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une introduction, trois arguments à développer et une conclusion.

Description de la cible : Texte argumentatif réalisé par Thomas, élève de 3^e année secondaire, né le 20 avril 2007.

Selon moi, le mariage forcé n'est pas quelque chose de bon et n'est donc pas à prôner. Ce sont souvent les femmes de pays occidentaux qui en sont victimes. Je vais donc expliquer mon point de vue vis-à-vis du mariage forcé.

Tout d'abord, je ne trouve pas que le mariage forcé soit une chose juste et digne de l'être humain. Le fait qu'un père oblige sa fille à se marier avec un inconnu souvent beaucoup plus âgé qu'elle est cruel. Sa progéniture devrait avoir le droit de se marier avec la personne de son choix et avec qui elle désire passer le restant de sa vie. Tout simplement, avec celui qu'elle aime.

Ensuite la jeune épouse ne sera plus libre, libre de ses choix, d'elle-même. En effet, elle devra vivre la maison de son mari à qui elle devra une obéissance absolue. Elle vivra loin de sa famille sans jamais la revoir.

Par contre, le mariage peut être bénéfique pour la jeune femme.

Par conséquent, elle peut épouser un homme riche et doux qui ne la battra jamais et qui l'aimera pour ce qu'elle est.

En conclusion, le mariage forcé selon moi est un acte cruel envers sa progéniture. Les femmes devraient avoir le choix de se marier avec la personne qu'elle souhaite et ne pas avoir une obéissance absolue envers son conjoint. Cependant, il peut être bénéfique pour la personne qui se marie si elle est entre de bonnes mains.

Annexe 3. 1^{ère} version de la grille d'évaluation

Critères, indicateurs, niveaux de maîtrise et pondération

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 -1 0-1-2-3 0-1-2-3 0-1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	3.1 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0
4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes.	6
		<input type="checkbox"/> De 90% à 94% de formes correctes	4
		<input type="checkbox"/> De 80% à 89% de formes correctes	2
		<input type="checkbox"/> Moins de 80% de formes correctes	0

	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			/50

Annexe 4. 2^e version de la rédaction avec ancrage du stéréotype « redoublant »

Texte argumentatif sur le mariage forcé

Consigne générale de rédaction : rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une thèse, trois arguments à développer et une conclusion.

Description de la cible : Texte argumentatif réalisé par Thomas, élève de 3^e année secondaire, né le 20 avril 2006, ayant redoublé en 2^e année secondaire.

Selon moi, le mariage forcé n'est pas quelque chose de bon et n'est donc pas à prôner. Ce sont souvent les femmes de pays occidentaux qui en sont victimes. Je vais donc expliqué mon point de vue vis-à-vis du mariage forcé.

Tout d'abord, je ne trouve pas que le mariage forcé soit une chose juste et digne de l'être humain. Le fait qu'un père oblige sa fille à se marier avec un inconnu souvent beaucoup plus âgé qu'elle est cruel. Sa progéniture devrait avoir le droit de se marier avec la personne de son choix et avec qui elle désire passer le restant de sa vie. Tout simplement, avec celui qu'elle aime.

Ensuite la jeune épouse ne sera plus libre, libre de ses choix, d'elle-même. En effet, elle devra vivre la maison de son mari à qui elle devra une obéissance absolue. Elle vivra loin de sa famille sans jamais la revoir.

Par contre, le mariage peut être bénéfique pour la jeune femme.

Par conséquent, elle peut épouser un homme riche et doux qui ne la battra jamais et qui l'aimera pour ce qu'elle est.

En conclusion, le mariage forcé selon moi est un acte cruel envers sa progéniture. Les femmes devraient avoir le choix de se marier avec la personne qu'elle souhaite et ne pas avoir une obéissance absolue envers son conjoint. Cependant, il peut être bénéfique pour la personne qui se marie si elle est entre de bonnes mains.

Annexe 5. 2^e version de la rédaction sans ancrage du stéréotype « redoublant »

Texte argumentatif sur le mariage forcé

Consigne générale de rédaction : rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une thèse, trois arguments à développer et une conclusion.

Description de la cible : Texte argumentatif réalisé par Thomas, élève de 3^e année secondaire, né le 20 avril 2007.

Selon moi, le mariage forcé n'est pas quelque chose de bon et n'est donc pas à prôner. Ce sont souvent les femmes de pays occidentaux qui en sont victimes. Je vais donc expliquer mon point de vue vis-à-vis du mariage forcé.

Tout d'abord, je ne trouve pas que le mariage forcé soit une chose juste et digne de l'être humain. Le fait qu'un père oblige sa fille à se marier avec un inconnu souvent beaucoup plus âgé qu'elle est cruel. Sa progéniture devrait avoir le droit de se marier avec la personne de son choix et avec qui elle désire passer le restant de sa vie. Tout simplement, avec celui qu'elle aime.

Ensuite la jeune épouse ne sera plus libre, libre de ses choix, d'elle-même. En effet, elle devra vivre la maison de son mari à qui elle devra une obéissance absolue. Elle vivra loin de sa famille sans jamais la revoir.

Par contre, le mariage peut être bénéfique pour la jeune femme.

Par conséquent, elle peut épouser un homme riche et doux qui ne la battra jamais et qui l'aimera pour ce qu'elle est.

En conclusion, le mariage forcé selon moi est un acte cruel envers sa progéniture. Les femmes devraient avoir le choix de se marier avec la personne qu'elle souhaite et ne pas avoir une obéissance absolue envers son conjoint. Cependant, il peut être bénéfique pour la personne qui se marie si elle est entre de bonnes mains.

Annexe 6. 2^e version de la grille d'évaluation

Critères, indicateurs, niveaux de maîtrise et pondération

Pour vous assurer une bonne compréhension de cette grille, nous vous invitons à consulter le guide de correction disponible à la suite de cette grille d'évaluation.

CRITÈRES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher lorsque c'est nécessaire)	PONDÉRATION (à entourer)
1. Respect de l'intention : convaincre	<p>1.1 Le texte intègre toutes les caractéristiques d'une argumentation</p> <p>1.2 Pertinence de l'argumentation</p> <p>a) à la situation de communication</p> <p>b) aux informations données et à la réalité vécue</p> <p>1.3 Cohérence de l'argumentation</p>	<p>Présence de la thèse</p> <p>Présence de 3 arguments</p> <p>Présence d'une conclusion</p> <p>a) Arguments recevables</p> <p>b)</p> <p><input type="checkbox"/> Aucune erreur de compréhension des systèmes</p> <p><input type="checkbox"/> Erreurs(s) de compréhension des systèmes</p> <p><input type="checkbox"/> Pas de contradiction entre la conclusion et les divers arguments</p> <p><input type="checkbox"/> Contradiction(s) entre la conclusion et les divers arguments</p>	<p>0 -1</p> <p>0-1-2-3</p> <p>0-1</p> <p>0-1-2-3</p> <p>3</p> <p>0</p> <p>4</p> <p>0</p>
2. Cohérence textuelle	<p>2.1 Segmentation en paragraphes</p> <p>2.2 Développement des arguments</p> <p>2.3 Pertinence de la connexion (connecteurs logiques)</p>	<p>Un argument par paragraphe</p> <p>Les exemples permettent de développer les arguments</p> <p>Les arguments sont bien formulés, clairs, précis</p> <p><input type="checkbox"/> Utilisation de connexions correctes et efficaces</p> <p><input type="checkbox"/> Une ou deux erreur(s) de connexion</p> <p><input type="checkbox"/> Plus de deux erreurs de connexion</p>	<p>0-1-2-3</p> <p>0-1-2-3</p> <p>0-1-2-3</p> <p>4</p> <p>2</p> <p>0</p>

3. Respect des normes linguistiques	3.1 Maitrise de l'orthographe	<input type="checkbox"/> De 0 à 1 erreur <input type="checkbox"/> De 2 à 3 erreurs <input type="checkbox"/> De 4 à 5 erreurs <input type="checkbox"/> Plus de 5 erreurs	6 4 2 0
	3.2 Maitrise de la ponctuation et de la syntaxe	<input type="checkbox"/> De 0 à 2 erreur(s) <input type="checkbox"/> Plus de 2 erreurs	3 0
	3.3 Maitrise du vocabulaire	<input type="checkbox"/> De 0 à 2 erreur(s) <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL (faire la somme de toutes les pondérations et inscrire le total ici)			/40

Annexe 7. Guide de correction

Au travers de cette rédaction, l'auteur se devra de communiquer, par écrit, à autrui un jugement personnel argumenté sur le sujet du mariage forcé.

1. Respect de l'intention de convaincre

Une argumentation doit respecter différentes règles, à savoir : comprendre une thèse, comprendre des arguments (dans ce cas-ci trois), et une conclusion.

Quant aux arguments, ils doivent posséder les qualités suivantes : être développés (illustrés par un exemple, être justifiés), être pertinents (tenir la route, être plausibles), et être cohérents (en lien avec la thèse et ne pas être en contradiction avec la conclusion).

Plus précisément, voici à quoi se réfèrent chacun des indicateurs ci-dessous :

1.1 Le texte intègre toutes les caractéristiques d'une argumentation

- Présence d'une thèse : la thèse est le noyau du texte argumentatif, l'idée qui sera défendue par l'auteur tout au long de son texte.
- Présence de 3 arguments : c'est au travers des arguments que l'auteur développe, soutient sa thèse.
- Présence d'une conclusion : la conclusion reprend les différentes idées énoncées dans les arguments, c'est l'opinion globale.

1.2 Pertinence de l'argumentation

La pertinence d'une argumentation est liée à la situation de communication mais également aux informations données et à la réalité vécue.

- a) à la situation de communication : dans le cadre de cette rédaction, l'auteur doit se positionner par rapport au mariage forcé. Si un des arguments ne concerne pas le mariage forcé mais un autre thème, alors il ne pourra être considéré comme pertinent.
- b) aux informations données et à la réalité vécue : dans le cadre de cette rédaction, si un argument repose sur une erreur de compréhension du système actuel et/ou du nouveau système proposé, il ne pourra être considéré comme pertinent.

De plus, si l'argumentation dénigre ou valorise un ou des individus, elle ne pourra être acceptée. Elle doit impérativement rester au niveau du système.

Exemples :

- *Je veux changer de classe car je ne supporte pas mon professeur de géographie.*
- *Je veux plus d'heures d'anglais car le professeur est sympathique*

1.3 Cohérence de l'argumentation

La cohérence implique la non-contradiction entre les arguments et la conclusion. La conclusion doit découler logiquement de l'argumentation qui précède.

2. Cohérence textuelle

2.1 Segmentation en paragraphes

L'argumentation développée par l'auteur est articulée sous forme de paragraphes. Autrement dit, dans le cas de cette rédaction, voici la segmentation qui doit être obtenue : un paragraphe pour l'exposition de la thèse, un paragraphe par argument développé, un paragraphe pour la conclusion.

2.2 Développement des arguments

Si l'auteur développe plus en profondeur un argument, on peut lui accorder un point bonus, mais ce qui est attendu en 3^e année secondaire, c'est qu'il y ait une justification à l'aide d'au moins un exemple pour chacun des arguments.

De plus, il est important que les arguments soient formulés de manière compréhensible, claire et précise. Nous devons pouvoir comprendre ce qu'a voulu dire l'élève au travers de ses arguments sur base de la simple lecture de ceux-ci.

2.3 Pertinence de la connexion

La connexion est jugée comme correcte et efficace lorsqu'elle favorise la lisibilité. Par conséquent, l'absence d'un connecteur là où le lecteur peut inférer facilement une relation logique ne peut donc être considérée comme une erreur. On considère comme erreur un connecteur mal choisi en fonction du rapport logique.

3. Respect des normes linguistiques

3.1. Maîtrise de l'orthographe

Dans les socles de compétences, les seuils de maîtrise orthographique sont fixés à 80% de formes correctes (= 5 erreurs) avec outils à 12 ans et à 90% (3 erreurs) à l'âge de 14 ans. Si on applique cette logique de progression au 2^e degré, on obtient un seuil de 95% de formes correctes (= 1 erreur).

3.2. Maîtrise de la ponctuation et de la syntaxe

Étant donné qu'il n'est pas toujours évident pour un correcteur de décider si telle erreur relève de la syntaxe ou de la ponctuation, nous avons fait le choix de regrouper ces deux critères en un seul.

Attention, il ne faut pas confondre les maladresses stylistiques avec des erreurs syntaxiques.

Voici une série d'exemples d'erreurs : une erreur dans le choix du mode ou du temps verbal, l'emploi incorrect d'un mot, une rupture de construction, une phrase incomplète, un oubli de virgule (après un complément en tête de phrase, dans une énumération, pour isoler une apposition, etc.), l'emploi de la virgule pour le point.

En matière de ponctuation, il s'agira d'être très prudent car il est fréquent que deux évaluateurs ponctuent de manière différente un même texte.

3.3. Maîtrise du vocabulaire

Il s'agira ici de pénaliser l'auteur pour l'emploi d'un mot à la place d'un autre et pour l'utilisation non volontaire d'un terme. Toutefois, on ne le pénalisera pas pour un « manque de richesse » du vocabulaire.

Annexe 8. Questionnaire contextuel

Questionnaire contextuel

Veillez répondre aux quatre questions suivantes :

- 1) Quel est votre **genre** ?

- 2) Dans quelle **discipline** enseignez-vous ?

- 3) Depuis combien de **temps** enseignez-vous ?

- 4) Avez-vous **déjà redoublé** durant votre parcours scolaire ?

Annexe 9. Évaluation initiale de la rédaction

28/05/2018

français :
L'argumentation,
Le mariage forcé

Consignes:

- Rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une introduction, trois arguments à développer et une conclusion.

Selon moi, le mariage forcé n'est pas une
bonne chose.

~~orientaux~~ ≠ ! Ce sont souvent des femmes de pays
occidentaux qui en sont victimes.

Je vais donc expliquer mon point de vue
vis-à-vis du mariage forcé.

Veille à préciser l'introduction
re.

Tout d'abord, je ne trouve pas que le
mariage forcé soit une chose juste et digne de l'être
humain.

Le fait qu'un père oblige sa fille à se marier
avec un inconnu souvent beaucoup plus âgé qu'elle
est cruel.

La progéniture devrait avoir le droit de se
marier avec la personne de son choix et avec qui elle
désire passer le restant de sa vie. Tout simplement,
avec celui qu'elle aime.

Ensuite, la jeune épouse ne sera plus libre,
libre de ses choix, d'elle-même.

En effet, elle devra vivre dans la maison
de son mari à qui elle devra une obéissance absolue.
Elle vivra loin de sa famille sans jamais
la revoir.

Par contre, le mariage peut être bénéfique
pour la jeune femme. En effet,
Elle peut épouser un homme riche et doux
qui ne la battra jamais et qui l'aimera pour ce
qu'elle est.

Tu es très paik. Tu n'as développé d'arguments.

En conclusion, le mariage forcé, selon moi,
est un acte cruel envers sa progéniture.

Les femmes devraient avoir le choix de se
marier avec la personne qu'elle souhaite et ne
pas ^{devoir} avoir une obéissance absolue envers son conjoint.

Cependant, il peut être bénéfique pour la
personne qui se marie si elle est entre de bonnes mains.

Saps apem !!

Critères, indicateurs, niveaux de maîtrise et pondération

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			6/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - 1 0-1-2-3 0-1-2-3 0-1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	3.2 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0

4. Respect des normes linguistiques	4.4 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	6 4 2 0
	4.5 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 0
	4.6 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			32/48 → 26,67/40 → 6,67/10 → Rédaction d'un niveau satisfaisant

Annexe 10. Outil d'évaluation du portail de l'enseignement en FW-B

Critères, indicateurs, niveaux de maîtrise et pondération

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à faire)	PONDERATION PROPOSEE (à faire)
1. Respect de l'intention : convaincre 22/50	1.1 Développement d'une argumentation 1.2. Pertinence de l'argumentation - aux informations données et à la réalité vécue - à la situation de communication 1.3 Cohérence de l'argumentation	<input type="checkbox"/> Présence d'une conclusion (opinion globale) <input type="checkbox"/> Présence des 3 avis <input type="checkbox"/> Présence d'une justification pour chaque avis <input type="checkbox"/> Aucune erreur de compréhension des systèmes <input type="checkbox"/> Erreur(s) de compréhension des systèmes <input type="checkbox"/> Arguments recevables <input type="checkbox"/> Arguments non recevables <input type="checkbox"/> Pas de contradiction entre la conclusion et les divers avis <input type="checkbox"/> Contradiction entre la conclusion et les divers avis	OUI/NON ⁷ 3 ou 0 3+3+3 De 4 à 0 De 4 à 0 2 0
2. Respect du genre : Formulaire 6/50	2.1. Formulaire rempli complètement et correctement 2.2. Respect des contraintes matérielles (espaces disponibles)	<input type="checkbox"/> Tous les éléments ⁸ sont présents et correctement positionnés <input type="checkbox"/> Un ou des éléments sont absents ou mal positionnés <input type="checkbox"/> Contraintes respectées <input type="checkbox"/> Contraintes non respectées	2 0 2 0

⁷ Si le niveau de maîtrise de cet indicateur est négatif, on considèrera la tâche demandée comme non accomplie et on n'attribuera aucun point aux autres indicateurs du critère 1.

⁸ Hormis la conclusion déjà évaluée en 1.1

	2.3. Énonciation adaptée à la situation	<input type="checkbox"/> Énonciation adaptée <input type="checkbox"/> Énonciation inadaptée	2 0
	2.4. Utilisation de procédés susceptibles de persuader le lecteur (non activé)		
3. Cohérence textuelle 6/50	3.1. Pertinence de la segmentation en paragraphes (non activé) 3.2. Pertinence de la connexion 3.3. Correction des anaphores	<input type="checkbox"/> Connexion correcte et efficace <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion <input type="checkbox"/> Aucune erreur <input type="checkbox"/> Une erreur <input type="checkbox"/> Deux erreurs	3 2 0 3 1 0
4. Respect des normes linguistiques (pour un texte d'environ 200 mots) (1) 16/50	4.1. Correction de l'orthographe 4.2. Correction de la ponctuation et de la syntaxe 4.3. Correction du vocabulaire	<input type="checkbox"/> Plus de 95 % de formes correctes <input type="checkbox"/> 95 % de formes correctes <input type="checkbox"/> Moins de 95 % de formes correctes <input type="checkbox"/> De 1 à 3 erreurs <input type="checkbox"/> Entre 4 et 6 erreurs <input type="checkbox"/> Plus de 7 erreurs <input type="checkbox"/> Une erreur <input type="checkbox"/> Entre 2 et 4 erreurs <input type="checkbox"/> 5 erreurs ou plus	6-5 4-3 De 2 à 0 6-5 4-3 De 2 à 0 4 3-2 1-0
TOTAL			/50

(1) Les niveaux de maîtrise proposés pour ce critère respectent une cohérence d'ensemble induite par les socles de compétences (cf. guide de correction). Le même souci de cohérence amène à proposer une pondération modulée mais le correcteur peut bien entendu utiliser un autre type de pondération : ainsi, pour l'orthographe par exemple, 6-5 peut devenir 6, 4-3 peut devenir 3 et 2-0 peut devenir 0.

Guide de correction

Le guide qui suit complète et précise la grille d'évaluation. En proposant une définition de certaines notions, il contribuera sans doute à harmoniser l'évaluation de différents correcteurs et à déterminer les apprentissages à mettre en place en amont.

1. Respect de l'intention : convaincre

L'argumentation doit posséder trois qualités :

- elle doit être développée
- elle doit être pertinente
- elle doit être cohérente

1.1. Développement d'une argumentation

Une argumentation comporte nécessairement une thèse (opinion, avis) et un ou des arguments plus ou moins développés. En 3^e on peut se satisfaire de la justification de chacun des avis, sans se priver pour autant d'accorder un bonus à l'élève qui développerait un argument. Il est à noter que, dans la présente épreuve, l'élève est amené à exprimer son opinion globale sous forme de conclusion en cochant la dernière case du formulaire (je suis pour/je suis contre/j'hésite).

1.2. Pertinence de l'argumentation

La pertinence de l'argumentation est liée à la situation de communication.

En l'occurrence, l'élève doit ici se positionner par rapport à une innovation scolaire.

Un argument ne sera pas pertinent s'il repose sur une erreur de compréhension du système actuel et/ou du nouveau système proposé. Par ailleurs, on ne pourra accepter que l'argumentation dénigre ou valorise des personnes ; elle doit rester au niveau du système.

Exemples

Je veux plus d'heures d'allemand parce que le professeur est sympathique.

Je veux le nouveau système parce que je ne supporte plus mon professeur de mathématiques.

1.3. Elle doit être cohérente

La cohérence implique la non-contradiction entre les arguments. La conclusion doit découler logiquement de l'argumentation qui précède même si cela n'implique pas obligatoirement un choix identique pour chacun des avis.

2. Respect du genre

2.1. Formulaire rempli complètement et correctement

Par définition un formulaire doit être complètement rempli. Pour éviter d'évaluer la même chose deux fois, on n'évalue pas ici la présence de l'avis global (conclusion) et des 3 avis (cf 1.1.)

2.2. Respect des contraintes matérielles

Le formulaire définit l'espace disponible pour les réponses.

2.3. Énonciation adaptée à la situation

Marques de l'énonciateur et du destinataire, modalisateurs, registre de langue, marques de courtoisie doivent être adaptés à la situation. Le destinataire n'étant pas un pair, aucune familiarité ni écart de courtoisie ne peut être toléré.

Exemple d'énonciation inadaptée tiré d'une copie d'élève

Encore contre ! Nous commençons comme d'habitude, mais nous terminons trop tôt. On est en général et pas en technique de transition ou de qualification ! Je trouve aussi que c'est une mauvaise idée de regrouper deux heures d'un même cours. On ne tiendrait pas en géographie et en anglais ! Non, c'est aussi une mauvaise idée.

2.4. Utilisation de procédés susceptibles de persuader le lecteur

Le genre formulaire ne se prête guère à l'usage de procédés rhétoriques.

3. Cohérence textuelle

3.1. Pertinence de la segmentation en paragraphes

Cet indicateur n'est pas activé pour le formulaire.

3.2. Pertinence de la connexion

La connexion est correcte lorsqu'elle favorise la lisibilité. L'absence d'un connecteur là où le lecteur peut inférer facilement une relation logique ne peut donc pas être considérée comme une erreur. Une erreur manifeste, c'est surtout un connecteur mal choisi en fonction du rapport logique.

3.3. Correction des anaphores

Exemples d'erreurs

- un pronom est employé pour un autre (par exemple, *il* pour *celui-ci* ou *que* pour *dont...*),
- deux termes sont considérés à tort comme synonymes,
- ...

4. Respect des normes linguistiques

4.1. Correction de l'orthographe

Les socles de compétences fixent des seuils de maîtrise orthographique à 80% de formes correctes (avec outils) à 12 ans et à 90% à 14 ans. Si on transfère cette logique de progression aux deuxième et troisième degrés on obtient les seuils suivants : 95% de formes correctes au 2^e degré, 97% de formes correctes au 3^e degré.

4.2. Correction de la ponctuation et de la syntaxe

Si la ponctuation et la syntaxe sont confondues dans un seul indicateur, c'est parce qu'il est parfois difficile pour un correcteur de décider si telle erreur relève de l'un ou de l'autre poste.

Les maladresses stylistiques ne sont pas à confondre avec des erreurs syntaxiques.

Voici une série d'exemples d'erreurs : une rupture de construction, une erreur dans le choix du mode ou du temps verbaux, une phrase incomplète, un oubli de virgule (après un complément en tête de phrase, dans une énumération, pour isoler une apposition, etc.), l'emploi de la virgule pour le point. En matière de ponctuation, on sera extrêmement prudent : il est très courant que deux experts ne ponctuent pas le même texte de la même manière.

4.3. Correction du vocabulaire

On sanctionnera l'emploi d'un mot pour un autre et l'utilisation non volontaire d'un terme. On ne sanctionnera pas le « manque de richesse » du vocabulaire.

Annexe 11. Annonce de recrutement sur les réseaux sociaux (Facebook et Instagram)



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 production écrite, en vue d'améliorer les pratiques évaluatives.

Vous êtes enseignants dans l'enseignement secondaire ordinaire en FW-B et partants pour m'aider dans ma recherche ?

Alors j'ai besoin de vous ! Cela ne vous prendra pas plus de 30 minutes. Pour ce faire, commentez cette publication en y indiquant votre adresse mail, je vous contacterai en retour dans les plus brefs délais.

D'avance merci, et n'hésitez pas à partager, cela m'aidera beaucoup !

Annexe 12. Mail à l'intention de la direction

Objet : Demande de participation à une étude en Sciences de l'Éducation

Monsieur/Madame,

Je m'appelle Kim Rauw et je suis actuellement étudiante en Sciences de l'Éducation à l'Université de Liège. Je me permets de vous contacter, car, dans le cadre de mon mémoire, je mène un projet de recherche ayant comme objectif général de mesurer les facteurs qui interviennent dans l'évaluation d'une production écrite, en vue d'améliorer les pratiques évaluatives.

Pour mener à bien cette recherche, il me faut recueillir au minimum 120 enseignants secondaire, à qui je ferai passer deux épreuves en ligne qui ne prendront pas plus de 30 minutes. Dès lors, seriez-vous d'accord pour que j'interroge les enseignants de votre école ? Si oui, pouvez-vous faire suivre le document ci-joint en l'envoyant aux enseignants de votre école.

Si vous avez des questions, je reste disponible par téléphone ou par email.

En vous remerciant d'avance pour votre participation, je vous prie d'agréer, Monsieur/Madame, mes salutations distinguées.

Kim Rauw.

Contacts :

Adresse mail : kim.rauw@student.uliege.be

Numéro de téléphone : 0491/ 18 03 06

Document à joindre au mail :



Kim Rauw
kim.rauw@student.uliege.be
0491/ 18 03 06

Objet : Demande de participation à une étude en Sciences de l'Éducation

Monsieur/Madame,

Je m'appelle Kim Rauw et je suis actuellement étudiante en Sciences de l'Éducation à l'Université de Liège. Je me permets de vous contacter, car, dans le cadre de mon mémoire, je mène un projet de recherche ayant comme objectif général de mesurer les facteurs qui interviennent dans l'évaluation d'une production écrite, en vue d'améliorer les pratiques évaluatives.

Vous êtes enseignants dans l'enseignement secondaire ordinaire en FW-B et partant pour m'aider dans ma recherche ?

Alors j'ai besoin de vous et cela ne prendra pas plus que de 30 minutes de votre temps. Pour ce faire, envoyez-moi un mail sur l'adresse : kim.rauw@student.uliege.be. Vous recevez ensuite un formulaire d'information présentant la recherche, un consentement à signer ainsi que deux petites épreuves à réaliser.

Par ailleurs, sachez que la participation à cette étude est totalement anonyme et que les données récoltées resteront confidentielles.

Si vous avez des questions, je reste disponible par téléphone ou par email.

En vous remerciant d'avance pour l'intérêt que vous porterez à mon projet et pour votre participation, je vous prie d'agréer, Monsieur/Madame, mes salutations distinguées.

Kim Rauw

Annexe 13. Compte rendu des modifications apportées suite au laboratoire cognitif

Pour attester des résultats de ce laboratoire cognitif nous allons reprendre la grille d'évaluation, critère par critère, et mettre ces derniers en parallèle avec les commentaires qui ont été formulés par les participants.

Tout d'abord, au niveau du premier critère relatif à *l'impression globale*, ce dernier a fait l'objet de nombreux commentaires. La plupart des participants (6/8) ont mis en évidence la subjectivité de ce critère – puisqu'il ne se base sur aucun indicateur, – et donc la difficulté qu'ils ont éprouvée pour attribuer une note à ce dernier. Certains (2/8) ont d'ailleurs suggéré qu'il serait préférable de positionner ce critère en fin de grille afin que les évaluateurs puissent évaluer celui-ci en se basant sur les autres critères de la grille d'évaluation. D'après eux, en procédant de cette manière, le jugement des évaluateurs en regard de ce critère serait moins biaisé et donc plus objectif.

Initialement, nous avons choisi ce critère car, dans le cadre de cette recherche, il nous semble important de laisser une certaine liberté dans la notation aux participants qui évaluent à l'aide de la grille d'évaluation. En effet, nous sommes partie du principe que, sans cette liberté, il nous serait difficile d'observer des variations entre les jugements évaluatifs des évaluateurs. Ce critère relatif à *l'impression globale* nous semblait donc être idéal pour permettre d'attester de ces variations. Concernant sa position en début de grille, nous avons fait ce choix car notre objectif au travers de ce critère est d'avoir l'impression globale et spontanée de l'évaluateur à la suite de la lecture de la rédaction. Si nous le positionnons à la fin de la grille d'évaluation, l'impression globale ne sera plus spontanée car elle risquerait de subir l'influence des autres critères. Par exemple, nous pouvons supposer que l'enseignant pourrait être amené à dévaloriser le contenu d'une rédaction s'il constate, en évaluant plus en détail cette dernière, qu'elle présente de nombreuses erreurs orthographiques et syntaxiques.

Néanmoins, sur base des résultats obtenus lors du laboratoire cognitif, nous constatons la présence de variations inter-codeurs au niveau de plusieurs critères : seuls les niveaux de maîtrise « présence de la thèse », « présence d'une conclusion » de l'indicateur 2.1 et l'indicateur 4.3 ont obtenu une note identique de la part des évaluateurs. Dès lors, étant donné les raisons qui nous ont motivée à intégrer le critère *impression globale* et compte tenu des résultats précédemment cités, il nous semble que ce critère, problématique pour la plupart des enseignants interrogés, peut être supprimé.

Ensuite, sur base des remarques des quelques participants nous avons choisi de remplacer certains mots par d'autres :

2.3 Cohérence de l'argumentation	<input checked="" type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis arguments.	2
	<input checked="" type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis arguments	0

Dans les niveaux de maîtrise de l'indicateur 2.3, certains participants nous ont fait remarquer que l'emploi du mot « avis » pouvait sous entendre que divers avis étaient acceptés, en l'occurrence les contre-arguments. Cependant, ayant fait le choix de ne pas accepter ces derniers dans le cadre de ce travail, nous avons remplacé le mot « avis » par « arguments ».

4. Respect des normes linguistiques	4.1 Correction Maîtrise de l'orthographe	<input type="checkbox"/> De 0 à 1 erreurs <input type="checkbox"/> De 2 à 3 erreurs <input type="checkbox"/> De 4 à 5 erreurs <input type="checkbox"/> Plus 5 erreurs	6 4 2 0
	4.2 Correction Maîtrise de la ponctuation et de la syntaxe	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
	4.3 Correction Maîtrise du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0

Dans les indicateurs 4.1, 4.2, 4.3, nous avons décidé de remplacer le mot « correction » par le mot « maîtrise ». En effet, celui-ci perturbant certains participants (2/8), nous avons préféré le remplacer par un autre mot que nous jugeons plus approprié.

Consigne générale de rédaction : rédige une argumentation sur le mariage forcé. Celle-ci devra comporter une introduction **une thèse**, trois arguments à développer et une conclusion.

Dans la consigne générale de rédaction, nous nous sommes rendue compte que parler d'une « introduction » pouvait prêter à confusion. En effet, un participant nous a fait prendre conscience que dans les critères de la grille d'évaluation nous ne parlons pas « d'introduction » mais bien de « thèse ». Ainsi, pour rester en cohérence avec les termes de la grille d'évaluation, nous avons modifié la consigne générale de rédaction en remplaçant le mot « introduction » par le mot « thèse ».

Par ailleurs, la plupart des participants ont soulevé le caractère imprécis de plusieurs critères de cette grille, et voici ce que certains ont spécifiquement mis en évidence : préciser le mot thèse (3/8), préciser ce qu'on entend par « arguments recevables » (2/8) (à partir de quand un argument peut-il être considéré comme recevable ? Sur base de quels critères ?), préciser si les contre-arguments sont acceptés ou non (1/8), préciser ce qu'on entend par « correction du vocabulaire » (1/8). Par conséquent, puisque cette grille d'évaluation est amenée à être utilisée par des enseignants tout venant, il nous a semblé important de tenir compte de ces diverses remarques et de l'aménager en fonction.

C'est pourquoi, pour nous assurer de la bonne compréhension et utilisation de cette grille d'évaluation, nous avons joint à cette dernière un guide de correction qui complète et précise chacun des critères qui y sont présents (annexe n°7, p.100).

	2.2 Pertinence de l'argumentation a) à la situation de communication b) aux informations données et à la réalité vécue	a) Arguments recevables b) <input type="checkbox"/> Aucune erreur de compréhension des systèmes <input type="checkbox"/> Erreurs(s) de compréhension des systèmes	0-1-2-3 3 0
--	--	--	-----------------------

En outre, tout en réalisant notre guide de correction, pour plus de précision, nous avons décidé de scinder l'indicateur 2.2 en deux parties : a) et b) comme cela est suggéré dans la grille de correction disponible sur le portail de l'enseignement en FW-B dans la famille de tâches « convaincre » destinée aux élèves du 2^e degré.

Au niveau de système de notation, de manière générale, il a été mis en évidence que la pondération était trop catégorique et ne laissait pas place à un continuum plus large des points. Ainsi, voici les modifications qui ont été appliquées :

1. Impression globale			10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	Présence de la thèse	0-1
		Présence de 3 arguments	0-1-2-3
		Présence d'une justification pour chaque argument (exemples)	0-1-2-3
		Présence d'une conclusion (opinion globale)	0-1

Premièrement, un participant a mis en évidence que l'auteur de cette rédaction ne devrait pas obtenir la moitié des points étant donné que le but d'une argumentation est de convaincre, d'exemplifier et de détailler alors qu'ici, le développement de l'élève est pauvre et présente certaines contradictions. Malgré cela, le participant déclare que l'élève obtient tout de même la moitié des points, ce qu'il explique par le fait qu'une grande partie de la pondération est accordée à la « présence de... », et à l'impression globale. Autrement dit, 2/5^e de la note sont consacrées à de l'appréciation ou à la simple présence des éléments demandés. Cette remarque ainsi formulée appuie d'une part notre décision précédente de supprimer le critère *impression globale* et, d'autre part, nous amène à supprimer le niveau de maîtrise relatif à « la présence d'une justification pour chaque arguments » étant donné que celui-ci est déjà repris dans l'indicateur 3.2.

2.2 Pertinence de l'argumentation	a) à la situation de communication	a) Arguments recevables (bien choisis par rapport à la thèse). Arguments non recevables (mal choisis par rapport à la thèse)	4 → 0-1-2-3
	b) aux informations données et à la réalité vécue	b) <input type="checkbox"/> Aucune erreur de compréhension des systèmes <input type="checkbox"/> Erreurs(s) de compréhension des systèmes	0 3 0

Deuxièmement, nous avons changé le système de notation relatif à l'indicateur 2.2. Initialement, nous avons développé deux niveaux de maîtrise, soit : « arguments recevables » auquel on attribuait la note de 4 et « argument non recevables » auquel on attribuait la note de 0. Toutefois, plusieurs participants (6/8) ont rencontré des difficultés pour effectuer un choix entre ces deux extrêmes. Effectivement, ils trouvaient que deux arguments sur trois étaient recevables, et voulaient donc inscrire une note intermédiaire mais le système de notation ne le permettait pas : soit, tous les arguments étaient considérés comme recevables, soit aucun. C'est pourquoi nous avons décidé de supprimer le niveau de maîtrise « arguments non recevables » et avons gardé uniquement celui « arguments recevables » auquel nous avons attribué la pondération suivante : 0-1-2-3. Ainsi, les participants sont en mesure de choisir si aucun, un, deux ou tous les arguments sont recevables.

4. Respect des normes linguistiques	4.1 Maîtrise de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes → de 0 à 1 erreurs	6
		<input type="checkbox"/> De 90% à 94% de formes correctes → de 2 à 3 erreurs	4
		<input type="checkbox"/> De 80% à 89% de formes correctes → de 4 à 5 erreurs	2
		<input type="checkbox"/> Moins de 80% de formes correctes → plus de 5 erreurs	0

Troisièmement, plusieurs participants (3/8) nous ont exposé la difficulté qu'ils avaient ressentie pour attribuer une note à l'indicateur 4.1. Dès lors, sur base des recommandations de ces derniers, nous avons choisi d'évaluer la maîtrise orthographique en s'inspirant de la méthode utilisée pour les indicateurs 4.2 et 4.3, à savoir parler en termes de nombre d'erreurs.

	4.2 Maîtrise de la ponctuation et de la syntaxe	<input type="checkbox"/> De 0 à 2 erreurs	5 → 3
		<input type="checkbox"/> Plus de 2 erreurs	0

Quatrièmement, nous avons modifié la notation attribuée au niveau de l'indicateur 4.2 suite aux commentaires de plusieurs participants (3/8) soulignant que beaucoup de points étaient accordés à cet indicateur, et que la transition entre le 5 et le 0 était trop radicale. En effet, l'étudiant qui fait 2 erreurs, à tous les points et celui qui fait seulement, une faute de plus, perd 5 points. Par conséquent, nous avons fait le choix de changer le système de notation en attribuant 3 points à l'élève qui commet entre 0 et 2 erreurs de ponctuation et de syntaxe, et 0 point à l'étudiant qui commet plus de 2 erreurs.

2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers arguments. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers arguments	2 → 4 0
----------------------------------	--	------------

Cinquièmement, si nous tenons compte de toutes ces modifications apportées, nous arrivons à une évaluation notée sur 38. Or, certains évaluateurs n'ayant pas de grille d'évaluation pour appuyer leur jugement évaluatif, nous avons décidé de changer la pondération du niveau de maîtrise associée à l'indicateur 2.3 en passant d'une pondération de 2 à une pondération de 4. En effet, nous avons estimé que ce critère revêtait une grande importance compte tenu de l'objectif de l'évaluation et qu'il méritait donc qu'on y accorde une plus grande pondération dans la note finale.

Sixièmement, au niveau de la note attribuée par chacun des participants, nous observons de grands écarts entre évaluateurs (voir tableau n°1, p.147) : les notes varient de 26/50 à 39/50 alors qu'il s'agissait pourtant de la même copie et que tous ont reçu une grille d'évaluation pour émettre leur jugement évaluatif.

De plus, nous constatons qu'en moyenne, les évaluateurs ont attribué une note de 34,75/40 à cette rédaction ce qui correspond à une note de 6,95/10, et donc à une copie d'un niveau satisfaisant (Bureau du Secrétaire général des Écoles européennes. s.d), ce qui rejoint la note que nous avons initialement attribuée à l'aide de la première version de la grille d'évaluation à savoir celle de 6,67/10.

Enfin, certains participants (3/8) ne comprenant pas toujours l'utilité des cases à cocher, nous les avons laissées uniquement lorsque cela était pertinent. C'est-à-dire lorsqu'il fallait choisir entre différents niveaux de maîtrise, comme c'est le cas pour les niveaux de maîtrise relatifs aux indicateurs 2.2, 2.3, 3.3, 4.2, et 4.3.

Pour conclure, en mettant en application toutes les modifications précédemment citées, nous obtenons une 2^e version de la rédaction (annexe n°4, p.96), ainsi qu'une 2^e version de la grille d'évaluation (annexe n°5, p.97) qui nous amène à revoir la note que nous avons initialement attribuée à cette copie. En effet, en mettant en application cette nouvelle grille, nous attribuons à la rédaction une note de 5,75/10. Nous passons donc d'une copie initialement évaluée comme étant d'un niveau satisfaisant (6,67/10) à un copie d'un niveau suffisant (5,75) (Bureau du Secrétaire général des Écoles européennes. s.d).

Annexe 14. Résultats du laboratoire cognitif

Critères, indicateurs, niveaux de maîtrise et pondération (participant 1)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale ⁹			8/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 -1 0-1-2-3 0-1-2-3 0-1
	2.2 Pertinence ¹⁰ de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	3.3 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe (→ il y en a 2 dans la conclusion)	0-1-2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0

⁹ je trouve que c'est assez vague/subjectif de donner une impression globale en début de grille. Du coup soit je le mettrais en dernier soit je ne le mettrais pas comme critère parce qu'il n'y a aucun indicateur qui le définissent clairement. Mnt c'est peut être le but de ton étude et du coup ça a tout son sens de le laisser...

¹⁰ je trouve que c'est assez vague/subjectif de donner une impression globale en début de grille. Du coup soit je le mettrais en dernier soit je ne le mettrais pas comme critère parce qu'il n'y a aucun indicateur qui le définissent clairement. Mnt c'est peut être le but de ton étude et du coup ça a tout son sens de le laisser...

4. Respect des normes linguistiques	4.1 Correction ¹¹ de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes ¹³	6 4 2 0
	4.2 Correction de la ponctuation et de la syntaxe ¹² (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			39/50

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?

J'ai répondu au point 2.2

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

J'ai répondu au point 4.1 et 4.2

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

J'ai répondu au point 4.1

4) Le système de notation est-il compréhensible ?

Pour moi il est tout à fait compréhensible oui.

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

J'ai éprouvé des difficultés dans certains points de l'évaluation car je ne sais pas vraiment ce qui peut être attendu pour un élève de 3ème secondaire. Dès lors, il m'est difficile de savoir si les exemples sont assez détaillés ou si plus d'informations auraient été nécessaires. De plus, mettre une appréciation globale me semble être qqch d'assez subjectif. Aucun indicateur précis ne peut détailler ce critère.

¹¹ je ne mettrais peut être pas "correction de l'orthographe" mais plus "maîtrise de l'orthographe" et pareil pour le vocabulaire au point 4.3.

¹² et du coup ici, je mettrais plus respect de la ponctuation et de la syntaxe personnellement

¹³ ce ne serait pas plus simple de parler en nbre d'erreurs comme en dessous? Parce qu'avec les % il faut prendre le nbre de mot puis faire une règle de trois pour savoir dans quelle partie on se situe (ou alors je n'ai encore rien compris ahah)

Critères, indicateurs, niveaux de maîtrise et pondération (participant 2)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale	La dissertation de l'étudiant répond à la consigne générale (présence d'une introduction, de trois arguments et d'une conclusion). On perçoit que l'élève a compris les différents éléments qui doivent être contenus dans chacune des parties cependant la formulation est quelque peu ambiguë. Par exemple, le troisième argument est difficilement pertinent puisque, dans la manière dont il est rédigé, on a l'impression que l'argument en tant que tel est « bénéfique » et l'exemple « être marié à un époux bienveillant ». Or, dire que c'est bénéfique implique qu'on explique pourquoi, ce qui relève, selon moi, d'un argument et non d'un exemple. De plus, cet argument en faveur du mariage forcé n'est pas présent dans la thèse. La conclusion est en lien avec les arguments développés précédemment. Quelques erreurs de syntaxe, de ponctuation et d'orthographe sont présentes.		6/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - <input type="checkbox"/> 1 0-1-2- <input type="checkbox"/> 3 0-1- <input type="checkbox"/> 2-3 0- <input type="checkbox"/> 1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 <input type="checkbox"/> 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	<input type="checkbox"/> 2 0
3. Cohérence textuelle	3.4 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1- <input type="checkbox"/> 2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1- <input type="checkbox"/> 2-3 0-1-2- <input type="checkbox"/> 3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	<input type="checkbox"/> 3 2 0

4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	<input checked="" type="checkbox"/> 6 4 2 0
	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 <input checked="" type="checkbox"/> 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	<input checked="" type="checkbox"/> 3 0
TOTAL			34/50

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?

Les 4 critères utilisés sont compréhensibles et ne comportent pas d'ambiguïté

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

Pour le **point 1** « Impression globale », j'expliquerai peut-être ce qu'on entend par « thèse » (mettre quelques mots entre parenthèses comme pour « argumentation » ou « conclusion ») si la grille est donnée à un enseignant tout venant (donc pas un professeur de français du secondaire supérieur).

Pour les autres critères, les mots utilisés sont compréhensibles et appropriés.

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

Pour le **point 1** « Impression globale », je n'avais pas compris qu'il fallait indiquer quelque chose dans la case. C'est uniquement lorsque j'ai vu la notation « /10 » que j'ai cru comprendre que nous devons rédiger un texte afin de donner notre impression globale sur le travail de l'élève. Je ne vois pas très bien ce que je pourrais indiquer dans la case : j'ai l'impression que mon commentaire va reprendre les différents éléments à cocher. Je ne vois pas non plus comment noter ce point puisque je trouve que c'est assez vague et j'aurais envie de me référer à des critères pour donner une note (critères ci-dessous en réalité).

Je pense que pour remédier à ce souci, je n'attribuerai pas de point à cette case et je la positionnerai à la fin de la grille, plus pour avoir un avis général sur la production (un feedback écrit à destination de l'élève).

Pour le **point 2.2** « Pertinence de l'argumentation », j'ai rencontré des difficultés dans le système de notation parce que je trouvais que deux arguments sur trois étaient recevables (le troisième était un argument plutôt en faveur alors que dans la thèse, ce n'était pas mentionné), or nous devons noter si tous étaient recevables ou aucun.

Je pense que pour remédier à ce souci, je noterai uniquement « Arguments recevables (bien choisis par rapport à la thèse) » et j'ajouterai un système de notation « 0-1-2-3 », afin que les enseignants puissent choisir si aucun, un, deux ou tous les arguments sont recevables.

Pour les autres critères, le système de notation est selon moi approprié.

4) Le système de notation est-il compréhensible ?

Le système de notation est compréhensible, cependant je ne vois pas trop l'intérêt de cocher dans la case « niveaux de maîtrise » : nous entourons déjà dans la case « pondération » les notes que nous souhaitons attribuer aux critères. Nous pouvons donc déjà voir si le critère est présent ou pas.

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

N'étant pas enseignante de français dans le secondaire supérieur, j'ai éprouvé une difficulté à évaluer la présence de la thèse puisque je ne me rappelais pas concrètement de cette notion.

Critères, indicateurs, niveaux de maîtrise et pondération (participant 3)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher) ¹⁴	PONDERATION (à entourer + faire total)
1. Impression globale			7/10 ¹⁵
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input checked="" type="checkbox"/> Présence de la thèse <input checked="" type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input checked="" type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - 1 0-1-2- 3 0-1- 2 -3 ¹⁶ 0- 1
	2.2 Pertinence de l'argumentation	<input checked="" type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input checked="" type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	3.5 Segmentation en paragraphe	<input checked="" type="checkbox"/> Un argument par paragraphe	0-1-2- 3
	3.2 Développement des arguments	<input checked="" type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input checked="" type="checkbox"/> Les exemples permettent de développer les arguments	0-1- 2 -3 0-1- 2 -3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input checked="" type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0

¹⁴ J'ai été embêtée par le fait de devoir cocher les cases. Pour certains critères, le niveau de maîtrise n'est pas total donc je ne savais pas si je devais quand même cocher la case...

¹⁵ Je valide !! Ça permet d'évaluer sans tenir compte des coquilles qui, des fois, peuvent amener l'enseignant à dévaloriser un texte qui, sans faute, pourrait être extra niveau contenu.

¹⁶ J'aime bien l'idée de ne pas pénaliser totalement l'élève s'il lui manque par exemple un argument ou une justification.

4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input checked="" type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	6 ¹⁸ 4 2 0
	4.2 Correction de la ponctuation et de la syntaxe ¹⁷ (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input checked="" type="checkbox"/> Plus de 2 erreurs	5 0
	4.3 Correction du vocabulaire	<input checked="" type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			38/50 ¹⁹

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?

Voir point en dessous

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

Voir point 4.1

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

Voir point 4.2

4) Le système de notation est-il compréhensible ?

J'ai été embêtée par le fait de devoir cocher les cases. Pour certains critères, le niveau de maîtrise n'est pas total donc je ne savais pas si je devais quand même cocher la case...

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ? Globalement, ta grille est bien ! Maintenant, je suis instit primaire donc, les grilles d'évaluation, j'en ai vu défiler et je sais les « comprendre » facilement. Ce serait peut-être intéressant d'avoir l'avis de quelqu'un qui n'est pas du tout dans le contexte de l'enseignement...

¹⁷ Je trouve que ce serait plus intéressant de séparer les « sous-critères » de ce point. Au moins, si l'élève a accès à la grille d'évaluation, il sait dire s'il est plus faible en ponctuation ou en syntaxe. Ça permettrait aussi à l'enseignant de peut-être lui proposer des pistes de remédiation.

Ici, un élève qui maîtrise peut-être parfaitement la syntaxe va se retrouver avec un 0/5 parce qu'il n'a mis aucune virgule dans son texte. Je trouve ça un peu « dommage ».

¹⁸ Je ne comprends pas pourquoi tu as à chaque fois mis « correction de ... » Pourquoi pas simplement « Orthographe », « syntaxe », « vocabulaire » ?

¹⁹ Globalement, ta grille est bien ! Maintenant, je suis instit primaire donc, les grilles d'évaluation, j'en ai vu défiler et je sais les « comprendre » facilement. Ce serait peut-être intéressant d'avoir l'avis de quelqu'un qui n'est pas du tout dans le contexte de l'enseignement...

Critères, indicateurs, niveaux de maîtrise et pondération (participant 4)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			6/10 ²⁰
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 ① 0-1-2 ③ 0-1 ②-3 0 ①
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 ② 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 ①
3. Cohérence textuelle	3.6 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1 ②-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0 ①-2-3 0-1 ②-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	③ 2 0

²⁰ Difficile de donner une note... De plus, l'impression peut dépendre d'un enseignant à l'autre. Ce critère pourrait être gardé mais alors on pourrait peut-être donner des sous-critères qui permettraient de s'y retrouver et avoir des « repères » comme pour la correction de l'orthographe.

²¹ Je trouve que c'est difficile de choisir entre les deux extrêmes pour ce cas-ci.

4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. ²³ <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	⑥ 4 2 0
	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	⑤ 0
	4.3 Correction du vocabulaire ²²	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	③ 0
TOTAL			39/50

Laboratoire cognitif

- 1) **Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?**
Voir remarques critère 1 et critère 4.2.
- 2) **Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?**
Oui ils sont opportuns.
- 3) **Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)**
Voir remarque critère 2.2.
- 4) **Le système de notation est-il compréhensible ?**
Oui.
- 5) **Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?**
A part les différents points plus problématiques que j'ai mis en évidence, non.

²² J'ai trouvé l'exercice difficile pour ces deux derniers points comparé aux critères de l'orthographe qui est clair. Il suffit d'appliquer les règles orthographiques et puis de compter le nombre de fautes. Pour moi, les critères sont assez subjectifs et cela dépend un peu de la personne qui corrige (par exemple : la relation entre deux mots pourraient convenir pour une personne mais pas pour une autre. Le vocabulaire pourrait être adéquat pour un enseignant mais pas pour un autre, il aurait choisi un autre mot).

²³ Cette manière de donner un « repère » pour évaluer est vraiment bien. Ça nous permet d'être juste avec chaque élève. Il n'y a pas de place pour la subjectivité.

Critères, indicateurs, niveaux de maîtrise et pondération (participant 5)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			7/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - <input type="checkbox"/> 1 0-1-2- <input type="checkbox"/> 3 0-1- <input type="checkbox"/> 2-3 0- <input type="checkbox"/> 1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	<input type="checkbox"/> 4 ²⁴ 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 <input type="checkbox"/> 0
3. Cohérence textuelle	3.7 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2- <input type="checkbox"/> 3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1- <input type="checkbox"/> 2-3 0-1- <input type="checkbox"/> 2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	<input type="checkbox"/> 3 2 0
4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes	<input type="checkbox"/> 6 4 2

²⁴ J'ai trouvé l'exercice difficile pour ces deux derniers points comparé aux critères de l'orthographe qui est clair. Il suffit d'appliquer les règles orthographiques et puis de compter le nombre de fautes. Pour moi, les critères sont assez subjectifs et cela dépend un peu de la personne qui corrige (par exemple : la relation entre deux mots pourraient convenir pour une personne mais pas pour une autre. Le vocabulaire pourrait être adéquat pour un enseignant mais pas pour un autre, il aurait choisi un autre mot).

	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> Moins de 80% de formes correctes <input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	0 5 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			37/50

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?

Pour le critère 4.3, peut-être ajouter des exemples comme pour le 4.2. Est-ce que c'est l'utilisation d'un vocabulaire qui est adapté au public ? Vocabulaire soutenu ? Bon choix de mots ?

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ? Les mots me semblent bien choisis.

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

Le point « présence de la thèse » ne doit pas exister, si la thèse n'est pas présente, on ne sait pas la corriger et l'épreuve est annulée mais si elle est présente, la correction peut avoir lieu et je n'attribuerais pas de point pour la simple présence de la thèse.

2.2. → J'aurais bien voulu choisir entre 0 et 4 pour noter ce point. On pourrait considérer qu'un seul des arguments n'est pas acceptable, que fait-on dans ce cas là ? Soit on met « 4 » mais pour moi ça ne correspond pas à la copie ou on met « 0 » et ce n'est pas juste car certains arguments sont acceptables. Si 3 arguments → je mettrais le critère sur trois points → 0 – 1 – 2 – 3

Je trouve qu'il y a beaucoup de points accordés à la ponctuation (4.2.). A nouveau, je trouve la transition entre le « 5 » et le « 0 » radicale. L'étudiant qui fait 2 erreurs à tous les points et celui qui fait seulement une faute de plus perd 5 points d'un coup. → PISTE : 0 erreur de ponctuation → 3 / 1 erreur → 2 / 2 erreurs → 1 / 3 erreurs ou plus → 0 (par exemple).

4) Le système de notation est-il compréhensible ?

Par ex pour le critère 2.3. → je mettrais dans la colonne de pondération « 0 – 2 ». S'ils sont sur la même ligne, on est sûr que le correcteur n'entour qu'un des deux. Normalement si on lit bien la grille, on comprend que ce n'est qu'un seul à entourer mais ce serait plus visuel comme ça enfin je trouve.

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

Pour le premier critère « impression globale » → OUI. Car il n'y a pas de critères de correction pour justifier un « 10 », un « 5 » ou un « 0 ». L'impression globale → BCP trop subjectif. Du coup, j'ai un peu tapé dans le tas pour ce critère car je manquais d'infos pour corriger. J'ai attribué 7/10 → je suis sûr que ma note est totalement subjective.

Critères, indicateurs, niveaux de maîtrise et pondération (participant 6)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			6/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - 1 0-1-2-3 0-1-2-3 0-1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	4.4 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0
4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	6 4 2 0

	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			31/50

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïtés ?

- Pour le premier critère (impression globale) : à quoi correspond « l'impression générale » ? Je trouve que c'est fort subjectif.
- Pour le deuxième critère (respect de l'intention de convaincre) : dans les consignes l'évaluateur parle d'introduction et puis dans la grille on parle de thèse. Il me semble qu'il faudrait uniformiser le propos et donner un mot d'explication afin que l'élève soit au clair avec les attendus de sa dissertation.
L'évaluateur parle également d'arguments. Compte-t-il le contre argument dans les trois « arguments » demandés ? Autorise-t-il les contre arguments ? Je ne comprends pas pourquoi les critères sont en liens les uns avec les autres. L'évaluateur parle de la thèse et des arguments en fonction de celle-ci. Je pensais que les critères (sous critères également) devaient être indépendants des uns des autres... Pourquoi aussi passer de plus de 4 (0-1-2-3) propositions de points et puis plus que 2 (4-0) ? Pour moi, il devrait y avoir une note intermédiaire entre le 4 et le 0. Pourquoi parler des différents avis ? Cela veut-il dire que les élèves peuvent utiliser des contre arguments ?
- Pour le troisième critère (cohérence textuelle) : Je n'ai pas grand-chose à ajouter. Cela me semble clair et cohérent.
- Pour le quatrième critère (respect des normes linguistiques) : Pas facile d'évaluer le pourcentage. Il faudrait donner un autre indicateur que le pourcentage. Peut-être parler en nombre de fautes ? Même remarque que pour le critère deux : pourquoi ne pas avoir des cotations intermédiaires au lieu du 5 et du 0 par exemple ?

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

Pour moi, il faut davantage préciser le mot thèse ou bien appeler cela introduction et être en cohérence avec les consignes que les élèves ont reçues. En effet, dans leurs consignes, on parle de l'introduction et pas de la thèse. S'ils ont accès à la grille de correction, ils ont très certainement dû demander des éclaircissements du terme « thèse ». Et puis le terme « impression générale » fait appel à beaucoup de subjectivité je trouve... Quand pouvons nous conclure qu'un argument est recevable ou non ?

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

Je pense qu'il ne faut pas utiliser des « points binaires ». C'est-à-dire des propositions à deux choix, soit 0 ou 5. Il faut des catégories intermédiaires pour moi. Et puis peut-être permettre à l'élève d'avoir des le moyen de s'auto évaluer et de lui prévoir des critères. Cela pourrait être utile autant pour l'élève que pour l'enseignant.

4) Le système de notation est-il compréhensible ?

Non. J'ai bien compris qu'il fallait entourer ce qui nous semble représenter chaque critère, mais je ne comprends la cotation de « impression générale ». Quels sont les critères pour plus d'objectivité ? Mais pour moi, à partir du moment où on demande des impressions, nous sommes dans la subjectivité. Est-ce que les 10 points sont à part des 50 ou bien compris dedans ? Moi je ne les ais pas compris dedans car le critère 1 a l'air à part et ne pas être sur le même pied que les autres critères. De plus, le concepteur de la grille parle de pondération, mais nous n'avons aucun moyen de savoir comment la réaliser ? S'agit-il d'une moyenne ? Quel calcul effectué ?

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

Je me suis questionnée à plusieurs endroits ne sachant pas trop quelle note donner. Je pense aussi que c'est parce qu'elle manque de clarification dans les concepts ou bien peut-être d'exemples concrets comme nous retrouvons sur les corrections d'épreuves (non) certificatives réalisées en Fédération Wallonie Bruxelles (dans le carnet de corrections).

Critères, indicateurs, niveaux de maîtrise et pondération²⁵ (participant 7)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			6/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 - <u>1</u> 0-1-2- <u>3</u> 0-1-2- <u>3</u> 0- <u>1</u>
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	$\frac{4}{0}$ ²⁶
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	$\frac{2}{0}$
3. Cohérence textuelle	4.2 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2- <u>3</u>
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1- <u>2</u> -3 0-1- <u>2</u> -3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 $\frac{2}{0}$

²⁵ J'ai trouvé l'exercice difficile pour ces deux derniers points comparé aux critères de l'orthographe qui est clair. Il suffit d'appliquer les règles orthographiques et puis de compter le nombre de fautes. Pour moi, les critères sont assez subjectifs et cela dépend un peu de la personne qui corrige (par exemple : la relation entre deux mots pourraient convenir pour une personne mais pas pour une autre. Le vocabulaire pourrait être adéquat pour un enseignant mais pas pour un autre, il aurait choisi un autre mot).

²⁶ Le deuxième argument, pour moi, est non recevable car il n'est pas prouvé. Le choix d'un mari par le père ne peut pas assurer un mariage sans violence.

4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	6 <u>4</u> 2 0
	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 <u>0</u>
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	<u>3</u> 0
TOTAL			34 /50

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïté.s ?

Oui compréhensibles, et non ils ne comportent aucune ambiguïté.

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

Le mot thèse peut être interprété de plusieurs façon, comme étant le sujet de la dissertation ou comme étant la présence de l'opinion de l'étudiant dans l'introduction.

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

Le point sur la recevabilité des arguments devrait être plus large. La consigne demande 3 arguments, la cotation devrait être alors divisée en 3 (1 point par argument recevable par exemple et -1/0 point par argument non recevable).

Réaliser un travail d'argumentation est fort personnel, comment évaluer la recevabilité d'un argument ? Avec quels critères ? Il y a-t-il une limite à ne pas franchir sur ce sujet ? Ce critère devait peut être être plus précis, par exemple « les arguments sont-ils vérifié par de la documentation, par des interviews ? »

La correction de l'orthographe pourrait être évaluée de la façon suivante « de 0 à 1 faute : 6 points, de 2 à 3 fautes : 5 points, etc. »

La correction de la syntaxe ne me semble pas assez souple pour un étudiant de troisième année. Prévoir de scindé les points ; sens → 1, ordre des mots → 1, relation entre les mots → 1 ? etc

4) Le système de notation est-il compréhensible ?

Oui mais devrait être plus précis.

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

Oui car la dissertation est fort brève. La manière d'écrire de l'étudiant est claire mais les arguments ne sont pas assez approfondis et son avis est déjà tranché alors même que la thèse n'a pas fini d'être énoncée.

Critères, indicateurs, niveaux de maîtrise et pondération (participant 8)

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Impression globale			4/10
2. Respect de l'intention : convaincre	2.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 -1 0-1-2-3 0-1-2-3 0-1
	2.2 Pertinence de l'argumentation	<input type="checkbox"/> Arguments recevables (bien choisis par rapport à la thèse). <input type="checkbox"/> Arguments non recevables (mal choisis par rapport à la thèse)	4 0
	2.3 Cohérence de l'argumentation	<input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	2 0
3. Cohérence textuelle	4.3 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2-3
	3.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	3.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0
4. Respect des normes linguistiques	4.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes.	6
		<input type="checkbox"/> De 90% à 94% de formes correctes	4
		<input type="checkbox"/> De 80% à 89% de formes correctes	2
		<input type="checkbox"/> Moins de 80% de formes correctes	0

	4.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	5 0
	4.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
TOTAL			26/50

Laboratoire cognitif

1) Est-ce que les critères utilisés sont compréhensibles ou comportent-ils une des ambiguïté.s ?

Je ne relève aucune réelle ambiguïté au niveau de la catégorie des critères.

2) Les mots utilisés pour formuler chacun des critères sont-ils opportuns, ou faudrait-il remplacer certains par d'autres plus appropriés ?

Ici, vous indiquez le mot « critères ». Je me réfère donc à la première colonne. Or, vous ne parlez pas des indicateurs ou des niveaux de maîtrise. Par « critères », entendez-vous l'ensemble des colonnes ?

Si non, je n'observe aucun problème de formulation dans la première colonne.

3) Le système de notation attribué à chaque critère est-il approprié ? (Si pas, montrez comment vous changeriez cela)

À mes yeux, le système utilisé est trop catégorique. Il ne laisse pas place à un continuum plus large des points. Lorsque l'on est face à la cotation de la présence d'un élément, il est normal d'avoir un système binaire (c'est présent ou ça ne l'est pas) mais le 2.2., par exemple, est trop catégorique à mes yeux : soit tout est valable, soit tout ne l'est pas. Or, j'ai trouvé qu'il y avait des nuances à apporter au niveau de la rédaction que je viens de corriger. Il y a des contradictions mais tout n'est pas réellement à jeter à la poubelle.

À mes yeux, l'élève dont je viens de corriger la copie ne devrait pas avoir la moyenne dans le sens où le but de l'argumentation est de convaincre, d'exemplifier, de détailler. Ici, les développements sont très pauvres et les quelques phrases se contredisent parfois. Or, il obtient une note de +1 la moyenne. Cela est dû au fait de la présence d'une pondération accordée trop importante pour tout ce qui est « présence de... ». Je comptabilise 11 points, rien que pour la présence des éléments demandés. Ajoutons à cela les 10 points de l'appréciation globale, cela fait 2/5^e de la note consacrée à de l'appréciation ou à de la simple présence d'éléments demandés. Pour moi, cela n'évalue pas assez la pertinence des éléments. J'accorderais donc une plus grande importance (avec plus de possibilités d'accorder des points médians et non juste un système binaire). Enfin, je placerais l'appréciation globale à la fin de la grille de correction car l'on pourrait avoir, à mon sens, moins de biais de la part du correcteur s'il est amené à évaluer après s'être basé sur des critères objectifs. Placer cet item au départ, qui plus est sur 10 points (que je ramènerais à 4 points d'ailleurs car 1/5^e du travail, cela me paraît tout de même beaucoup), sans s'imprégner de la grille, amène davantage de subjectivité.

Ici-bas, je propose une autre pondération et un changement dans la grille (modifications en rouge).

4) Le système de notation est-il compréhensible ?

Oui, il l'est. C'est en termes de cohérence que j'émet des réserves (explicitées plus haut).

5) Avez-vous éprouvé des difficultés pour évaluer cette dissertation ?

Oui, pour les raisons exactes et exhaustives citées précédemment.

Enfin, pourquoi devoir cocher les niveaux de maîtrise alors que la pondération est présente ? Est-ce utile ? C'est ambigu.

CRITERES	INDICATEURS	NIVEAUX DE MAITRISE (à cocher)	PONDERATION (à entourer + faire total)
1. Respect de l'intention : convaincre	1.1 Le texte intègre toutes les caractéristiques d'une argumentation	<input type="checkbox"/> Présence de la thèse <input type="checkbox"/> Présence de 3 arguments développés <input type="checkbox"/> Présence d'une justification pour chaque argument (exemples) <input type="checkbox"/> Présence d'une conclusion (opinion globale)	0 -1 0-1-2-3 0-1-2-3 0-1
	1.2 Pertinence de l'argumentation	<input type="checkbox"/> La majorité des arguments est recevable (bien choisis par rapport à la thèse). <input type="checkbox"/> Une partie des arguments est recevable	3 2
	1.3 Cohérence de l'argumentation	<input type="checkbox"/> La majorité des arguments est irrecevable (mal choisis par rapport à la thèse) <input type="checkbox"/> Pas de contradiction(s) entre la conclusion et les divers avis. <input type="checkbox"/> Contradiction(s) entre la conclusion et les divers avis	0 2 0
2. Cohérence textuelle	2.1 Segmentation en paragraphe	<input type="checkbox"/> Un argument par paragraphe	0-1-2-3
	2.2 Développement des arguments	<input type="checkbox"/> Les arguments sont bien formulés, clairs, précis <input type="checkbox"/> Les exemples permettent de développer les arguments	0-1-2-3 0-1-2-3
	2.3 Pertinence de la connexion (connecteurs logiques)	<input type="checkbox"/> Utilisation de connexions logiques correctes et efficaces <input type="checkbox"/> Une ou deux erreur(s) de connexion <input type="checkbox"/> Plus de deux erreurs de connexion	3 2 0
3. Respect des normes linguistiques	3.1 Correction de l'orthographe	<input type="checkbox"/> 95% et plus de formes correctes. <input type="checkbox"/> De 90% à 94% de formes correctes <input type="checkbox"/> De 80% à 89% de formes correctes <input type="checkbox"/> Moins de 80% de formes correctes	6 4 2 0

	3.2 Correction de la ponctuation et de la syntaxe (sens, présence et ordre des mots, relation entre les mots, etc.)	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
	3.3 Correction du vocabulaire	<input type="checkbox"/> De 0 à 2 erreurs <input type="checkbox"/> Plus de 2 erreurs	3 0
4.Impression globale			/4
TOTAL			/40

Je tiens à préciser que cette conception est subjective et qu'il faudrait la travailler en collaboration afin d'obtenir une fidélité inter-codeurs qui donne plus de crédits à cette grille.

Annexe 15. Formulaire de consentement



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

Comité d'éthique

PRESIDENTE : Fabienne COLLETTE

SECRETAIRE : Annick COMBLAIN

CONSENTEMENT ECLAIRE

POUR DES RECHERCHES IMPLIQUANT DES PARTICIPANTS HUMAINS

Un exemplaire du formulaire d'information et de consentement signé doit être donné au participant, et le chercheur doit en conserver une copie. Le consentement étant nominatif, il doit être considéré comme une donnée sensible et devra être protégé contre toute « violation ». Cela implique qu'il soit conservé sous clé avec des accès réglementés et qu'il soit également détruit en cas de demande d'effacement (sauf raison à justifier).

CE-Cons_écl-1	Participant adulte, pas de transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.
CE-Cons_écl-2	Participant adulte, transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.
CE-Cons_écl-3	Participant mineur ou dans l'incapacité de donner son consentement, pas de transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.
CE-Cons_écl-4	Participant mineur ou dans l'incapacité de donner son consentement, avec transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.
CE-Cons_écl-5	Accord pour enregistrement vidéo de personnes n'étant pas des participants directs à l'étude mais se trouvant dans le champ de la caméra.
CE-Cons_écl-6	Participation via internet, regroupant le formulaire d'information et de consentement. Pour les mineurs, inclure la procédure permettant d'avoir l'accord parental et modifier la partie du consentement en vert.
CE-Cons_écl-7	Information et consentement oraux adaptés aux participants mineurs ou présentant des difficultés de compréhension.
CE-Cons_écl-8	Information et consentement via <i>voie téléphonique</i> . Participant adulte, pas de transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.
CE-Cons_écl-9	Information et consentement <i>oral</i> . Participant adulte, pas de transmission de résultats vers un professionnel des soins de santé, avec ou sans enregistrement audio/vidéo.

CONSENTEMENT ECLAIRE

POUR DES RECHERCHES IMPLIQUANT DES PARTICIPANTS HUMAINS

Titre de la recherche	Mesurer les déterminants de la notation dans le cadre de l'évaluation d'une rédaction en vue de modéliser les pratiques évaluatives.
Chercheur responsable	Rauw Kim
Promoteur	Ariane Baye
Service et numéro de téléphone de contact	Département dans les Sciences de l'Éducation (Analyse et Intervention dans le domaine du Décrochage et de l'Exclusion) +32 43 66 20 63

- Je, *soussigné(e)*

déclare :
- avoir reçu, lu et compris une présentation écrite de la recherche dont le titre et le chercheur responsable figurent ci-dessus ;
- avoir pu poser des questions sur cette recherche et reçu toutes les informations que je souhaitais.
- avoir reçu une copie de l'information au participant et du consentement éclairé.

J'ai compris que :

- je peux à tout moment mettre un terme à ma participation à cette recherche sans devoir motiver ma décision ni subir aucun préjudice que ce soit. Les données codées acquises resteront disponibles pour traitements statistiques.
- je peux demander à recevoir les résultats globaux de la recherche mais je n'aurai aucun retour concernant mes performances personnelles.
- je peux contacter le chercheur pour toute question ou insatisfaction relative à ma participation à la recherche.

- des données me concernant seront récoltées pendant ma participation à cette étude et que le chercheur/mémorant responsable et le promoteur de l'étude se portent garants de la confidentialité de ces données. Je conserve le droit de regard et de rectification sur mes données personnelles (données démographiques). Je dispose d'une série de droits (accès, rectification, suppression, opposition) concernant mes données personnelles, droits que je peux exercer en prenant contact avec le Délégué à la protection des données de l'institution dont les coordonnées se trouvent sur la feuille d'information qui m'a été remise. Je peux également lui adresser toute doléance concernant le traitement de mes données à caractère personnel. **Je dispose également du droit d'introduire une réclamation auprès de l'Autorité de protection des données (<https://www.autoriteprotectiondonnees.be>, contact@apd-gba.be).**
- les données à caractère personnel ne seront conservées que le temps utile à la réalisation de l'étude visée, c'est-à-dire durant deux années.

Je consens à ce que :

- les données anonymes recueillies dans le cadre de cette étude soient également utilisées dans le cadre d'autres études futures similaires, y compris éventuellement dans d'autres pays que la Belgique.
- les données anonymes recueillies soient, le cas échéant, transmises à des collègues d'autres institutions pour des analyses similaires à celles du présent projet ou qu'elles soient mises en dépôt sur des répertoires scientifiques accessibles à la communauté scientifique uniquement.
- mes données personnelles soient traitées selon les modalités décrites dans la rubrique traitant de garanties de confidentialité du formulaire d'information.

En conséquence, je donne mon consentement libre et éclairé pour être participant à cette recherche.

Lu et approuvé,

Date et signature


Chercheur responsable

- Je soussigné, Rauw Kim, chercheur responsable, confirme avoir fourni oralement les informations nécessaires sur l'étude et avoir fourni un exemplaire du document d'information et de consentement au participant.
- Je confirme qu'aucune pression n'a été exercée pour que la personne accepte de participer à l'étude et que je suis prêt à répondre à toutes les questions supplémentaires, le cas échéant.
- Je confirme travailler en accord avec les principes éthiques énoncés dans la dernière version de la « Déclaration d'Helsinki », des « Bonnes pratiques Cliniques » et de la loi belge du 7 mai 2004, relative aux expérimentations sur la personne humaine, ainsi que dans le respect des pratiques éthiques et déontologiques de ma profession.

Nom, prénom du chercheur responsable : RAUW Kim

Date et signature :

29/10/21



Annexe 16. Formulaire d'information



Faculté de Psychologie, Logopédie et des Sciences de l'Éducation
Comité d'éthique
PRESIDENTE : Fabienne COLLETTE
SECRETAIRE : Annick COMBLAIN

Formulaire d'information au volontaire

TITRE DE LA RECHERCHE

Mesurer les facteurs qui interviennent dans l'évaluation d'une production écrite, en vue d'améliorer les pratiques évaluatives

CHERCHEUR / ETUDIANT RESPONSABLE

Nom/prénom : Rauw Kim

Statut : Étudiante en dernière année Master en Sciences de l'Éducation ;

Téléphone : 0491180306 ;

Email : rauwkim@gmail.com.

PROMOTEUR

Ariane Baye

Université de Liège

Service : Sciences de l'Éducation

Adresse : Ariane.Baye@uliege.be

DESCRIPTION DE L'ÉTUDE

L'objectif général de cette étude est de mesurer les facteurs qui interviennent dans l'évaluation d'une production écrite, en vue d'améliorer les pratiques évaluatives. Pour ce faire, il me faut recueillir un échantillon d'enseignants secondaire qui accepteront de participer à deux épreuves en ligne : d'une part de corriger une rédaction (accompagnée ou non d'une grille d'évaluation) (+-25 min) et d'autre part de répondre à un questionnaire contextuel (+-5 min) afin de recueillir certaines informations sur les participants telles que leur âge, leur nombre d'années d'expérience dans l'enseignement, etc. Afin d'assurer la bonne réalisation de ces épreuves, il s'agira de réaliser ces dernières seule, et si possible, dans un endroit calme.

Vos données personnelles (c'est-à-dire les données qui permettent de vous identifier comme votre nom ou vos coordonnées) seront conservées durant la réalisation de l'étude dans un endroit sûr durant 2 ans, après quoi elles seront détruites.

Avant de participer à l'étude, nous attirons votre attention sur un certain nombre de points.

Votre participation est conditionnée à une série de droits pour lesquels vous êtes couverts en cas de préjudices. Vos droits sont explicités ci-dessous.

- Votre participation est libre. Vous pouvez l'interrompre sans justification.
- Aucune divulgation de vos informations personnelles n'est possible même de façon non intentionnelle. Seules les données codées pourront être transmises à la communauté des chercheurs. Ces données codées ne permettent plus de vous identifier et il sera impossible de les mettre en lien avec votre participation.
- Le temps de conservation de vos données personnelles est réduit à son minimum. Par contre, les données codées peuvent être conservées *ad vitam aeternam*.
- Les résultats issus de cette étude seront toujours communiqués dans une perspective scientifique et/ou d'enseignement.
- En cas de préjudice, sachez qu'une assurance vous couvre.
- Si vous souhaitez formuler une plainte concernant le traitement de vos données ou votre participation à l'étude, contactez le responsable de l'étude et/ou le DPO et/ou le Comité d'éthique (cf. adresses à la fin du document).

Tous ces points sont détaillés aux pages suivantes. Pour toute autre question, veuillez vous adresser au chercheur ou au responsable de l'étude. Si ces informations sont claires et que vous souhaitez participer à l'étude, nous vous invitons à signer le formulaire de consentement. Conservez bien une copie de chaque document transmis afin de pouvoir nous recontacter si nécessaire.

INFORMATIONS DETAILLEES

Toutes les informations récoltées au cours de cette étude seront utilisées dans la plus stricte confidentialité et seuls les expérimentateurs, responsables de l'étude, auront accès aux données récoltées. Vos informations seront codées. Seul le responsable de l'étude ainsi que la personne en charge de votre suivi auront accès au fichier crypté permettant d'associer le code du participant à son nom et prénom, ses coordonnées de contact et aux données de recherche. Ces personnes seront tenues de ne JAMAIS divulguer ces informations.

Les données codées issues de votre participation peuvent être transmises dans le cadre d'une autre recherche en lien avec cette étude-ci. Elles pourront être compilées dans des bases de données accessibles uniquement à la communauté scientifique. Seules les informations codées seront partagées. En l'état actuel des choses, aucune identification ne sera possible. Si un rapport ou un article est publié à l'issue de cette étude, rien ne permettra votre identification. Vos données à caractère personnel conservées dans la base de données sécurisée sont soumises aux droits suivants : droits d'accès, de rectification et d'effacement de cette base de données, ainsi que du droit de limiter ou de s'opposer au traitement des données. Pour exercer ces droits, vous devez vous adresser au chercheur responsable de l'étude ou, à défaut, au délégué à la protection des données de l'Université de Liège, dont les coordonnées se trouvent au bas du formulaire d'information. Le temps de conservation de vos données à caractère personnel sera le plus court possible ; durant deux années. Les données issues de votre participation à cette recherche (données codées) seront également conservées durant deux années, après quoi elles seront détruites.

Si vous changez d'avis et décidez de ne plus participer à cette étude, nous ne recueillerons plus de données supplémentaires vous concernant et vos données d'identification seront détruites. Seules les données rendues anonymes pourront être conservées et traitées.

Les modalités pratiques de gestion, traitement, conservation et destruction de vos données respectent le Règlement Général sur la Protection des Données (UE 2016/679), les droits du patient (loi du 22 août 2002) ainsi que la loi du 7 mai 2004 relative aux études sur la personne humaine. Toutes les procédures sont réalisées en accord avec les dernières recommandations européennes en matière de collecte et de partage de données. Le responsable du traitement de vos données à caractère personnel est l'Université de Liège (Place du XX-Août, 7 à 4000 Liège), représentée par son Recteur. Ces traitements de données à caractère personnel seront réalisés dans le cadre de la mission d'intérêt public en matière de recherche reconnue à l'Université de Liège par le Décret définissant le paysage de l'enseignement supérieur et l'organisation académique des études du 7 novembre 2013, art.2. Vous disposez également du droit d'introduire une réclamation auprès de l'Autorité de protection des données (<https://www.autoriteprotectiondonnees.be>, contact@apd-gba.be).

Une assurance a été souscrite au cas où vous subiriez un dommage lié à votre participation à cette recherche. Le promoteur assume, même sans faute, la responsabilité du dommage causé au participant (ou à ses ayants droit) et lié de manière directe ou indirecte à la participation à cette étude. Dans cette optique, le promoteur a souscrit un contrat d'assurance auprès d'Ethias, conformément à l'article 29 de la loi belge relative aux expérimentations sur la personne humaine (7 mai 2004).

Vous signerez un consentement éclairé avant de prendre part à l'expérience. Vous conserverez une copie de ce consentement ainsi que les feuilles d'informations relatives à l'étude.

Cette étude a reçu un avis favorable de la part du comité d'éthique de la faculté de psychologie, logopédie et des sciences de L'Éducation de l'Université de Liège. En aucun cas, vous ne devez considérer cet avis favorable comme une incitation à participer à cette étude.

Personnes à contacter

Vous avez le droit de poser toutes les questions que vous souhaitez sur cette recherche et d'en recevoir les réponses.

Si vous avez des questions ou en cas de complication liée à l'étude, vous pouvez contacter les personnes suivantes :

ARIANE Baye

Téléphone de service : +32 436 620 63

Service : Département dans les Sciences de l'Éducation (Analyse et Intervention dans le domaine du Décrochage et de l'Exclusion)

Email : Ariane.Baye@uliege.be

Adresse : Place des Orateurs 2, 4000 Liège

DYLAN Dacht

Téléphone du service : +32 436 620 63

Service : Département dans les Sciences de l'Éducation (Analyse et Intervention dans le domaine du Décrochage et de l'Exclusion)

Email : ddacht@uliege.be

Adresse : Place des Orateurs 2, 4000 Liège

ou l'investigateur principal du projet :

RAUW Kim

Téléphone: + 32 0491 18 03 06

Email: kim.rauw@student.uliege.be

Pour toute question, demande d'exercice des droits ou plainte relative à la gestion de vos données à caractère personnel, vous pouvez vous adresser au délégué à la protection des données par e-mail (dpo@uliege.be) ou par courrier signé et daté adressé comme suit :

Monsieur le Délégué à la protection des données

Bât. B9 Cellule "GDPR",

Quartier Village 3,

Boulevard de Colonster 2,

4000 Liège, Belgique.

Vous disposez également du droit d'introduire une réclamation auprès de l'Autorité de protection des données (<https://www.autoriteprotectiondonnees.be>, contact@apd-gba.be).

Figure 1. Les étapes du processus d'évaluation des apprentissages (Pellegrino, Chudowsky et Glaser, 2001, cités par Berthiaume et al., 2011)

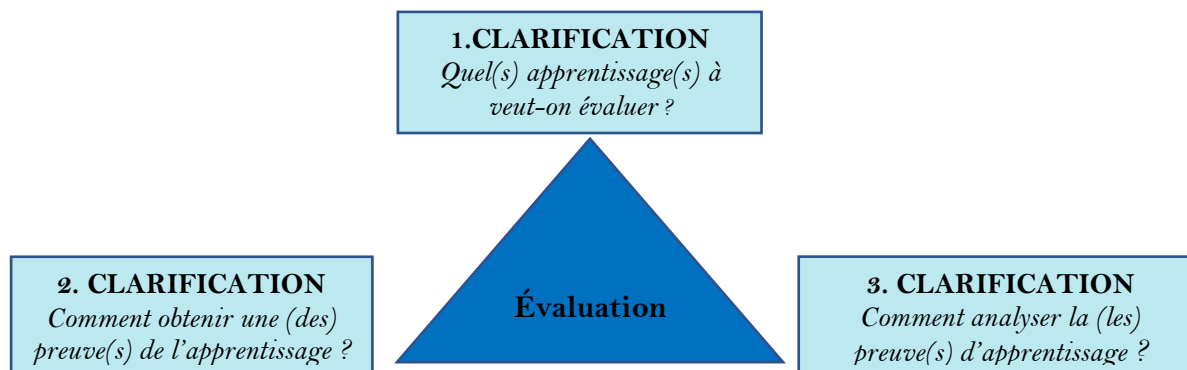


Figure 1. Les étapes du processus d'évaluation des apprentissages (Pellegrino et al., 2001, cités par Berthiaume et al., 2011)

Figure 2. L'usage de la grille critériée inscrite dans le temps (Berthiaume et al., 2011)



Figure 2. L'usage de la grille critériée inscrite dans le temps (Berthiaume et al., 2011)

Tableau 1. Résultats du laboratoire cognitif

Participants	Métier	Note attribuée
Participant n°1	Institutrice primaire	39/50 → 7,4/10
Participant n°2	Institutrice primaire	34/50 → 6,8/10
Participant n°3	Institutrice primaire	38/50 → 7,6/10
Participant n°4	Institutrice primaire	39/40 → 7,8/10
Participant n°5	Instituteur primaire	37/50 → 7,4/10
Participant n°6	Institutrice primaire	31/50 → 6,2/10
Participant n°7	Institutrice primaire	34/50 → 6,8/10
Participant n°8	Professeur de français dans l'enseignement secondaire inférieur.	26/50 → 5,2/10
Moyenne		34,75 → 6,95/10

Tableau 1. Résultats du laboratoire cognitif

Tableau 2. Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes

<i>Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes</i>																														
	SEXE (a)		ANCIENNETÉ (b)					VÉCU PAR RAPPORT REDOUBLEMENT (c)		DISCIPLINE (d)																				
	F	H	A	B	C	D	E	A	B	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
Groupe 1 (N= 35)	26	9	15	7	3	2	8	16	19	9	9	2	1	6		1		1	3	1					1	1				
Groupe 2 (N=35)	24	11	21	5	4	1	4	15	20	8	10	1	4	4	1		1		2		1	1		1					1	
Groupe 3 (N=35)	25	10	13	5	6	3	8	18	17	7	8	2	2	6		2	1	1	2				1						2	1
Groupe 4 (N=35)	29	6	16	4	2	3	10	15	20	14	6		1	7	1	1						1			1	1	1			

Tableau 2. Caractéristiques de l'échantillon d'enseignants répartis dans les quatre groupes

-Note a : F= Femme, H= Homme

-Note b : A = 0an < x ≤ 5ans, B = 5ans < x ≤ 10ans, C = 10ans < x ≤ 15ans, D = 15ans < x ≤ 20ans, E = x > 20ans

-Note c : A = connu le redoublement, B= pas connu le redoublement

-Note d : A = français, B = mathématiques, C = histoire, D = sciences (chimie, biologie, physique), E = langues modernes (anglais, néerlandais, allemand, espagnol, français langue étrangère), F = langues anciennes (latin), G = religion (catholique, protestante etc), H = philosophie et citoyenneté, I = cours artistiques (art plastique, éducation musicale), J = éducation physique, K = agriculture, L = mécanique, M = hôtellerie, N = horticulture, O = informatique, P = économie, Q = sciences économiques, R = sciences humaines, S = sciences sociales, T= agencement intérieur.

Tableau 3. Tableau récapitulatif reprenant les résultats au test de normalité pour les notes totales attribuées à chacun des groupes de copies

Variables	Test	Statistique		Valeur de P	
Copies promus sans grille (groupe 4)	Shapiro-Wilk	W	0,966872	Pr<W	0.3632
Copies promus avec grille (groupe 2)	Shapiro-Wilk	W	0,960966	Pr<W	0.2444
Copies redoublants sans grille (groupe 3)	Shapiro-Wilk	W	0.972903	Pr<W	0.5278
Copies redoublants avec grille (groupe 1)	Shapiro-Wilk	W	0.976834	Pr<W	0.6543

Tableau 3. Tableau récapitulatif reprenant les résultats au test de normalité pour les notes totales attribuées à chacun des groupes de copie

Tableau 4. Tableau récapitulatif reprenant les résultats au test de normalité pour les notes attribuées à chacun des blocs de critères selon qu'il s'agit d'une « copie redoublant » ou d'une « copie promu »

Variables Conditions	B1			B2			B3		
	Test	Statistique (W)	Valeur de P (Pr<W)	Test	Statistique (W)	Valeur de P (Pr<W)	Test	Statistique (W)	Valeur de P (Pr<W)
Promu	Shapiro-Wilk	0,930476	0,0289	Shapiro-Wilk	0,952769	0,1378	Shapiro-Wilk	0,939504	0,0540
Redoublant	Shapiro-Wilk	0,948197	0,0997	Shapiro-Wilk	0,937956	0,0485	Shapiro-Wilk	0,942722	0,0678

Tableau 4. *Tableau récapitulatif reprenant les résultats au test de normalité pour les notes attribuées à chacun des blocs de critères selon qu'il s'agit d'une « copie redoublant » ou d'une « copie promu »*

Tableau 5. Analyse de la variance simple pour la note totale

Source	DF	Somme des carrés	Carré moyen	Valeur de F	Pr > F
Modèle	7	798,61	114,09	3,79	0,0009
Erreur	132	3973,11	30,10		
Total corrigé	139	4771,72			

R-carré	Coefficient de variation	Écart quadratique moyen	Note moyenne
0,16	21,76	5,52	25,37

Tableau 5. *Analyse de la variance simple pour la note totale*

Tableau 6. Analyse de la variance simple pour la note attribuée à chacun des blocs de critères

	B1					B2					B3				
	DF	Somme des carrés	Carré moyen	Valeur de F	Pr > F	DF	Somme des carrés	Carré moyen	Valeur de F	Pr > F	DF	Somme des carrés	Carré moyen	Valeur de F	Pr > F
Modèle	5	69,72	13,94	1,72	0,14	5	28,45	5,69	1,56	0,18	5	50,22	10,04	1,61	0,1702
Erreur	64	519,78	8,12			64	232,82	3,64			64	399,27	6,24		
Total corrigé	69	589,50				69	261,27				69	449,49			

B1				B2				B3			
R-carré	Coefficient de variation	Écart quadratique moyen	Note moyenne	R-carré	Coefficient de variation	Écart quadratique moyen	Note moyenne	R-carré	Coefficient de variation	Écart quadratique moyen	Note moyenne
0,12	27,14	2,85	10,50	0,11	22,29	1,91	8,56	0,11	33,37	2,50	7,49

Tableau 6. Analyse de la variance simple pour la note attribuée à chacun des blocs de critère

Tableau 7. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée à la rédaction selon la présence ou non du stéréotype relatif au statut redoublant

Redoublant	N	Note totale attribuée à la rédaction	
		Moyenne (écart-type)	Ampleur de l'effet (1) vs (2)
Oui (1)	70	24,42 (5,47)	-0,33
Non (2)	70	26,33 (6,11)	

Tableau 7. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée à la rédaction selon la présence ou non du stéréotype relatif au statut redoublant

Tableau 8. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale

	DL	Valeur de F	PR>F
Redoublant	1	3,45	0,0654
Grille	1	2,76	0,0990
Redoublant * grille	1	0,05	0,8320
Ancienneté	1	5,96	0,0160
Connu le redoublement	1	7,73	0,0062
Genre	1	0,51	0,4781
Français	1	4,25	0,0411

Tableau 8. Résultats des sommes de carrés de type III de l'analyse de variance simple sur la note totale

Tableau 9. Moyennes, écarts-types et ampleurs de l'effet des notes attribuées à chacun des blocs de critères de la grille d'évaluation selon la présence ou non du stéréotype relatif au statut redoublant

Variables	B1			B2			B3		
	N	μ (σ)	Ampleur de l'effet (1) vs (2)	N	μ (σ)	Ampleur de l'effet (1) vs (2)	N	μ (σ)	Ampleur de l'effet (1) vs (2)
Redoublant (1)	35	10,34 (2,85)	-0,11	3	8,86	0,31	35	6,69	-0,66
				5	(2,29)			(2,73)	
Promu (2)	35	10,66 (3,03)		3	8,26		35	8,29	
				5	(1,50)			(2,11)	

Tableau 9. Moyennes, écarts-types et ampleurs de l'effet des notes attribuées à chacun des blocs de critères de la grille d'évaluation selon la présence ou non du stéréotype relatif au statut redoublant

Tableau 10 : Tableau récapitulatif reprenant les résultats des sommes de carrés de type III de l'analyse de variances simples pour chacun des blocs de critères

Variables Conditions	B1			B2			B3		
	DL	Valeur de F	PR > F	DL	Valeur de F	PR > F	DL	Valeur de F	PR > F
Redoublant	1	0,39	0,5361	1	1,31	0,2564	1	7,36	0,0086
Ancienneté	1	0,62	0,4344	1	0,16	0,6870	1	0,35	0,5537
Connu le redoublement	1	0,92	0,3418	1	2,76	0,1014	1	0,49	0,4871
Genre	1	0,04	0,8516	1	2,55	0,1153	1	0,00	0,9748
Français	1	6,40	0,0139	1	0,93	0,3395	1	0,02	0,8752

Tableau 10. Tableau récapitulatif reprenant les résultats des sommes de carrés de type III de l'analyse de variances simples pour chacun des blocs de critères

Tableau 11. Calcul de la variance des notes totales selon la présence ou non d'une grille d'évaluation

Conditions Variable	Pas grille			Grille		
	N	μ (σ)	σ^2	N	μ (σ)	σ^2
Note totale	70	24,76 (6,18)	38,23	70	25,99 (5,49)	30,16

Tableau 11. Calcul de la variance des notes totales selon la présence ou non d'une grille d'évaluation

Tableau 12. Moyenne et écarts-types de la note totale pour chacune des 4 copies

Conditions Variable	Grille et redoublant (1)		Grille et promu (2)		Pas de grille et redoublant (3)		Pas de grille et promu (4)	
	N	μ (σ)	N	μ (σ)	N	μ (σ)	N	μ (σ)
Note totale	35	25,40 (6,17)	35	26,57 (4,74)	35	23,43 (4,56)	35	26,09 (7,29)

Tableau 12. Moyenne et écarts-types de la note totale pour chacune des 4 copies

Tableau 13. Calcul de l'ampleur de l'effet entre les 4 groupes de copies en fonction de la note totale

Conditions	Grille et redoublant (1)		Grille et promu (2)		Pas de grille et redoublant (3)		Pas de grille et promu (4)		(1)vs(2)	(1)vs(3)	(1)vs(4)	(2)vs(3)	(2)vs(4)	(3)vs(4)
	N	μ (σ)	N	μ (σ)	N	μ (σ)	N	μ (σ)	Ampleur de l'effet					
Note totale	35	25,40 (6,17)	35	26,57 (4,74)	35	23,43 (4,56)	35	26,09 (7,29)	-0,21	-0,36	-0,10	-0,68	-0,08	-0,44

Tableau 13. Calcul de l'ampleur de l'effet entre les 4 groupes de copies en fonction de la note totale

Tableau 14. Régression linéaire de la note totale selon la présence du stéréotype « redoublant », de la grille d'évaluation, de l'ancienneté des évaluateurs, du fait qu'ils aient connu ou non le redoublement, de leur genre et de s'il s'agit de professeurs de français ou non.

Effet	Estimation	Erreur standard	Valeur t	Pr > t
Intercept	25,20725	1,20516	20,916	< 2e-16
Redoublant	-1,95425	1,34635	-1,452	0,14901
Grille	1,36474	1,35115	1,010	0,31432
Ancienneté	0,10346	0,04238	2,442	0,01595
Connu le redoublement	-2,61730	0,94146	-2,780	0,00623
Genre	-0,76703	1,07818	-0,711	0,47808
Français	2,20474	1,06895	2,063	0,04112
Redoublant*grille	0,40319	1,89689	0,213	0,83200

Tableau 14. Régression linéaire de la note totale selon la présence du stéréotype « redoublant », de la grille d'évaluation, de l'ancienneté des évaluateurs, du fait qu'ils aient connu ou non le redoublement, de leur genre et de s'il s'agit ou non de professeurs de français

Tableau 15. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée aux « copies redoublants » selon que l'évaluateur a connu ou non le redoublement

Variables \ Condition	Redoublant		
	N	μ (σ)	Ampleur de l'effet (1) vs (2)
Note totale attribuée par des enseignants ayant connu le redoublement (1)	34	23,26 (4,51)	-0,41
Note totale attribuée par des enseignants n'ayant pas connu le redoublement (2)	36	25,50 (6,12)	

Tableau 15. Moyennes, écarts-types et ampleur de l'effet de la note totale attribuée aux « copies redoublants » selon que l'évaluateur a connu ou non le redoublement

Tableau 16. Résultats au test d'homogénéité des variances pour la variable note totale (test de Levene)

Test de Levene pour l'homogénéité de la variance NOTE ANOVA des écarts au carré par rapport aux moyennes de groupe					
Source	DF	Somme des carrés	Carré moyen	Valeur de F	Pr > F
Groupe	3	22723.2	7574.4	3.74	0,0128
Erreur	136	275622	2026.6		

Tableau 16. Résultats au test d'homogénéité des variances pour la variable note totale