

Les jugements scolaires des professeurs de sciences en Fédération Wallonie-Bruxelles concordent-ils avec les résultats aux épreuves externes certificatives (CE1D) ?

Auteur : Rigutto, Lorenzo

Promoteur(s) : Lafontaine, Dominique

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

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

Année académique : 2020-2021

URI/URL : <http://hdl.handle.net/2268.2/13401>

Avertissement à l'attention des usagers :

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

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



Faculté de psychologie, logopédie et des sciences de l'éducation.

Département éducation et formation.

Les jugements scolaires des professeurs de sciences en Fédération Wallonie-Bruxelles concordent-ils avec les résultats aux épreuves externes certificatives (CE1D) ?

Lorenzo RIGUTTO

Mémoire présenté en vue de l'obtention du grade de Master en Sciences de l'Éducation

Promotrice : Dominique LAFONTAINE

Lectrices : Stéphanie NOEL et Juliane VERVIER

Année académique 2020-2021

Remerciements

Je tiens tout d'abord à remercier Madame Lafontaine, ma promotrice, pour sa disponibilité, sa bienveillance et son expertise dont elle a fait preuve tout au long de l'encadrement de ce mémoire.

Je remercie également Mesdames Virginie Dupont et Doriane Jaegers pour la réalisation des analyses statistiques et leur aide précieuse dans l'interprétation des résultats.

Il est également important pour moi de remercier Monsieur Fabian Pressia pour la relecture soutenue de la partie pratique et ses précieux conseils.

Je remercie d'avance Mesdames Stéphanie Noël et Juliane Vervier pour l'intérêt et le temps qu'elles consacreront à la lecture de ce travail.

Je tiens à remercier les directions, les enseignants et les élèves qui ont accepté de participer à notre recherche et sans qui la collecte des données n'aurait pas été possible. J'en profite pour mentionner Guillaume Martin, Pauline Delsalle et Caroline Delnaux pour notre collaboration sans faille lors des différentes étapes liées à la partie pratique, que ce soit pour les démarches réalisées auprès du comité d'éthique, la conception des questionnaires ou la récolte des données.

Enfin, je terminerai par remercier ma famille et mes amis pour leur soutien tout au long de ce Master. Je me permets de citer ma sœur, Alicia Rigutto, sans qui je ne me serais sans doute jamais inscrit.

Table des matières

I. Introduction	9
II. Revue de littérature	11
1. Le système éducatif en Fédération Wallonie-Bruxelle (FW-B)	11
A. Description	11
B. Les différents constats	12
C. Les évaluations externes certificatives	14
2. Le jugement scolaire	15
A. La définition du jugement scolaire	15
B. La subjectivité du jugement scolaire	17
3. Les biais du jugement scolaire	24
A. Selon les caractéristiques de l'élève	24
B. Selon les caractéristiques de l'enseignant	33
C. Selon les caractéristiques liées au contexte	36
4. Conclusion	38
III. Question de recherche et hypothèses	39
1. Question de recherche	39
2. Hypothèses de recherche	40
IV. Méthodologie	44
1. Description de la population et de l'échantillon	44
2. Méthodes et instrument.	44
A. Collecte des données	44
B. Modalités d'administration	47
C. Comité d'éthique	47
3. Traitement des données	47
A. Variables	47
B. Analyses statistiques	50
V. Présentation des résultats	52
1. Concordance entre les notes de l'enseignant et les scores au CE1D	52
A. Concordance toutes classes confondues	52
B. Concordance classe par classe	53
C. Dispersion des notes à l'intérieur des classes	54
2. La décomposition de la variance.	55
3. Lien entre les notes des enseignants et les variables étudiées.	56
A. Lien entre les notes des enseignants et les scores au CE1D	56

B.	Lien entre les notes des enseignants et les variables liées aux caractéristiques des élèves ...	57
C.	Lien entre les notes des enseignants et les variables liées aux caractéristiques des enseignants	59
D.	Lien entre les notes des enseignants et les variables liées aux caractéristiques du contexte .	60
VI.	Interprétation et discussion des résultats.	62
1.	Concordance entre les notes des enseignants et les scores au CE1D	62
2.	Dispersion des notes de l'enseignant et des scores au CE1D	63
3.	Lien entre les variables étudiées et les notes de l'enseignant	65
A.	Variables liées aux caractéristiques de l'élève	65
B.	Variables liées aux caractéristiques de l'enseignant.....	67
C.	Variables liées aux caractéristiques du contexte	68
4.	Biais et limites de notre recherche.....	70
VII.	Conclusion et perspectives.	72
VIII.	Bibliographie.	75
IX.	Annexes	78
1.	Formulaire pour la retranscription des points.....	78
2.	Questionnaire « élèves ».....	79
3.	Questionnaire « enseignant ».....	83

Table des illustrations

Figures

Figure 1 Les 3 étapes du jugement scolaire.....	16
Figure 2 Heuristic model of teacher judgment accuracy (Sudkamp <i>et al.</i> , 2012)	23

Tableaux

Tableau 1 La répartition de la variance entre classes et entre élèves dans les classes (Crahay, 2007) .	22
Tableau 2 Corrélations de Spearman classe par classe entre les notes des enseignants et les scores au CE1D.	53
Tableau 3 Moyennes et écart-types des notes de l'enseignant et des scores au CE1D.....	54
Tableau 4 Régression du score au CE1D sur les notes des enseignants.....	56
Tableau 5 Régressions multiniveaux entre les variables "élèves" et les notes des enseignants	57
Tableau 6 Régressions multiniveaux entre les variables "enseignant" et les notes de l'enseignant	59
Tableau 7 Régressions multiniveaux entre les variables "contexte" et les notes de l'enseignant.....	60

I. Introduction

En Fédération Wallonie-Bruxelles (FW-B), les enseignants évaluent et attribuent une cote aux élèves dans le but de vérifier les acquis avant d’octroyer l’autorisation de passation à l’année supérieure. Cette cote attribuée correspond donc à un jugement porté par le professeur (Merle, 1996), nous parlerons dans ce travail du jugement scolaire. L’habileté d’évaluer les performances des élèves avec précision est considérée comme un aspect important de la compétence professionnelle des enseignants (Ready et Wright, 2011 cités par Sudkamp *et al.*, 2012).

A ce sujet, Bressoux et Pansu (2003) ont réalisé une étude sur les éléments qui influencent le jugement scolaire. Même s’ils ont pu établir un lien significatif entre celui-ci et le score obtenu aux épreuves standardisées, ils constatent que près de la moitié de la variance n’est pas justifiée par la performance de l’élève, la note du professeur ne dépendrait pas exclusivement de leur compétence. Plus récemment, Sudkamp *et al.* (2012) qualifient le jugement professoral comme « loin d’être parfait » et affirment que plus de deux tiers de la variance de ces jugements ne peut être expliqué par la performance des élèves. Dans le but d’apporter des éclaircissements à cette constatation qui peut sembler surprenante, nous nous sommes posé la question de recherche suivante : « *Les jugements scolaires des professeurs de sciences en Fédération Wallonie-Bruxelles concordent-ils avec les résultats aux épreuves externes certificatives (CEID) ?* ».

Concrètement, en nous basant sur les facteurs influençant le jugement scolaire déterminés par Bressoux et Pansu (2003), nous allons tenter d’étudier la concordance entre le jugement scolaire des enseignants et les résultats obtenus aux épreuves externes certificatives au cours d’ « initiation scientifique ». Nous allons tenter de mettre en évidence des tendances qui renseigneraient davantage sur la variance des jugements scolaires inexplicée par les performances en nous focalisant sur des paramètres propres à l’élève, à l’enseignant et à l’établissement fréquenté. Pour ce faire, nous allons comparer les points au dernier bulletin en sciences des élèves de 2^{ème} année secondaire en FW-B avec le score qu’ils ont obtenu au CEID.

Ces données ont été récoltées grâce à la participation de 23 enseignants et de 465 élèves de 2^e année secondaire de l’enseignement de la FW-B, le nombre d’adolescents constituant notre échantillon s’élève à 465. Afin d’identifier les variables pouvant expliquer la variance des jugements scolaires, deux types de questionnaires ont été soumis : un pour les élèves (afin de

récolter des informations sur leur vie personnelle) et un pour les enseignants (de façon à recueillir des indications sur leur contexte de travail).

La revue de la littérature sera présentée en 3 volets. Le premier volet décrira le système éducatif de la FW-B, système dans lequel s'inscrit cette étude. Un point théorique se focalisera sur les évaluations externes et nous y retracerons brièvement l'histoire. Ensuite, le second volet sera consacré au jugement scolaire et à la difficulté de l'établir en toute objectivité. Enfin, un troisième volet passera en revue les différents biais liés aux caractéristiques de l'élève, du professeur et du contexte susceptibles d'influencer le jugement que portent les enseignants. Après la conclusion de cette partie théorique, nous reviendrons sur notre question de recherche et les hypothèses formulées à la suite de la revue de la littérature seront listées. Après cela, nous apporterons quelques précisions concernant la méthodologie employée avant de faire apparaître les résultats. Ceux-ci seront ensuite analysés et discutés en les mettant en lien avec la revue de la littérature. Pour terminer, la conclusion de ce travail comprendra les éventuels apports de cette recherche, ses limites et les perspectives de continuité qui pourraient sembler pertinentes.

II. Revue de littérature

1. Le système éducatif en Fédération Wallonie-Bruxelle (FW-B)

A. Description

Le système éducatif en Belgique francophone est fortement décentralisé puisque les établissements scolaires bénéficient de subventions de l'État et doivent se soumettre à son contrôle via l'inspection quel que soit le réseau (Lafontaine, 2001). En effet, 3 réseaux d'enseignement coexistent, à savoir :

- la fédération Wallonie-Bruxelles (FW-B), sous la responsabilité du Ministère de l'Education.
- l'officiel subventionné, organisé par les communes ou les provinces.
- le « libre » subventionné, organisé en majorité par l'Eglise catholique (Mottier Lopez et Crahay, 2009).

Afin de mieux piloter et orienter ces unités décentralisées ainsi que pour rendre le système plus juste et de meilleure qualité, le décret « Missions » apparaît en 1997 (Maroy *et al.*, 2002). Puisque le système manquait de moyens de pilotage, ce décret visait à réduire l'hétérogénéité des pratiques en matière de contenu transmis, d'évaluation et de pratiques pédagogiques (Crahay, 1997, cité par Maroy *et al.*, 2002). Les missions du système sont désormais précisées et les compétences à atteindre sont définies, quel que soit le réseau dans lequel s'inscrit l'établissement scolaire (Maroy *et al.*, 2002).

Malgré cette homogénéisation, les programmes d'enseignement restent différents selon le réseau, ce qui offre une marge de manœuvre relativement importante aux établissements scolaires (Mottier Lopez et Crahay, 2013). Le choix d'une « bonne » école ou du moins, adaptée au profil de l'élève, est une préoccupation constante des parents (Mottier Lopez et Crahay, 2013), Maroy *et al.* (2002) parlent de la réalité du « marché scolaire » souvent évoquée en Belgique francophone. Contrairement à ce que l'on pourrait penser, la concurrence n'apparaît pas essentiellement entre les réseaux mais plutôt en fonction de la situation géographique et de l'offre scolaire (Maroy *et al.*, 2002).

En FW-B, l'offre est d'autant plus diversifiée en raison des filières d'enseignement qui apparaissent à partir de la 3^e secondaire. En effet, au terme de la 2^e année secondaire, apparaissent 4 formes d'enseignement à savoir, le général, le technique de transition, le technique de qualification et le professionnel (Crahay *et al.*, 2019). En 2007, Mons (cité par

Crahay *et al.*, 2019) distingue le modèle de séparation composé de filières rigides comme en Allemagne, du système d'intégration caractérisé par un tronc commun et fortement ancré dans les pays du Nord de l'Europe ainsi que dans certains pays asiatiques. Compte tenu de l'orientation en filières relativement précoce, des taux de redoublement important et de l'existence d'un enseignement spécialisé fréquenté par 4-5% des élèves, la FW-B s'inscrit dans le modèle de la séparation, tout comme les deux autres communautés belges d'ailleurs.

Quoi qu'il en soit, l'obligation scolaire s'étend de 5 à 18 ans. Depuis septembre 2020, les enfants sont obligés de fréquenter l'enseignement maternel au moins un an puisqu'il s'agit d'un facteur important pour la réussite de l'enseignement primaire (Fédération Wallonie-Bruxelles, 2021). Après le préscolaire, le système éducatif se décline en 6 années d'enseignement primaire et 6 années dans le secondaire. Ces dernières sont déclinées en 3 cycles de 2 ans, le premier degré étant qualifié d'observation, le second d'orientation et le 3^e de détermination (Crahay *et al.*, 2019). L'enseignement fondamental correspondrait au tronc commun puisqu'il s'étend de la 3^e maternelle à la 2^e année secondaire, période durant laquelle le choix d'une filière n'intervient pas encore (Crahay *et al.*, 2019).

B. Les différents constats

Si nous nous intéressons au taux de redoublements dans les différents pays de l'OCDE selon les études PISA, nous pouvons constater que le FW-B détient le record avec 46 % en 2015 (Crahay *et al.*, 2019). Non seulement, notre système est celui qui a le plus recourt au redoublement mais en plus, il a connu la hausse la plus importante. En 2003, le taux de redoublement des enfants de 15 ans s'élevait à 38,6 % (Crahay et Marcoux, 2019). Ce pourcentage équivaut 46,0 % en 2009 avant d'atteindre 47,8 % en 2012 (Crahay et Marcoux, 2019). En 2018, le taux de pourcentage d'élèves en retard redescend à 41,1 % alors que la moyenne des pays de l'OCDE correspond à 10,6 % (Lafontaine *et al.*, 2019). Malgré cette diminution, cela signifie qu'en moyenne, quasiment 1 élève sur 2 a connu le redoublement à l'âge de 15 ans au sein de notre système éducatif. Lafontaine *et al.* (2019) qualifient d'exceptionnelle, cette tendance qu'a la FW-B à recouvrir massivement au redoublement

Lafontaine *et al.* (2019) se sont interrogés sur les effets néfastes du redoublement et ont utilisé les données PIRLS pour déterminer si le redoublement amplifiait les inégalités de performances liées à l'origine socio-culturelle de l'élève et s'il contribuait à la ségrégation scolaire. Dans un premier temps, ils ont déterminé la corrélation entre les taux de retard et la part de variance des performances en lecture expliquée par le nombre de livres à la maison. Il

apparaît que le lien entre l'origine socio-culturelle et les performances est plus fort dans les systèmes où le taux de retard est élevé. En ce qui concerne la ségrégation scolaire, ils ont déterminé les corrélations entre les taux de retard et la part de la variance des performances entre écoles puis entre élèves à l'intérieur des écoles. Il en résulte que le redoublement provoquerait une homogénéisation des élèves au sein des écoles et une exacerbation des différences entre écoles. En d'autres termes, la pratique du redoublement contribuerait à la ségrégation scolaire.

Crahay et Marcoux (2019) explorent un possible lien entre le taux de retard scolaire et les différents modèles définis par Mons (2007). En appariant leurs informations, il s'avère que tous les systèmes scolaires qui ont opté pour le modèle de séparation se situent au-dessus de la barre des 15% de redoublement.

Selon Felouzis (2015), les filières renforcent les inégalités scolaires. Cet auteur a mené une étude à ce propos en 2013 avec Charillot et Fouquet-Chauprade à Genève. Puisque certains établissements étaient régis par un système à classe hétérogènes avec le même enseignement pour tous et que d'autres suivaient un système de filières, ils ont pu comparer l'efficacité et l'équité de ces 2 systèmes sur des données longitudinales au niveau des acquis des élèves et de leur orientation. Il en résulte pour les 3 domaines testés par PISA, que l'écart de score entre les élèves les plus favorisés et les moins favorisés au plan socio-économique est plus fort dans le modèle de séparation que dans celui d'intégration. De plus, pour un élève moyen, les chances de s'orienter dans le secondaire général est de 10 % s'il est scolarisé dans la filière à exigence faible alors que s'il était scolarisé dans un système sans filière, ses chances de poursuivre dans le général s'élèveraient à 50 % (Felouzis, 2015).

En analysant les résultats de PISA 2018, Lafontaine *et al.* (2019) montrent que l'écart de performance en fonction de l'origine sociale des élèves est nettement plus marqué pour les pays qui ont adopté un modèle de séparation avec un système de filières, notamment l'Allemagne, le Luxembourg, la Flandre ou la Hongrie. En FW-B, la différence de scores en lecture entre les 25 % d'élèves les moins favorisés et les 25 % plus favorisés s'élève à 107 points alors que la moyenne de l'OCDE correspond à 89 points. Cet écart est relativement stable depuis les années 2000 et équivaut à environ 3 années d'étude (Lafontaine *et al.*, 2019). La FW-B se classe ainsi parmi les systèmes éducatifs où ces inégalités sont les plus marquées.

Par contre, il est important de souligner que la différence de performances entre élèves natifs et immigrés de 1^{re} ou 2^e génération est moins marquée. En effet, même si notre système

compte une proportion importante d'élèves issus de l'immigration, ceux-ci présentent une différence de performances limitée par rapport à leurs condisciples d'origine belge. Lafontaine *et al.* (2019) affirment qu'il est important de souligner ce constat positif et que rares sont les systèmes éducatifs qui affichent un tel résultat. Nous savons que la FW-B se distingue par des inégalités sociales fortes mais elles ne sont pas davantage marquées selon l'origine ethnique et culturelle des élèves (Lafontaine *et al.*, 2019).

C. Les évaluations externes certificatives

Dans le but de construire progressivement une forme de régulation plus centralisée des établissements, le décret « Missions » prévoyait la mise en place d'évaluations externes des performances du système (Maroy *et al.*, 2002). Néanmoins, ce dispositif ne peut être instauré sans l'articuler avec des « socles de compétences » ou référentiels communs à l'ensemble des écoles (Lafontaine, 2001). D'après Lafontaine (2001), le Rapport sur l'état du système éducatif rédigé par les experts de l'OCDE serait à l'origine de la mise en place du dispositif d'évaluation externe. En effet, ce rapport datant de 1991 a mis en évidence deux caractéristiques inquiétantes du système éducatif en FW-B à savoir, l'importance du taux d'échecs et la disparité de réussite en fonction de l'école fréquentée (Lafontaine, 2001). Il s'avère que ce dispositif d'évaluation permettrait d'améliorer l'efficacité et de réduire les inégalités scolaires (Grissmer *et al.*, 2000 ; Hong et Yongs, 2008 ; cités par Maroy *et al.*, 2013). Imposer des standards communs à tous les élèves obligerait les enseignants à développer des attentes uniformes à l'égard de tous ses élèves, quelles que soient leurs caractéristiques individuelles et le contexte scolaire dans lequel ils évoluent (Maroy *et al.*, 2013).

Avant 2008, il n'existe aucune forme de certification externe commune dans le primaire et dans le secondaire, il revenait donc à l'école de décider l'octroi ou non des diplômes (Mottier Lopez et Crahay, 2009). Depuis la mise en place du décret relatif à l'évaluation externe des acquis des élèves de l'enseignement obligatoire (2006), les élèves obtiennent une certification sur base de leur performance aux évaluations externes certificatives. Ces épreuves sont communes à tous les élèves au même stade dans leur scolarité. Elles sont élaborées par un groupe de travail constitué d'inspecteurs, de conseillers pédagogiques, d'enseignants et d'un représentant du Service de Pilotage du système éducatif, ce groupe de travail est présidé par l'Inspecteur général du niveau d'enseignement concerné (Fédération Wallonie-Bruxelles, 2021).

A la fin de l'enseignement primaire, les enfants doivent se soumettre à cette épreuve dans le but d'obtenir leur Certificat d'Etude de Base (CEB). Certains élèves du 1^{er} degré, n'ayant pas obtenu leur CEB en 6^{ème} primaire, sont amenés à le repasser, qu'ils soient dans une année différenciée, commune ou complémentaire (Fédération Wallonie-Bruxelles, 2021). Ensuite, depuis l'année académique 2013-2014, les élèves doivent passer le Certificat d'Etude du 1^{er} Degré de l'enseignement secondaire (CE1D) à la fin de leur 2^{ème} secondaire. Cette épreuve externe porte sur le français, les mathématiques, les langues modernes et les sciences, les élèves doivent obtenir 50 % dans chacune de ces disciplines puisque ce test porte sur l'ensemble des compétences telles qu'elles sont décrites par les Socles de compétences (Fédération Wallonie-Bruxelles, 2021). Dans le cadre de ce travail, c'est à cette épreuve que nous comparerons les notes attribuées par l'enseignant pour évaluer l'exactitude du jugement scolaire. Et enfin, au terme de leur dernière année d'étude dans le secondaire, l'épreuve externe est prise en compte pour l'octroi du Certificat d'Enseignement Secondaire Supérieur (CESS). Pour celui-ci, le décret relatif aux évaluations externes (2021) indique que les résultats obtenus à ce test doivent être pris en considération pour les compétences évaluées mais il précise également l'importance de se référer aux évaluations internes pour les autres compétences relatives au français et à l'histoire. En effet, cette épreuve ne porte que sur l'une ou l'autre compétence de ces deux cours. Le CESS est organisé depuis l'année académique 2010-2011 (Fédération Wallonie-Bruxelles, 2021).

2. Le jugement scolaire

A. La définition du jugement scolaire

Lafortune (2006, p.22, citée par Mottier et Allal, 2008, p. 465) définit le jugement professionnel de l'enseignant comme « un processus qui mène à une prise de décisions, laquelle prend en compte différentes considérations issues de son expertise (expérience et formation) professionnelle. Ce processus exige rigueur, cohérence et transparence. En ce sens, il suppose la collecte d'informations à l'aide de différents moyens, la justification du choix des moyens en lien avec les visées ou intentions et le partage des résultats de la démarche dans une perspective de régulation ». Le jugement professoral n'est donc pas propre à l'évaluation puisqu'il concerne l'ensemble des activités mises en place par l'enseignant (Mottier & Allal, 2008) mais dans le cadre de ce mémoire, il est préférable de se limiter à une définition contextualisée dans le domaine de l'évaluation.

En nous référant à la définition de Stufflebeam (1986) reprise par Romainville (2011, p.1), évaluer correspond 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. ». L'évaluation peut donc être décomposée en 3 phases successives : la prise d'informations suivie de l'interprétation et enfin, la prise de décision. Lors de cette dernière étape, l'enseignant prend des initiatives qu'il estime justes dans le but que l'élève poursuive sa scolarité dans des conditions optimales. En fonction du niveau d'enseignement et de la période de l'année académique, la prise de décision peut consister à diriger un élève vers des séances de remédiation, vers une autre option voire vers une filière que l'ensemble du conseil de classe jugerait plus adéquate. Parfois, en cas d'échec, l'équipe éducative peut conclure qu'il est préférable d'avoir recours au redoublement, sinon les enseignants déterminent s'il est nécessaire que l'élève passe un examen à la rentrée. Meissel *et al.* (2017) précisent que l'intervention de certains biais dans ces décisions à enjeux élevés doit être prise particulièrement au sérieux mais nous nous attarderons sur les biais ultérieurement. Nous aborderons également les biais liés la seconde phase, la prise d'informations à travers les caractéristiques liées à l'enseignant. Concrètement, elle correspond au moment où l'enseignant corrige la copie de ses élèves. A l'aide d'une grille construite au préalable, le professionnel relève des informations en corrigeant afin d'établir un jugement en se basant sur la note. Même si Stufflebeam (1986) fait clairement apparaître le terme « jugement » lors de la troisième phase, Mottier (2008) reconnaît qu'il est difficile de démarquer le jugement de l'interprétation et de la prise de décision. Le jugement se situerait donc plutôt entre les phases 2 et 3.

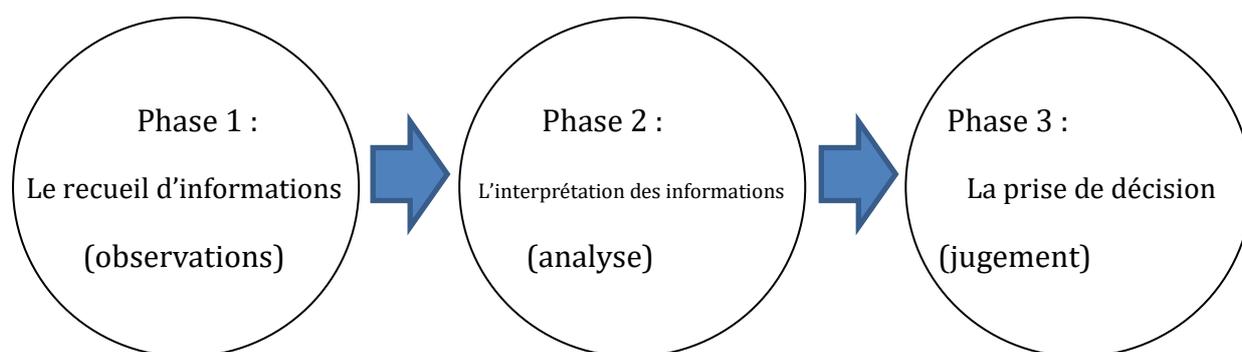


Figure 1 Les 3 étapes du jugement scolaire

Tout comme les appréciations, la note chiffrée fait partie intégrante du jugement professoral. Tout au long de ce mémoire, ce système de cotation correspondra au jugement des enseignants puisque nous n'avons aucune information sur les décisions prises concernant la scolarité des élèves ayant participé à notre étude. De plus, selon Bressoux et Pansu (2003), la note que ces derniers reçoivent occupe une place privilégiée dans leur parcours scolaire. Ces auteurs affirment que même si le jugement professoral peut se présenter sous des formes plus informelles à travers des feed-back qu'ils soient verbaux ou non, la note renseigne les élèves sur leur niveau de performance, leur permettant de se juger eux-mêmes plus ou moins compétents. Felouzis (1997) affirme que les notations sont considérées comme le reflet de leurs compétences tandis que Bressoux et Pansu (2003) déclarent qu'elles sont tellement ancrées dans la mémoire collective que la note constituerait une véritable forme objectivée de la valeur scolaire. La note correspond-elle réellement au reflet des compétences effectives ? Autrement dit, le jugement scolaire se fonde-t-il exclusivement sur les productions scolaires ? Si ce n'est pas le cas, comment se construisent ces jugements professoraux ? Quoiqu'il en soit Romainville (2011, p.1) qualifie l'objectivité de « chimère » dans le domaine de l'évaluation.

B. La subjectivité du jugement scolaire

a) Une tâche complexe pour les enseignants

Tout d'abord, nous pourrions remettre en question la faculté des professeurs à porter ce type de jugements. En effet, même si ceux-ci sont considérés comme « les personnes légitimement habilités à produire ces jugements » (Felouzis, 1997, p. 71), Hadji (1990, cité par Bressoux & Pansu, 2003) souligne la problématique du degré d'assurance de ces jugements auxquels on ajoute un élément qui n'est plus d'ordre factuel. En effet, Nicaise (2002) définit « le fait de noter » comme « une action proprement rationnelle qui trouve ses fondements à la fois dans les propres intérêts et les propres valeurs de l'enseignant » (p. 139). Ce chercheur souligne l'importance des « filtres interactifs » dans l'évaluation et distingue trois types d'« arrangements évaluatifs » que le professeur doit prendre en considération avant d'attribuer une note. En ce qui concerne les « arrangements internes », l'enseignant doit participer à une « négociation didactique » avec ses élèves au sein de sa classe (Chevallard, 1991, cité par Nicaise, 2002, p.140). Concrètement, ces arrangements régulent les interactions didactiques puisque la note peut contribuer à entretenir un climat propice aux apprentissages, à encourager les élèves en difficulté, à restaurer l'autorité ou encore, à maintenir une moyenne de points habituelle (Nicaise, 2002). Bressoux et Pansu (2003) avaient également déclaré que

les fonctions internes du système éducatif à savoir, les fonctions didactiques influençaient les jugements évaluatifs et que ceux-ci participaient au maintien de la paix sociale en classe. Nicaise (2002) parle également d' « arrangements externes », cette deuxième catégorie prend en considération toutes les personnes qui interagissent avec la classe sans y participer directement comme la direction, l'administration, les collègues et la famille des élèves. L'enseignant doit leur présenter une image satisfaisante de sa classe, il doit donc veiller au niveau de difficulté et au pourcentage de réussite. L'enseignant « vache » peut directement impacter le parcours scolaire d'élèves moyens qui auraient obtenu la moyenne si le correcteur avait été une autre personne car l'équipe éducative se base sur les notes pour prendre des décisions en conseil de classe (Felouzis, 1997). Pourtant, ce type de professeurs est généralement considéré plus favorablement par les collègues en raison de la logique éducative bien ancrée dans notre système ; il est préférable de faire échouer un élève dont le niveau de performance est à la limite de ce qui est exigé dans le programme (Crahay, 2007). D'après Crahay (2007), cette « culture de l'échec » suppose qu'un enseignant, avec qui tous les élèves réussissent serait davantage susceptible d'être victime de reproches par ses collègues le qualifiant de « trop facile » et donc, sa compétence et son identité professionnelle pourraient être remises en cause. Suivant cette logique, les notes attribuées à l'élève ne reflèteraient pas la juste valeur de la performance car celle-ci serait sous-estimée afin que le professionnel garde une certaine crédibilité auprès de ses collègues. Cependant, Nicaise (2002) souligne l'importance de concevoir des évaluations ni trop sévères ni trop laxistes afin que cela ne nuise pas à la réputation de l'enseignant. Enfin, le dernier type d'arrangements correspond aux « *arrangements pour soi* », ceux-ci dépendent directement de l'histoire et de la personnalité de l'évaluateur (Nicaise, 2002). Cette catégorie recense une série de représentations personnelles relatives à son idéal pédagogique, sa conception de l'éducation, son propre parcours scolaire ou encore ses engagements politiques et associatifs. Par conséquent, Nicaise (2002) affirme que l'objectivité et l'équité ne sont pas la norme dans l'acte évaluatif puisqu'il dépend de ces trois catégories d' « arrangements » qui sont étroitement liées.

Nous pouvons également retrouver ces arrangements dans les quatre caractéristiques du jugement professionnel déterminées par Mottier Lopez et al (2009). Ces chercheurs ont réalisé une étude à Genève en 2006 ; ils se sont entretenus avec dix enseignants de profils différents en se focalisant sur deux élèves pour lesquels ils ont hésité entre la note 4 et la note 3, cette dernière qualifie les objectifs de « presque » acquis alors que la note 4 certifie qu'ils sont

acquis. A l'aide du matériel utilisé par l'enseignant dans l'attribution de ces notes et de l'analyse qualitative des entretiens, ils ont pu mettre en évidence ces quatre caractéristiques du jugement professionnel :

- « Le jugement professionnel en évaluation comprend deux démarches de base : réunir et confronter des sources d'information de nature diverses ; les combiner à travers un raisonnement interprétatif qui peut inclure un algorithme de calcul mais qui ne s'y réduit pas. (...) »
- Le jugement professionnel s'appuie sur des procédures d'évaluation que l'enseignant élabore et applique à l'ensemble des élèves, mais une de ses fonctions essentielles est d'assurer l'adaptation des procédures aux situations singulières qui l'exigent. (...)
- Le jugement professionnel en évaluation est un acte cognitif individuel mais il s'élabore et s'intègre dans un réseau de communication au sein d'un contexte social et institutionnel. (...)
- Le jugement professionnel en évaluation implique une mise en relation entre les exigences du système scolaire, la culture de l'établissement, les valeurs personnelles et les stratégies pédagogiques de l'enseignant. » (Mottier Lopez *et al.*, 2009, p. 31-32-33)

Non seulement, nous retrouvons la présence des filtres interactifs qui mènent à l'élaboration de certains arrangements mais en plus de cela, Mottier Lopez *et al.* (2009) y apportent quelques précisions en abordant notamment l'importance des situations singulières. Concrètement, si l'enseignant pense que le résultat d'un élève à un contrôle ne représente pas ses compétences, il est en droit de proposer une autre interrogation ou d'accorder une importance supplémentaire sur les tâches ordinaires réalisées en classe avant d'inscrire la note dans le bulletin. En plus de cela, ces chercheurs soulignent l'importance de ne pas se focaliser exclusivement sur la note. Selon Mottier Lopez *et al.* (2009), lorsque l'enseignant doit prendre une décision à haut-enjeux, il doit confronter ses notes avec d'autres informations relatives à l'investissement de l'élève, sa progression et aux éventuelles informations recueillies en interagissant avec l'élève ou en s'entretenant avec sa famille.

b) La considération pour la note chiffrée

La forme sous laquelle est transmis le jugement scolaire pourrait également être critiquée puisqu'elle renforce la comparaison entre élèves (Merle, 2018). Merle (2018) va plus loin en affirmant que la note chiffrée aurait des effets contre-productifs sur les apprentissages des élèves. Avant cela, Marchand (1996) s'était intéressé aux avantages et aux inconvénients de

ce système de cotation, il nuance donc les propos des chercheurs en affirmant que les points permettent des combinaisons, des totalisations, des moyennes et des classements. La notion de classement fait, elle aussi, directement référence à la comparaison. Selon Marchand (1996), l'action de « classer » est une activité inévitable dans la vie en société. Cette manière de procéder engendre un sentiment de compétition chez les enfants qui se retrouvent davantage stimulés et motivés pour performer, les plaçant inévitablement dans une visée d'excellence alors que les élèves devraient considérer leur performance dans un but de maîtrise (Crahay, 2007). Même si la motivation extrinsèque a mauvaise réputation dans le cadre scolaire, Marchand (1996) déclare que la compétition n'a pas toujours eu une valeur négative et qu'il serait délicat de supprimer les classements qui permettent aux professeurs et aux parents de relativiser la note. Merle (2018), quant à lui, estime que cette logique n'entretient aucun rapport avec les apprentissages scolaires et justifie la légitimité de la note au sein des classes par son usage dans les concours. En plus de cela, Marchand (1996) reconnaît que ces comparaisons risquent de créer un sentiment de maîtrise surestimé chez les élèves les plus forts et d'être néfastes pour les élèves se trouvant en bas de l'échelle. Sudkamp *et al.* (2012) rejoignent Marchand en affirmant que les notes ont des effets préjudiciables sur la perception des élèves de leurs capacités, ce qui impacterait considérablement leur apprentissage.

De plus, il s'interroge également sur la réelle signification d'un point à l'autre en fonction de l'enseignant et de la matière concernée. En effet, les critères de notation en mathématiques seraient plus stables et contrôlables qu'en français, ce qui entraînerait une plus grande dispersion de notes dans les domaines relatifs à la langue où le produit attendu est plus difficilement identifiable (Felouzis, 1997). Merle (2018) fait référence à une étude menée en 1978 où les divergences entre correcteurs des copies de mathématiques au baccalauréat étaient importantes. Parallèlement, ces différences étaient davantage marquées pour les épreuves relatives au français, Merle (2018) justifie cette différence en fonction de la précision du barème où sont codifiés les critères d'évaluation. En effet, la précision du barème serait moindre pour une dissertation que pour une démonstration mathématique. En ce qui concerne les sciences, il est difficile de dégager une tendance stable comme en mathématiques et en français. Ceci dit, nous pourrions comparer davantage les épreuves relatives à la chimie et à la physique aux critères utilisés dans le cadre du cours de mathématique, la biologie intégrant généralement plus d'aptitudes liées à la langue.

c) Une logique de comparaison

Felouzis (1997) affirme que les performances des élèves ne prennent du sens que dans la comparaison. En effet, l'enfant évalué se compare à ses condisciples, cela engendre une motivation extrinsèque directement liée aux performances du reste de la classe. Il est important pour eux de ne pas descendre dans le classement, leur but devient de se retrouver dans le haut ou de maintenir leur position par rapport aux autres. Leur satisfaction dépend donc directement de la performance de chaque élève de la classe, ils ne peuvent pas apprécier leurs éventuels progrès car ceux-ci ne sont pas pris en compte en tant que tel (Crahay, 2007). C'est pour cette raison que leur position relative reste approximativement identique alors que chaque élève est censé progresser entre le début et la fin d'une année académique (Felouzis, 1997). Merle (2018) se base sur la recherche de Caverni et Noizet (1979) pour déclarer que les évaluations soumises tout au long de l'année ne prennent que modérément en compte les améliorations des performances. Il va plus loin en affirmant que « les procédures d'évaluation privilégient le repérage des permanences plutôt que les changements dans les productions des élèves » (Merle, 2018, p. 160). Nous pouvons déduire que cette inertie scolaire est directement en lien avec le jugement des enseignants. Il n'est donc pas évident de concevoir une évaluation « objective » puisque cette pratique s'inscrit dans une logique de comparaison.

De plus, le professionnel conçoit la plupart du temps ses interrogations de sorte à obtenir une distribution gaussienne de notes obtenues par ses élèves (Crahay, 2007). Effectivement, comme l'avait déjà souligné Posthumus en 1947, les exigences sont adaptées de sorte à avoir une majorité d'élèves moyens et une minorité d'élèves forts et faibles. Cette loi se vérifie aussi bien dans les classes jugées comme « excellentes » que dans celles catégorisées de « médiocres » (Merle, 2018). Merle (2018) précise d'ailleurs que les moyennes de classe ne sont pas supérieures dans les meilleures classes par rapport aux classes faibles. Ce constat prouve la subjectivité du jugement scolaire en fonction de la classe fréquentée puisque le professeur adapte logiquement le contenu des apprentissages et la difficulté des exercices proposés (Merle, 2018). Cette distribution gaussienne mène à surestimer la performance d'un élève moyen si ses condisciples sont plus faibles que lui et inversement, à la sous-estimer si les autres élèves de la classe sont plus forts que l'élève moyen dont il est question (Crahay, 2007). Merle (2018) recommande une évaluation plus objective comme les évaluations externes pour que le score ne soit plus dépendant du niveau moyen de la classe mais plutôt de celui des élèves de sa génération.

Grisay *et al.* (1984, cités par Crahay, 2007) ont recueilli auprès de 1503 élèves répartis dans 53 classes des scores obtenus à un test standardisé et les notes à l'examen conçu par l'enseignant. La part respective de variation des scores au test standardisé et à l'examen a pu être calculée entre les élèves dans les classes dans un premier temps et, dans un second temps, entre les différentes classes participantes.

	Scores au test standardisé	Score à l'examen
Variation entre élèves	57,9 %	82,8 %
Variation entre classes	42,1 %	17,2 %

Tableau 1 La répartition de la variance entre classes et entre élèves dans les classes (Crahay, 2007)

Le tableau ci-dessus indique la répartition de la variance. Nous pouvons constater que pour les scores obtenus au test standardisé, la variation est attribuable pour 57,9 % aux différences interindividuelles et pour 42,1 % aux différences entre classes alors qu'en ce qui concerne les scores à l'examen, la dispersion des notes s'explique par les différences entre les élèves pour 82,8 % et par les différences entre classes pour 17,2 %. Grâce au test standardisé qu'on peut considérer comme une épreuve externe, la variation des scores imputable aux différences entre élèves passe de 82,8 % à 57,9 %. Pour Crahay (2007), les évaluations conçues par les enseignants exacerbent les différences entre les élèves au sein de leurs classes, ce qui correspond à la loi de Posthumus.

d) L'exactitude du jugement professoral

De nombreuses études ont été réalisées sur l'exactitude du jugement professoral (Bressoux et Pansu, 2003). Tout d'abord, dans la méta-analyse de Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003), ces chercheurs ont déterminé une corrélation médiane de 0,66 entre le jugement du professeur et la performance de l'évalué, cela qualifierait les jugements établis par les professionnels d'assez valides. Cependant, comme mentionné dans l'introduction de ce travail, Bressoux et Pansu (2003) affirment que près de 50% de la variance n'est pas expliquée par la performance et Sudkamp *et al.* (2012) déclarent que près de deux tiers de la variance reste inexpliquée. Cela laisse sous-entendre qu'il existerait une série de facteurs qui interviendraient pour biaiser le jugement professoral.

D'après leurs recherches, Sudkamp *et al.* (2012) ont proposé un modèle de l'exactitude du jugement professoral reprenant quatre domaines qui pourraient directement et indirectement impacter l'exactitude de la note attribuée à un élève. Parmi ces quatre domaines, nous retrouvons les caractéristiques propres :

- à l'enseignant : son expertise professionnelle ou encore les stéréotypes qu'il porte à propos des élèves.
- au jugement : l'enseignant est parfois amené à évaluer une compétence spécifique mais il se peut aussi qu'il soit amené à fournir un jugement plus global sur l'habilité scolaire.
- à l'élève : que ce soit ses connaissances antérieures, sa motivation ou son intelligence au sens large.
- au test standardisé : cela concerne plutôt la matière à proprement parler, l'ensemble des tâches spécifiques mais également, le niveau de difficulté de ces dernières.

Selon eux, l'interaction entre les caractéristiques de l'enseignant et celles de l'élève impacterait l'exactitude du jugement, Artelt et Rausch (2014, p.38) parlent notamment du « similar-to-me effect » mais nous y reviendrons ultérieurement. Il en va de même pour l'interaction entre les caractéristiques du jugement et celles du test standardisé, Sudkamp *et al.* (2012) illustrent leurs propos par un exemple ; selon eux, un test peut mesurer une capacité académique très spécifique comme les compétences arithmétiques, alors que l'objectif des enseignants dans leur jugement est plus large, ils pourraient par exemple vouloir évaluer la capacité globale des élèves en mathématiques, ce rapport pourrait altérer la précision du jugement.

Sudkamp *et al.* (2012) ont également identifié deux variables à savoir, le jugement des enseignants et la performance des élèves à un test standardisé. En se basant sur ce modèle, la correspondance entre ces deux variables représenterait l'exactitude du jugement. Notons que le jugement des enseignants serait influencé par les caractéristiques de l'évaluateur et par celles du jugement, alors que la performance de l'élève serait influencée par les caractéristiques de ce dernier et celles du test.

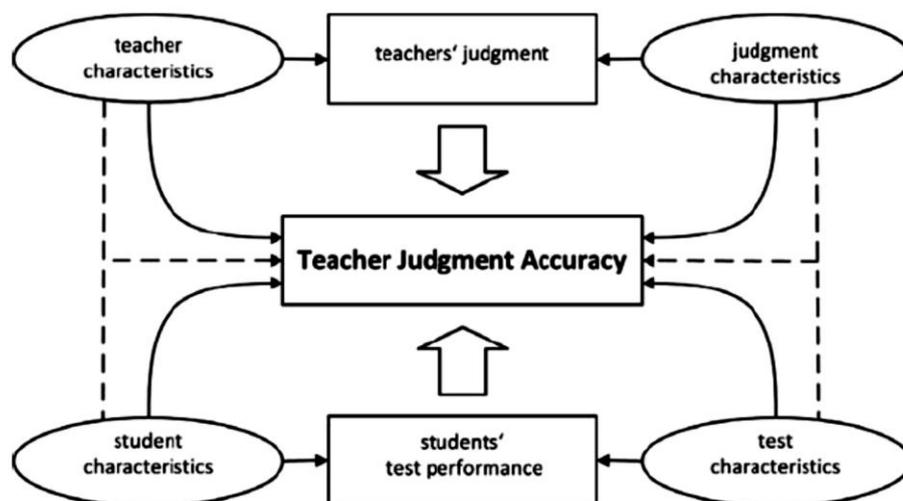


Figure 2 Heuristic model of teacher judgment accuracy (Sudkamp et al., 2012)

En se basant sur ce modèle et sur les recherches déjà effectuées à ce sujet, nous allons par la suite présenter les différents biais selon les caractéristiques de l'élève, du professeur et du contexte puisque nous avons déjà abordé l'influence de la classe dans laquelle un élève se trouve sur le jugement professoral. N'ayant pas d'information sur les interrogations des enseignants, nous ne nous focaliserons pas sur les biais liés à ce paramètre.

3. Les biais du jugement scolaire

A. Selon les caractéristiques de l'élève

Pour une même performance, une variation du jugement scolaire peut être observée en fonction des caractéristiques de l'élève évalué. Autrement dit, cela signifie qu'au préalable, les élèves peuvent être soit privilégiés soit défavorisés en raison d'une série de facteurs qui leur sont propres.

a) Le sexe

Plusieurs chercheurs se sont intéressés à la variation du jugement scolaire en fonction du sexe de l'élève. La littérature concernant les biais liés au genre a mis en évidence des résultats contradictoires (Kacprzyk *et al.*, 2019).

Même si Dusek et Joseph (1983, cités par Merle, 2018) ont montré que la relation entre le genre et les biais d'évaluation varie selon le niveau scolaire et la discipline dont il est question, Bressoux et Pansu (2003) déclarent que c'est surtout à un niveau de scolarité relativement avancé que nous pouvons dégager des tendances. Ainsi, selon Felouzis (1997) les filles de seconde sont surévaluées en français alors que les garçons bénéficient d'un bonus aux interrogations de mathématiques. Après avoir attribué des copies strictement identiques à des filles et à des garçons auprès de 48 enseignants, Lafontaine et Monseur (2009) nuancent les propos de Felouzis (1997) en affirmant que les enseignants de mathématiques ont effectivement tendance à surestimer les performances des élèves masculins si leur copie est bonne ou moyenne mais à l'inverse, ils valorisent les copies faibles des filles en étant plus indulgents. Ces chercheurs ont mené une deuxième étude de type corrélationnel complémentaire à leur première recherche expérimentale (2009) en mettant en relation les scores obtenus au test PISA 2000 et les notes attribuées par les enseignants en mathématiques. En ne s'intéressant qu'à la moyenne et à l'écart-type, les filles seraient moins performantes que les garçons à l'évaluation externe mais elles obtiendraient de meilleures notes scolaires de la part de leurs enseignants. Si nous regardons les corrélations, il apparaît que les garçons

seraient évalués plus objectivement que les filles puisque la corrélation entre leur score PISA et leur note s'élève à 0,30 comparativement à 0,20 pour les filles. Ceci dit, une analyse statistique plus poussée a permis de mettre en évidence que ce sont les garçons qui sont davantage victimes de sous-estimation. En effet, en ce qui concerne les élèves les plus performants, les filles sont légèrement sous-estimées par leurs professeurs alors que dans le cas des élèves moins performants, les garçons sont considérablement moins bien notés et se voient retirer davantage de points. Concrètement, l'écart de notes varie généralement de 5 points sur 100 pour les élèves performants tandis que pour les plus faibles, cet écart varie autour de 20 points. Parallèlement, Martinez *et al.* (2009, cités par Meissel *et al.*, 2017) ont montré que les filles étaient surévaluées par les échelles critériées de leurs professeurs par rapport aux tests standardisés. Lafontaine et Monseur (2009) avaient mené une deuxième étude similaire dans le domaine du français mais ils n'ont pas obtenu de différence significative en fonction du sexe de l'élève, ce résultat est en adéquation avec la recherche effectuée par Bressoux et Pansu en 2003.

De leur côté, Ready et Wright (2011, cités par Meissel *et al.*, 2017) ont trouvé un biais négatif dans l'évaluation des enseignants pour les garçons dans le domaine de la langue. En revanche, les résultats de Terrier (2014, cité par Merle, 2018) concordent davantage avec les travaux réalisés par Lafontaine et Monseur en 2009. La recherche de Terrier (2014) a été menée auprès de 4579 élèves de sixième scolarisés dans l'académie de Créteil. En comparant les scores au test standardisé aux notes attribuées par l'enseignant et en déterminant les écarts moyens, Terrier ne constate pas non plus de différence significative dans le domaine du français. En revanche, pour les mathématiques, les notes des professeurs sont en moyenne 6,2% plus élevées pour les filles que pour les garçons. Merle (2018) justifie ce biais d'évaluation positif en faveur des filles par une évaluation plus favorable des élèves faibles rappelant que le niveau moyen des filles dans ce domaine est plus faible (Terrier, 2014, cité par Merle, 2018).

Même si Felouzis déclare en 1997 qu'il serait pertinent de s'interroger sur le jugement scolaire en fonction de la branche scientifique évaluée, peu d'études ont été menées dans le cadre du cours de sciences.

Récemment, Kacprzyk *et al.* (2019) ont réalisé une étude dans le domaine de la biologie à l'Université de Dublin. Cette recherche a été effectuée pendant deux années consécutives (en 2014/2015 et en 2015/2016), ils étaient 188 à s'être inscrits en 2014 (121 filles et 67 garçons)

et 151 en 2015 (97 filles et 54 garçons). Les chercheurs ont analysé les résultats de ces 2 groupes d'étudiants pour 3 domaines à savoir, « La biologie cellulaire et la génétique », « La vie sur Terre » et « Les sciences biomédicales ». Sur base des données, les chercheurs ont construit un « female-to-male ratio » qui devrait être égal à 1 s'il n'y avait pas de différence de performance en fonction du sexe de l'étudiant. Si ce nombre est inférieur à 1, cela signifie que les garçons seraient plus performants que les filles et à contrario, si le ratio est supérieur à 1, ces dernières se montreraient plus performantes que leurs congénères masculins. Au sein de chaque module étudié, il n'y a pas de différence significative en fonction du sexe de l'étudiant que ce soit avec les rapports à réaliser à domicile ou via une plateforme, les travaux pratiques ou avec les interrogations sous-forme de QCM portant sur une partie de la matière. Par contre, que ce soit pour « La vie sur Terre » ou « La biologie cellulaire et la génétique », une différence significative apparaît lorsqu'on introduit le retrait de points aux QCM de fin d'année. Avec un ratio se situant entre 0,85 et 0,90, les filles seraient désavantagées par rapport aux garçons et deviendraient moins performantes. En ce qui concerne « Les sciences biomédicales », le ratio reste égal à 1 pour l'examen de fin d'année qui comprend également des questions de type QCM mais sans retrait de points. Dans ce cas, nous pouvons donc conclure que c'est le type de question et en particulier, le retrait de points par mauvaise réponse qui introduirait une inégalité en fonction du genre. C'est en tout cas ce que les chercheurs et l'Université ont conclu puisque cette étude a conduit à l'élimination des retraits de points pour les années académiques suivantes. Dans ce cas précis, l'effet du genre serait donc davantage marqué sur les pratiques d'évaluation que sur le sujet en tant que tel. Notons quand même que les 3 modules étaient relatifs à la biologie.

Breda et Ly (2014, cités par Merle, 2018) ont porté leur recherche sur le concours d'entrée à l'École nationale supérieure dans les filières scientifiques et littéraires en comparant les notes attribuées aux épreuves écrites et orales. Les résultats tendent à montrer que les filles seraient avantagées à l'oral dans les disciplines les plus dominées par les garçons à savoir, les mathématiques, la physique et la philosophie. Inversement, les garçons obtiennent de meilleurs résultats à l'oral dans les domaines où les filles sont surreprésentées comme la littérature ou l'histoire. Breda et Ly (2014, cités par Merle, 2018) en ont conclu qu'en fonction du genre de l'étudiant, la forme sous-laquelle est administrée l'examen (orale ou écrite) pouvait inverser la tendance des meilleurs résultats.

En nous référant aux deux dernières expériences citées précédemment, nous avons pu mettre en évidence que la forme sous-laquelle se présente l'évaluation a d'autant (voire plus) d'importance que la matière d'enseignement concernant les biais liés au genre.

b) L'origine ethnique

En plus d'être désavantagé par le fait de ne pas assister à des cours donnés dans sa langue maternelle (Crahay, 2007), un élève immigré serait davantage pénalisé par la note qu'il se voit attribuer. D'après Bressoux et Pansu (2003), l'origine ethnique de l'enfant influencerait également le jugement scolaire. La couleur de peau ainsi que la culture de l'enfant interviendraient dans la variance des notes qui n'est pas justifiée par la performance scolaire. D'après la méta-analyse de Dusek et Joseph (1983, cités par Bressoux et Pansu, 2003), à performances égales, les élèves blancs seraient mieux jugés que les autres. Merle (2018) précise que le résultat de cette étude anglo-saxonne n'est pas forcément transposable en France. Selon Merle (2018), il n'y aurait d'ailleurs aucune recherche statistique sur l'effet de l'origine ethnique sur les pratiques d'évaluation puisque cette caractéristique reste taboue.

En Belgique, nous pouvons mener ce type de recherches afin de vérifier l'influence de l'origine ethnique sur les jugements scolaires. Dans le cadre de son mémoire, Schépens (2017) avait mis en évidence qu'en tenant sous contrôle le milieu socio-économique, les enseignants faisaient preuve d'une plus grande sévérité envers les élèves issus de l'immigration, que ceux-ci soient forts ou faibles. Schépens (2017) précise que les élèves natifs sont privilégiés de deux points de performance pour les élèves forts et de moins d'un point pour les élèves faibles. Le public de cette recherche correspondait aux élèves de 1^e année de l'enseignement secondaire de la FW-B et leur enseignant de français.

Par contre, une étude française réalisée par Amigues *et al.* (1975, cités par Merle, 2018) s'est focalisée sur l'effet du patronyme français ou étranger sur le jugement scolaire. Les enseignants sollicités ont dû noter 8 rédactions dont 4 étaient censées être rédigées par des élèves français (par exemple Bertrand) et les 4 autres par des élèves étrangers (par exemple Hadjedi). Il en résulte que le patronyme de l'élève ne biaiserait pas l'évaluation puisqu'aucune différence significative n'a pu être observée.

Il nous paraît important de préciser que toutes les recherches anglo-saxonnes n'ont pas abouti à la même conclusion de Dusek et Joseph (1983, cités par Bressoux et Pansu, 2003). Si nous nous référons à l'étude menée par Feinberg et Shapiro en 2003 (cités par Meissel *et al.*, 2017),

l'origine ethnique des élèves ne biaisait pas le jugement des enseignants, toutes autres choses égales par ailleurs.

c) La classe sociale

En 1999, Oget (cité par Merle, 2018) a mené une recherche auprès d'adolescents qui passaient le baccalauréat afin de déterminer l'effet de l'origine sociale sur l'évaluation des compétences pendant l'année scolaire. En comparant les moyennes annuelles administrées par leurs enseignants et les résultats de l'épreuve certificative, il observe que les enfants d'ouvriers obtiennent généralement des meilleurs résultats au baccalauréat par rapport à la moyenne des notes obtenues tout au long de l'année. Il en conclut que les enfants d'origine populaire sont victimes des pratiques d'évaluation non-anonyme, contribuant ainsi à leur échec scolaire (Merle, 2018). Bressoux et Pansu (2001) s'étaient également intéressés au métier du père en affirmant que le fils d'un ouvrier était généralement jugé plus sévèrement en mathématiques que l'élève dont le père exerce une profession du cadre supérieur à performances égales.

Bressoux et Pansu (2003) font référence à plusieurs recherches en ce qui concerne les biais d'évaluation liés à l'origine sociale pour des élèves plus jeunes, notamment l'étude de Duru-Bellat et Mingat (1993) ainsi que celle menée par Jarlegan (1999). Pour la première, les chercheurs ont comparé les moyennes annuelles en mathématiques et en français avec leurs résultats aux épreuves standardisées de 2500 élèves de sixième répartis dans 17 collèges (Merle, 2018). Duru-Bellat et Mingat (1993) ont mis en évidence que les élèves du collège issus d'un milieu favorisé étaient généralement mieux notés que les autres. En ce qui concerne la deuxième recherche, Jarlegan en vient à la même conclusion mais précise que les élèves favorisés de CE1 sont mieux jugés au niveau de leurs résultats mais aussi de leur potentiel, leur intérêt, leurs efforts et leur attention. Jarlegan tire les mêmes conclusions pour les élèves de CM2 même si les résultats sont moins significatifs. Quoiqu'il en soit, Bressoux et Pansu (2003) s'accordent à dire, qu'à leur connaissance, aucune étude ne trouve d'effet négatif qui indiquerait un meilleur jugement quand le statut socio-économique de l'élève est plus faible. Merle (2018) souligne de son côté que le biais d'évaluation favorable aux enfants de catégories aisées est particulièrement constant dans la littérature scientifique internationale.

Ainsi, parmi les recherches anglo-saxonnes, Meissel *et al.* (2017) abordent une étude réalisée en 2011 par Ready et Wright qui avaient conclu que les jugements des enseignants étaient biaisés négativement dans le domaine de la langue pour les enfants de l'école maternelle issus d'un milieu social défavorisé. Hamilton *et al.* (1990) ainsi que Jussim (1986, 1991) ou encore

Rist (1970) en étaient arrivés à la même conclusion concernant les élèves scolarisés, quel que soit leur niveau d'étude.

C'est d'ailleurs auprès de 106 enseignants de niveaux différents (école élémentaire, « middle school » et « high school ») qu'Auwarter et Arugate ont mené leur recherche en 2008 au Missouri. Chaque professionnel participant se voyait recevoir une description fictive d'un élève recensant leur sexe, leur niveau socio-économique, leurs difficultés scolaires et des informations sur leur comportement. Il résulte que les professeurs prédisent un avenir moins prometteur aux élèves issus d'un milieu défavorisé, ce qui rejoint les études précédemment citées. En faisant référence à une étude réalisée par Pourtois (1978), Merle (2018) déclare que le « niveau d'expectation » du correcteur (p.153) est influencé par l'origine sociale des élèves. Selon lui, l'enseignant en associant des compétences intellectuelles plus grandes aux enfants des catégories aisées, contribue à la reproduction des inégalités. Merle (2018) affirme qu'un enfant avec un niveau socio-économique faible est catalogué et donc, désavantagé.

d) Le retard scolaire

Felouzis (1997) associe le retard scolaire à un handicap pour les élèves car ceux-ci verraient leur moyenne diminuée par année de retard. Toutes les études réalisées s'accordent pour conclure que les élèves dans une situation de retard scolaire sont plus sévèrement jugés par leur enseignant (Bressoux et Pansu, 2003). Puisque ce facteur prouve l'existence de difficultés scolaires, les enseignants construisent une étiquette négative à l'encontre des élèves concernés (Bressoux et Pansu, 2003). A contrario, Felouzis (1997) aborde l'importance des attentes positives sur l'évolution de l'élève. En d'autres termes, plus l'enseignant s'est construit une image positive de l'élève, meilleures sont ses attentes et plus son jugement sera favorable. En effet, selon les informations dont dispose l'enseignant sur le parcours scolaire de l'élève, des attentes vont ainsi se créer et le professionnel va adapter son jugement de sorte à les confirmer (Bressoux et Pansu, 2003). La réputation d'un élève considéré comme faible ou compétent est une caractéristique primordiale (Felouzis, 1997). Plus précisément, un élève en difficulté scolaire sera victime d'un effet de stéréotypie (Crahay, 2007). Selon ses résultats antérieurs, les filières fréquentées et si l'élève a rencontré ou non le redoublement, ce dernier sera jugé plus sévèrement. Puisque les enseignants ont tendance à attribuer les mêmes notes tout au long de l'année malgré les variations significatives des performances des élèves, il y a plus de risques qu'un élève soit dans une situation d'échec tout au long de l'année s'il n'a pas su se

montrer compétent dès les premiers mois, ses progrès n'étant pas pris en compte (Crahay, 2007).

Murat (1998, cité par Merle, 2018) a tenté de montrer que les élèves redoublants sont victimes de sous-notation systématique par rapport aux élèves dits « à l'heure ». Au test externe de compétences, les collégiens de 3^e ont à peu près la même note qu'ils aient connu ou non le redoublement. Par contre, dès qu'on s'intéresse aux notes attribuées par l'enseignant, la différence est de plus de 3 points sur 20 en faveur des élèves à l'heure. En ce qui concerne plus précisément les élèves faibles, la différence est plus petite et avoisine 1 point. Ces résultats sont donc en accord avec les recherches réalisées par Bressoux et Pansu (2003). L'étude d'Oget (1999, cité par Merle, 2018) va également dans ce sens concernant les élèves soumis au baccalauréat. Cette épreuve externe permettant un taux de réussite plus élevé des étudiants redoublants que si on se limitait au diplôme final administré selon les notes du professeur correcteur.

e) Le niveau scolaire

Au cours de la fin du siècle dernier, plusieurs études ont été menées dans le but de vérifier si le niveau scolaire de l'élève modifiait ou non l'exactitude du jugement professoral.

Après l'analyse de 16 études, Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) ont révélé que les jugements des enseignants apparaissaient plus valides lorsqu'un élève considéré comme fort était évalué par rapport à un élève faible. La conclusion d'une autre étude réalisée par Leinhardt (1983, cité par Sudkamp *et al.*, 2012) va dans le même sens : le degré de précision du jugement établi par l'enseignant est plus élevé dans le cas où les performances sont meilleures.

Merle (2018) fait référence à une étude réalisée en 1975 par Caverni, Fabre et Noizet auprès de 12 professeurs, il s'agit d'une des premières expériences se focalisant sur la connaissance de l'enseignant du niveau scolaire de leurs élèves sur l'évaluation de leur compétence. Dans un premier temps, 6 rédactions devaient être corrigées. Celles-ci étaient soit très bonnes soit très mauvaises dans le but d'obtenir un certain écart entre les notes pour le bon déroulement de la recherche, les cotes attribuées devaient avoisiner les 15/20 et les 5/20. Dans un second temps, les 6 autres professeurs recevaient une copie d'un niveau moyen où était indiquée la note que l'élève en question avait reçue pour la rédaction précédente. Notons que l'association entre les secondes copies et les premières notes étaient réalisées au hasard. Il

apparaît que lorsque la note précédente était élevée, la moyenne des copies s'élevait à 11,9/20 alors que pour les élèves qui avaient reçu une note médiocre à la première rédaction, la moyenne de leur résultat redescendait à 9,8/20.

Selon Merle (2018), cet écart de 2 points est intéressant pour 3 raisons :

- Il validerait l'existence de l'effet de Halo. En effet, l'information dont dispose le professeur au préalable modifie sensiblement son niveau d'appréciation. Merle (2018) parle d'un effet d'assimilation qui non seulement mettrait en évidence le flou inhérent aux critères de correction mais qui aussi, s'expliquerait par la nécessité inconsciente du correcteur d'éviter toute dissonance cognitive entre les deux notes. Cette dernière remarque nous renvoie directement aux « arrangements » développés dans notre 2^e chapitre.
- Il mettrait en évidence que l'évaluation des compétences d'un élève fondée sur le contrôle continu serait sensiblement biaisée par ses premières performances. Le professeur sous-estime la progression de ses élèves puisqu'il se fait une idée du niveau de chacun de ses élèves. Nous avons déjà abordé cette inertie scolaire (Caverni et Noizet, 1979, cités par Merle, 2018) lorsque nous nous intéressions à la pertinence de la note chiffrée.
- Les attentes des professeurs en termes de réussite scolaire sont supérieures pour les meilleurs élèves et favoriseraient les prophéties auto-réalisatrices (Bressoux et Pansu, 2004, cités par Merle, 2018). En effet, d'après Chaikin, Sigler et Derlega (cités par Merle, 2018), les professeurs ont des comportements spécifiques selon le niveau scolaire de leurs élèves et valoriseraient davantage les meilleurs élèves.

f) L'attractivité physique

De manière générale, les représentations communes sous-estiment largement l'impact de l'apparence physique dans les sentiments et jugements de valeurs interpersonnels (Dépret et Filisetti, 2001). Mais qu'en est-il pour les jugements portés dans un cadre scolaire ?

D'après Bressoux et Pansu (2003), les caractéristiques physiques ou socio-affectives constitueraient la base du noyau central des représentations à partir desquels l'enseignant jugerait ses élèves. Que ce soit la beauté, l'hygiène ou le capital de sympathie, il est évident que ces facteurs ne sont pas directement liés à la compétence de l'élève mais pourtant, ces

variables influencent significativement le jugement scolaire. En 1972, Dion *et al.* (cités par Bressoux et Pansu, 2003) avaient déjà énoncé cette phrase d'après leurs recherches : « Ce qui est beau est bon ». De plus, d'après leur méta-analyse, Dusek et Joseph (1983, cités par Bressoux et Pansu, 2003) ont conclu qu'à performances identiques, l'élève au physique attractif fait l'objet d'une évaluation plus favorable que les autres. Ritts *et al.* (1992, cités par Merle, 2018) vont plus loin en affirmant que l'apparence exerce également un effet positif sur l'appréciation de l'intelligence, du potentiel scolaire, des notes et des comportements de sociabilité. Toutefois, Bressoux et Pansu (2003) rappellent que Ritts *et al.* (1992) avaient souligné que la taille de l'effet est plus grande concernant les jugements sur les habiletés sociales que sur le jugement scolaire tel que nous l'avons défini dans ce travail.

Tarquinia et Tarquinia (2001) ont mené une recherche auprès d'enseignantes françaises d'une vingtaine d'années qui étaient en cours de formation initiale mais qui avaient déjà toutes eu la charge d'une classe depuis leur entrée en formation. Elles se sont vues administrer une vingtaine de dossiers d'élèves où apparaissaient une photo de l'enfant, son bulletin et une échelle d'intellectualité. Après avoir lu attentivement ce document, les sujets devaient graduer sur une échelle de type Likert leur jugement à propos du passage de l'élève en sixième, ses capacités d'adaptation, sa maturité, ses chances de réussite, ses compétences et sa motivation. Elles ont démontré que les élèves moyens bénéficiaient d'une évaluation plus favorable lorsque ceux-ci avaient un physique attractif, ce qui concorde avec la littérature précédemment citée. Par contre, en ce qui concerne les élèves dont le niveau est qualifié de moyen-faible, ce sont les élèves non-attractifs qui ont été privilégiés. Tarquinia et Tarquinia (2001) justifient ce résultat en déclarant que les enseignants ne pénalisent pas le stéréotype trop fortement ancré dans l'imaginaire collectif mais qu'ils ont plutôt tendance à sanctionner l'exemplaire. Les informations scolaires et physiques étant contradictoires, cela stigmatise le « faux-semblant » de la beauté des élèves. Cela apparaîtrait comme une forme de tromperie voire d'usurpation pour les enseignants, ce qui les amèneraient à pénaliser les élèves attractifs d'un niveau moyen-faible.

g) L'attitude en classe

D'après une étude menée par Bennett *et al.* en 1993, Bressoux et Pansu (2003) soulignent que si le comportement de l'élève est perçu comme positif au sein de la classe, son jugement scolaire sera meilleur. De plus, ils s'interrogent sur l'effet inverse : Le jugement sur la valeur scolaire influence-t-il le comportement des élèves ? Ces chercheurs parlent d'effets en boucle

qui se renforceraient mutuellement. En effet, puisqu'un professeur a tendance à surévaluer un élève qui a un comportement propice aux apprentissages, il est possible que les bonnes notes qu'il reçoit encouragent l'enfant à se comporter de la sorte. Suivant ce raisonnement, le comportement de l'élève influencerait les notes et, inversement, la cote revue inconsciemment à la hausse, aurait directement des conséquences sur l'attitude en classe. Autrement dit, il est possible que l'élève soit davantage motivé à continuer à se comporter de la sorte par les notes que lui attribue le professeur. C'est d'ailleurs ce qu'illustre une étude qualitative réalisée par Merle (1993, cité par Merle, 2018). Les garçons étant meilleurs en mathématiques qu'en français, ces derniers déclarent davantage se faire « bien voir » par le professeur qui dispense le cours où ils ont de meilleurs résultats. Merle (2018) parle alors de causalité circulaire et de renforcement mutuel entre les comportements scolaires adaptés, les compétences disciplinaires et les appréciations positives du professeur, ce qui rejoint le questionnement de Bressoux et Pansu (2003).

B. Selon les caractéristiques de l'enseignant

Même si Felouzis (1997, p.71) qualifient les enseignants comme étant « des juges habilités à prononcer un jugement scolaire », Bressoux et Pansu (2003) remettent en question l'exactitude de ce jugement selon le professeur. Suite à une étude assez ancienne, Helmke et Shrader (1987, cités par Bressoux et Pansu, 2003) ont conclu que certains professeurs avaient davantage tendance à modifier l'information relative à la valeur scolaire des élèves que d'autres professionnels exerçant la même fonction. Cette conclusion a amené certains chercheurs à se focaliser sur la variation du jugement scolaire en fonction des caractéristiques directement liées à l'enseignant.

a) Le genre

Dans leur étude expérimentale concernant l'influence du genre précédemment décrite, Lafontaine et Monseur (2009) mettent en évidence que les enseignants des deux sexes se montraient plus sévères avec les copies attribuées à des élèves de leur sexe. A contrario, ils seraient un peu plus laxistes avec les copies appartenant à des élèves du sexe opposé.

Même si Chouinard (2002) précise qu'il faut être prudent concernant les interprétations des études menées sur ce sujet puisqu'il y a généralement peu d'hommes dans les échantillons, il fait référence aux études menées par Brophy (1985), Duru-Bellat (1990) ainsi que celle de Parker-Price et Claxton (1996) qui avaient conclu que le sexe de l'enseignant n'avait pas de lien avec les notes qu'il attribue (Lafontaine et Monseur, 2009).

b) La personnalité professionnelle

L'évaluateur se construit des représentations sur son idéal pédagogique, sa conception générale de l'éducation, son propre parcours scolaire, son origine sociale ainsi que sur ses engagements politiques et associatifs qui risquent d'orienter son jugement professoral (Nicaise, 2002). D'après la méta-analyse de Hoge et Coladarci (1989, cités par Merle, 2018), les biais d'appréciation seraient plus importants parmi les professeurs qualifiés de « dogmatique » (p.158) comparativement aux enseignants plus ouverts. Artelt et Rausch (2014) soulignent l'importance de connaître les forces et les faiblesses de chaque élève par rapport à la tâche qui leur est proposée pour fournir un jugement précis. D'un autre côté, selon Bressoux et Pansu (2003), d'après une étude réalisée par Babad *et al.* (1982), les professeurs dogmatiques biaseraient davantage l'information en accentuant les différences entre les élèves. Merle (2018) qualifie les personnalités « dogmatiques » de conventionnelles et rigides, ce qui amènerait les professeurs à stigmatiser leurs élèves en fonction des stéréotypes sociaux en opposant et en simplifiant les catégories sociales, que ce soit les noirs et les blancs ou encore, les favorisés et les défavorisés.

c) Le niveau d'exigence

Au sein de notre système éducatif, un enseignant qui fait rater une certaine proportion d'élèves et qui se montre exigeant est davantage considéré comme un professeur compétent (Crahay, 2007). D'après une étude réalisée en 1992 par Lejong (citée par Crahay, 2007), tous les enseignants participants de mathématiques en quatrième année primaire avaient tendance à poser des questions d'examen du niveau supérieur à celui déterminé dans le programme. Il en est de même pour les professeurs de français ayant participé à la recherche APER-primaire, des écarts de niveau (estimés à 2 ans) selon l'enseignant ont été constatés dans des épreuves destinées à des élèves se situant au même stade dans leur scolarité. Or, les enseignants sont tenus de se fier au programme et d'évaluer les compétences déterminées selon le niveau du parcours scolaire. Du coup, il serait normal qu'un élève reçoive une note basse voire une note d'échec s'il n'a pas acquis les compétences sur lesquelles il n'est pas encore censé être évalué. Nous pouvons aller plus loin en affirmant que même pour les professeurs qui respectent les compétences du programme, il est possible de constater des niveaux de difficulté variables pour des questions évaluant un objectif similaire (Crahay, 2007). Ces constats expliquent le fait que des élèves, d'un même niveau de performance, obtiennent une note qui varie selon les épreuves auxquelles ils sont soumis par leur enseignant. Malgré

l'apparition des Socles de Compétences en 1997, le professeur, étant le seul maître au sein de sa classe, est toujours subjectif dans son évaluation (Nicaise, 2002).

Indépendamment du niveau de l'examen, la sévérité lors de la phase de correction joue un rôle majeur qui peut intervenir dans la justification de la variance du jugement pour une même performance. Effectivement, en plus de pouvoir s'interroger sur la réelle signification d'un point en fonction du professeur (Marchand, 1996), la pondération varie d'un enseignant à l'autre selon l'importance que ceux-ci accordent aux différentes sections d'une épreuve (Crahay, 2007). Rappelons que l'intensité de cette fluctuation dépend, entre autre, de la matière concernée, Felouzis (1997) jugeant les critères en mathématique plus stables et contrôlables que ceux relatifs au français. Quoiqu'il en soit, dans différentes commissions organisées selon la matière concernée, les enseignants sont parfois en désaccord concernant la compréhension de la consigne ou le niveau de justification et de démonstration, certains souhaitent un niveau d'exigence jugé excessif par d'autres. Les débats menés lors de ces commissions poussent à la catégorisation des professeurs à savoir, les « gentils » et les « non-gentils », malgré un barème (où sont codifiés les critères) très précis (Merle, 2018).

Quoiqu'il en soit, deux élèves ayant répondu de la même manière à un test identique, pourraient observer une différence de score obtenu selon l'attribution des points réalisée par le professeur lors de la phase de correction.

d) L'ancienneté

Même si Sudkamp *et al.* (2012) font référence aux recherches menées par Impara et Plake en 1998, ils affirment qu'il n'y a pas encore d'étude qui ont permis de mettre en évidence l'effet de l'ancienneté du professeur sur la précision du jugement scolaire.

Krolack-Schwerdt, Bohmer et Grasel (2009, 2012, cités par Artelt et Rausch, 2014) ont utilisé cette variable pour mettre en évidence l'importance de l'ancienneté du professeur dans la production d'un jugement scolaire précis. D'après leurs recherches, les enseignants les plus expérimentés seraient capables de mobiliser des ressources cognitives plus poussées afin d'émettre un jugement à haut enjeux. Par contre, les enseignants novices ne sembleraient pas adapter leur méthode selon l'importance de la décision. Autrement dit, plus l'enseignant est ancien dans la fonction, plus précis devrait être son jugement puisqu'ils ont plus facilement accès à des ressources cognitives leur permettant de poser un jugement plus juste. Artelt et Rausch (2014) rappellent que cette recherche a été menée en laboratoire et que les élèves sur

lesquels les enseignants devaient porter un jugement étaient fictifs, ce qui correspond à une limite de la recherche.

e) Le degré de similarité avec l'élève

D'un point de vue plutôt sociologique, le degré de ressemblance entre la source et la cible serait déterminant pour les recherches portant sur l'attraction et les attitudes interpersonnelles, cela pourrait découler d'une tendance narcissique généralisée (nous aimons ceux qui sont comme nous) mais aussi d'une satisfaction dérivée de la validation des attitudes propres (nous aimons ceux qui sont d'accord avec nous) (Dépret et Filisetti, 2001).

Dans le contexte scolaire, Bressoux et Pansu (2003) affirment que le degré de similarité entre l'élève (la cible) et l'enseignant (la source) peut affecter le jugement scolaire puisque les individus ont tendance à aimer davantage ceux qui leur ressemblent. Artelt et Rausch (2014) affirment que les élèves présentant le plus de points communs avec leur enseignant sont jugés plus positivement que le reste de la classe. En 1975, Rand et Wexley (cités par Artelt et Rausch, 2014) ont introduit le « similar-to-me effect », principe selon lequel la ressemblance mène à la sympathie, ce qui pourrait impacter positivement les jugements du professeur sur le niveau de compétence de l'élève.

C. Selon les caractéristiques liées au contexte

Les jugements des professeurs surviennent dans un contexte spécifique, ceux-ci sont donc au moins partiellement construits en fonction de la norme (Meissel *et al.*, 2017). L'importance du contexte fait donc directement référence à la classe et à l'établissement fréquenté.

a) Le niveau scolaire de la classe

Comme nous l'avons dit précédemment, les performances des élèves ne prennent du sens que dans la comparaison (Felouzis, 1997). Effectivement, c'est à travers le contexte-classe que l'enseignant va se faire une idée de la performance de ses élèves car son jugement sera influencé par le niveau moyen de la classe (Bressoux & Pansu, 2003). Le score moyen d'une classe a son importance puisque c'est en fonction de son niveau global que va varier la sévérité du jugement professoral.

D'après une étude menée par Bressoux et Pansu (2001) auprès de 19 classes de CE2, plus le niveau de la classe est élevé, plus l'enseignant a tendance à se montrer sévère dans l'évaluation des performances individuelles des élèves. Autrement dit, un élève peut recevoir

une note plus basse si le niveau moyen de ses condisciples est élevé. Au cours de la même étude, Bressoux et Pansu (2001) ont constaté que les jugements des enseignants variaient relativement peu en moyenne en fonction de la classe fréquentée puisque l'enseignant étalonne les valeurs scolaires en fonction du niveau de son groupe classe (Bressoux et Pansu, 2001). En effet, la part de variance inter-classe augmente significativement dans le cas des évaluations nationales, de 8,7 % en mathématique et de 10,4 % en français, elle atteint environ 25 % lorsque tous les élèves sont soumis à une épreuve standardisée. Cette sous-estimation de la variance inter-classe est une conséquence de la surestimation de la variance intra-classe. Autrement dit, la différence de performance est surestimée au sein d'une classe dans le but d'obtenir une dispersion adéquate mais entre les différents groupes-classes, leur moyenne de performance doit rester acceptable donc les différences sont sous-estimées. Crahay (2007) rappelle la logique suivant laquelle l'enseignant réalise des évaluations normatives visant à positionner le niveau de performance des élèves les uns par rapport aux autres afin d'obtenir une courbe de Gauss. Il fait référence à la loi de Posthumus que nous avons déjà développée dans ce travail. Par conséquent, un élève étant considéré comme faible dans une classe pourrait, à niveau de compétence égale, se retrouver parmi les plus forts d'un autre groupe. Bressoux et Pansu (2003) soulignent également l'importance de centrer la moyenne de la classe autour d'un score qui serait accepté socialement, ce qui nous renvoie aux différents arrangements auxquels l'enseignant doit faire face. Cette moyenne qui se doit d'être acceptable permet d'expliquer pourquoi le jugement des enseignants varie moins d'une classe à l'autre comparativement aux moyennes établies sur base de résultats obtenus à des épreuves standardisées (Bressoux et Pansu, 2003).

Plus récemment, dans une étude menée par Morlaix (2016, citée par Merle, 2018) auprès de 800 collégiens, la moyenne en mathématiques des élèves scolarisés en réseaux d'éducation prioritaire (REP) est de 12,3/20 comparativement à 11/20 pour les élèves scolarisés hors collèges REP. Alors qu'en nous intéressant aux tests PISA, la moyenne des scores des élèves bénéficiant d'un enseignement spécifique (5,59/20) devient inférieure à celle des élèves hors REP (6,12/20) ayant des caractéristiques sociales équivalentes. Ces observations rejoignent les études menées par Bressoux et Pansu (2001). Merle (2018) parle d'un effet de contexte pédagogique spécifique selon lequel, les enseignants sont amenés à évaluer différemment leurs élèves en se basant sur le niveau scolaire de la classe.

b) L'indice socio-économique de l'établissement

Meissel *et al.* (2017) ont mené une étude en Nouvelle-Zélande en 2012 et 2013 où l'un des objectifs était d'apporter des éclaircissements concernant l'impact éventuel du SES de l'établissement sur l'exactitude du jugement scolaire. Ils ont tenté de répondre à cette question pour le domaine de la lecture et de l'expression écrite. Concernant la lecture, leur échantillon était composé de 194 classes soit 4771 élèves répartis dans 44 écoles différentes. Même si tous les enfants participants étaient âgés de 8 à 13 ans, l'échantillon pour l'expression écrite s'élevait à 561 classes soit 11 765 élèves étant inscrits dans 105 établissements scolaires différents. Pour les deux domaines étudiés, il apparaît que le SES de l'établissement n'aurait pas d'effet significatif sur le jugement des enseignants.

Par contre, en 2011, Ready et Wright ont réalisé une recherche se focalisant également sur l'éventuel effet du SES de l'établissement sur les jugements professoraux. Cette recherche a été menée auprès de 1822 classes réparties dans 701 établissements scolaires. Dans ce cas-ci, le niveau socio-économique de l'école impacterait significativement les jugements des enseignants. Il en ressort que ceux-ci seraient plus favorables pour les élèves inscrits dans une école avec un SES moyen plus élevé.

4. Conclusion

Même si Bressoux et Pansu (2003) précisent que les jugements des enseignants sont valides puisque Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) avaient déterminé une corrélation de 0,66 entre les notes des enseignants et les scores représentant les performances des élèves, la revue de la littérature a mis en évidence la difficulté d'établir un jugement précis sur les performances des élèves. En effet, nous avons vu qu'un enseignant doit répondre à une série d'arrangements dans un contexte spécifique à la classe dans le but d'attribuer une note chiffrée aux élèves tout en veillant à ne pas altérer l'exactitude de ce jugement, ce procédé laisse place à une certaine subjectivité.

En nous basant sur le modèle construit par Sudkamp *et al.* (2012), nous avons listé les différents biais susceptibles d'influencer le niveau de précision du jugement scolaire en passant en revue les différentes recherches menées sur le sujet. Il en ressort que le sexe (Felouzis, 1997 ; Lafontaine et Monseur, 2009), l'origine sociale (Bressoux et Pansu, 2003 ; Merle, 2018), le retard scolaire (Bressoux et Pansu, 2003) de l'élève et la moyenne de la classe que ce dernier fréquente (Bressoux et Pansu, 2001) impactent le jugement scolaire.

III. Question de recherche et hypothèses

1. Question de recherche

Dans un contexte où les évaluations externes font partie intégrante du système éducatif, il nous paraît intéressant de poursuivre les recherches menées par Bressoux et Pansu en 2003 pour vérifier la concordance des notes que reçoivent les élèves par leur enseignant et les scores qu'ils obtiennent aux évaluations externes certificatives. Ces scores représentent les performances des élèves et leurs notes attribuées tout au long de l'année scolaire correspondent au jugement des enseignants.

Notre principale question de recherche correspond donc à : « Les jugements scolaires des professeurs de sciences en Fédération Wallonie-Bruxelles concordent-ils avec les résultats aux épreuves externes certificatives (CE1D) ? »

C'est en comparant les notes des enseignants et les scores obtenus au CE1D de sciences que nous allons tenter de répondre à cette question. Pour rappel, Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) avaient déterminé une corrélation de 0,66 entre ces 2 variables, ce qui avait permis de qualifier de « valide » le jugement des enseignants.

Mais Bressoux et Pansu (2003) avaient tenté d'expliquer la part de la variance non expliquée en menant une recherche qui visait à étudier l'impact des caractéristiques des élèves et du contexte spécifique de la classe sur le jugement professoral. Tout comme ces chercheurs et en nous basant sur la revue de la littérature de ce présent-travail, nous allons introduire des variables qui pourraient expliquer une éventuelle dissonance entre la note et la performance. Ces variables correspondent donc aux caractéristiques de l'élève, de l'enseignant et du contexte. Elles ont été présentées lorsque nous listions les biais pouvant altérer l'exactitude du jugement scolaire dans la partie théorique.

Cela nous mène à une sous-question que nous pouvons formuler de la sorte : « À performances égales, quelles sont les variables qui interviennent dans les notes que donnent les enseignants à leurs élèves ? »

A la suite de la détermination de ces variables et grâce aux analyses statistiques conduites sur les données que nous avons récoltées, les hypothèses que nous avons préalablement formulées pourront être vérifiées.

2. Hypothèses de recherche

Les hypothèses concernant notre principale question de recherche sont listées ci-dessous :

- **Hypothèse 1 :** En sciences, les élèves sont classés de la même façon selon que l'on considère les notes données par les enseignants pendant l'année et les scores obtenus au CE1D.

Dans leur étude de 2003, Bressoux et Pansu ont montré que les enseignants classaient relativement bien leurs élèves en respectant significativement le même ordre que celui dans lequel ils sont placés selon les résultats obtenus à une évaluation standardisée. Nous nous attendons donc à retrouver un classement similaire en observant notes et scores en sciences de élèves de la FW-B.

- **Hypothèse 2 :** Le classement réalisé par les enseignants concorde davantage avec celui obtenu sur base des scores des élèves au CE1D lorsqu'on l'observe pour chaque enseignant à l'intérieur de sa classe plutôt que pour l'échantillon complet.

Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) affirment que ces classements sont davantage similaires lorsqu'ils sont observés au sein d'une classe plutôt qu'au niveau de l'échantillon. Dans leur recherche, la corrélation médiane entre le jugement de l'enseignant et les performances aux épreuves standardisées passe de 0,64 à 0,70 lorsque l'appartenance aux classes est prise en compte.

- **Hypothèse 3 :** Au sein de chaque classe, la dispersion des notes de l'enseignant est plus importante que celle des résultats obtenus par les élèves au CE1D.

En nous référant à la recherche APER menée par Grisay (1984, cité par Crahay, 2007), les niveaux de compétence des élèves paraissent plus hétérogènes aux examens conçus par l'enseignant qu'aux tests standardisés. Cela nous renvoie à l'effet de Posthumus développé par De Landsheere (1980, cité par Crahay, 2007) qui nous dit qu'au sein des classes, les exigences sont adaptées de sorte à avoir une majorité d'élèves moyens et une minorité d'élèves forts et faibles.

- **Hypothèse 4 :** Il y a un lien entre la performance des élèves et les notes qui leur sont données par les enseignants.

Pour rappel, une corrélation assez forte (0,66) entre les notes des enseignants et le score aux épreuves standardisées ont permis de mettre en évidence que les professeurs jugeaient en

moyenne leurs élèves de manière juste (Hoge et Coladarci, 1989, cités par Bressoux et Pansu, 2003).

En ce qui concerne notre sous-question, des hypothèses plus spécifiques vont devoir être vérifiées :

- **Hypothèse 5 :** À performances égales, les notes données par l'enseignant ne diffèrent pas de manière statistiquement significative en fonction du sexe de l'élève pour la discipline des sciences.

Selon l'étude de Felouzis (1997), à performances égales, une fille obtient une meilleure note qu'un garçon en français. En ce qui concerne les mathématiques, Lafontaine et Monseur (2009) avaient déclaré que pour les garçons, les plus performants étaient surévalués et les moins performants étaient nettement sous-évalués. À l'inverse, les filles plus performantes étaient sous-évaluées alors que les moins performantes étaient surévaluées. Puisque ces constats diffèrent en fonction de la matière, il est intéressant de se poser cette question dans le cadre du cours de sciences, d'autant que peu d'études ont été réalisées dans ce domaine. Rappelons l'étude de Kacprzyk et al. (2019) qui s'étaient intéressés à l'effet du genre pour la biologie au niveau universitaire. Ils n'ont pas trouvé de différence significative entre les garçons et les filles sauf lorsque le retrait de point par mauvaise réponse était d'application, ce qui avait tendance à défavoriser les filles. En ce qui concerne le CE1D en sciences, il n'y a pas de retrait de point et le public est plus jeune. Selon les différentes études évoquées, nous pouvons émettre l'hypothèse qu'il n'y aura pas d'influence du genre sur les notes que reçoivent les élèves au cours de sciences.

- **Hypothèse 6 :** À performances égales, les notes données par l'enseignant diffèrent de manière statistiquement significative en fonction du milieu socioculturel de l'élève pour la discipline des sciences. Les notes sont surestimées pour les élèves dont le milieu socioculturel est favorisé.

Merle (2018) affirme que le biais d'évaluation favorable aux enfants de catégories aisées est particulièrement constant dans la littérature scientifique internationale. En effet, Bressoux et Pansu (2003) citent les recherches menées par Duru-Bellat et Mingat (1993) et de Jarlegan (1999) qui avaient mis en évidence que les élèves du collège issus d'un milieu favorisé étaient généralement mieux notés que les autres, à performances égales.

- **Hypothèse 7 :** À performances égales, les notes données par l'enseignant en sciences diffèrent de manière statistiquement significative en fonction du retard scolaire de l'élève pour la discipline des sciences. Les notes sont sous-estimées pour les élèves qui ont redoublé.

En 2003, Bressoux et Pansu déclarent que toutes les études réalisées (Duru-Bellat & Mingat, 1993 ; Jarlegan, 1999 ; Felouzis, 1997) s'accordent pour conclure que les élèves dans une situation de retard scolaire sont plus sévèrement jugés par leurs enseignants. De plus, Murat (1998, cité par Merle, 2018) avait mis en évidence que les collégiens en retard scolaire étaient victimes de sous-notation de la part de leurs enseignants.

- **Hypothèse 8 :** À performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction du sexe de l'enseignant pour la discipline des sciences.

Chouinard (2002) a fait référence aux études menées par Brophy (1985), Duru-Bellat (1990) ainsi que celle de Parker-Price et Claxton (1996) qui avaient conclu que le sexe de l'enseignant n'avait pas de lien avec les notes qu'il attribue (Lafontaine et Monseur, 2009).

- **Hypothèse 9 :** À performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction de l'ancienneté de l'enseignant pour la discipline des sciences.

D'après les recherches de Krolack-Schwerdt, *et al.* (2009, 2012, cités par Artelt et Rausch, 2014), les enseignants les plus expérimentés seraient capables de mobiliser des ressources cognitives plus poussées afin d'émettre un jugement à haut enjeux, ce qui signifierait que plus l'enseignant est ancien dans la fonction, plus précis devrait être son jugement. Ceci dit, puisque Artelt et Rausch (2014) rappellent que cette recherche a été menée en laboratoire et que les élèves sur lesquels les enseignants devaient porter un jugement étaient fictifs, nous n'irons pas dans ce sens pour la formulation de l'hypothèse. De plus, Impara et Plake (1998) affirmaient qu'il n'y a pas encore d'étude qui ont permis de mettre en évidence l'effet de l'ancienneté du professeur sur la précision du jugement scolaire (Sudkamp et al., 2012).

- **Hypothèse 10 :** À performances égales, les notes données par l'enseignant diffèrent de manière statistiquement significative en fonction de la moyenne de la classe de l'élève pour la discipline des sciences. Les notes sont sous-estimées dans les classes fortes et surestimées dans les classes faibles.

Crahay (2007) a mis en évidence l'importance du niveau scolaire de la classe sur la valeur scolaire d'un élève en rappelant l'importance de la comparaison intra-classe. Dans le cadre de leur étude, Bressoux et Pansu (2001) ont montré qu'un élève moyen reçoit une note diminuée si le niveau de sa classe est élevé. Inversement, Morlaix (2016, citée par Merle, 2018) avait mis en évidence que les élèves d'une classe faible avaient tendance à être surévalués par leur enseignant.

- **Hypothèse 11** : À performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction de l'indice socio-économique de l'établissement dans lequel se trouve l'élève pour la discipline des sciences.

Même si Ready et Wright (2011) ont trouvé un lien significatif entre le niveau socio-économique de l'école et les jugements scolaires, des études plus récentes n'ont pas abouti à la même conclusion. Selon Meissel *et al.* (2017), le SES de l'établissement n'aurait pas d'effet significatif sur les notes qu'attribuent les enseignants.

IV. Méthodologie

1. Description de la population et de l'échantillon

Puisque notre recherche se focalise sur les jugements des enseignants en sciences que nous comparons aux résultats du CE1D, notre population est constituée d'enseignants de 2^{ème} année commune de l'enseignement secondaire de la Fédération Wallonie-Bruxelles.

Notre échantillon comprend 23 enseignants en sciences travaillant tous dans des établissements différents ainsi que leurs 464 élèves. En raison de données manquantes, le nombre de participations des élèves s'abaisse à 441 sujets (soit un taux d'attrition de 4,95%). Il est important de préciser que cet échantillon n'est pas aléatoire et simple puisque nous avons dû obtenir l'accord des sujets avant de participer à cette recherche. Par conséquent, il ne serait pas correct d'élargir les résultats obtenus dans le cadre de ce travail à l'ensemble de la population.

Parmi les 23 enseignants, 18 travaillent dans le réseau libre subventionné, 3 dans l'enseignement organisé par la FW-B et 2 dans l'officiel subventionné. De plus, ils ont tous déclaré être tout à fait d'accord ou plutôt d'accord concernant l'adéquation du CE1D avec le socle des compétences. Initialement, nous voulions explorer l'effet éventuel de cette variable sur les notes qu'ils attribuent mais en raison du pourcentage d'enseignants étant d'accord (100%), nous n'avons pas pu traiter des données statistiques à ce sujet. Mis à part ces 2 indications sur les enseignants, les autres informations vous seront présentées lors de la vérification des hypothèses.

2. Méthodes et instrument.

A. Collecte des données

Puisque l'objectif de notre étude est de vérifier la concordance entre le jugement scolaire et la performance à l'épreuve externe, nous avons décidé de mener une étude quantitative pour laquelle nous avons dû récolter plusieurs données, à savoir :

- la note du bulletin obtenue en 3^{ème} période par l'enseignant dans le cours d'initiation scientifique ;
- le score obtenu au CE1D en sciences ;
- des informations sur les élèves (voir ci-dessous) ;
- des informations sur les enseignants et leur classe (voir ci-dessous).

a) Collecte des notes au bulletin et des scores au CE1D

Nous avons fait le choix de sélectionner les notes du bulletin de la 3^{ème} période puisque cette cote se rapproche le plus des scores obtenus à l'épreuve externe dans le temps. Notre choix étant déterminé par le souci d'utiliser des données les plus comparables possibles. Nous pouvons supposer que l'enseignant ne se soit pas focalisé exclusivement sur les derniers chapitres de l'année pour attribuer ses notes mais qu'il ait consacré du temps pour retravailler voire évaluer des compétences plus globales qui sont intégrées chaque année dans l'épreuve externe. Quoiqu'il en soit, si l'on se réfère à Crahay (2007), les notes entre les différentes périodes ne devraient pas être considérablement différentes. Pour rappel, les élèves ne verraient pas directement leurs progrès en se focalisant sur la note qu'ils reçoivent puisqu'ils sont victimes d'une inertie scolaire en raison de la comparaison et du classement intra-classe.

En ce qui concerne les scores obtenus aux épreuves externes, nous pensons que ces résultats sont les plus adéquats pour représenter leur performance. En effet, l'épreuve est commune à tous les élèves de notre échantillon et les enseignants reçoivent un guide de passation et de correction. Cet outil permet de limiter les biais de correction et ceux introduits par le test en lui-même, c'est d'ailleurs la raison pour laquelle ces types de biais n'ont pas été présentés dans la partie théorique. De plus, les enseignants sont assez libres pour organiser les corrections. Certains s'organisent en travaillant en équipe alors que d'autres sont amenés à corriger toutes les copies de leurs élèves. N'ayant pas d'emprise sur le déroulement de la correction, nous n'avons pas introduit de variables sur ce paramètre. Dans tous les cas, il est difficile d'obtenir la performance « vraie » d'un élève puisqu'il existe toujours des erreurs de mesure et d'interprétation.

La note et le score des élèves nous sont parvenus via le formulaire que nous avons préalablement construit (annexe 1) afin d'alléger la tâche des enseignants participant à notre recherche.

b) Collecte des informations sur les élèves

Afin de pouvoir introduire des variables pouvant expliquer les variations du jugement scolaire, nous avons construit un questionnaire destiné aux élèves (annexe 2) nous permettant de récolter différentes données sur leurs caractéristiques. Nous nous sommes basés sur l'outil utilisé dans le cadre de l'étude PIRLS 2016 pour sélectionner, adapter et présenter nos

questions. En fonction de nos hypothèses préalablement formulées, les différents items portent sur :

- l'année de naissance de l'élève ;
- le sexe de l'élève ;
- la fréquentation ou non de l'école maternelle ;
- le parcours scolaire (redoublement en primaire et fréquentation d'une année différenciée) ;
- la langue parlée à la maison ;
- l'âge d'arrivée en Belgique ;
- la profession des parents ;
- le nombre de livres à la maison ;
- la possession de différents biens tels qu'un système d'alarme, un ordinateur, un téléphone portable personnel, un bureau pour travailler, ...

c) Collecte des informations sur les enseignants

Dans le but de ne pas nous limiter aux variables liées aux caractéristiques des élèves, nous avons construit un deuxième questionnaire (annexe 3) devant être administré aux enseignants. Au vu de la revue de littérature, certaines hypothèses sont formulées de sorte à vérifier l'impact de certaines caractéristiques des enseignants et du contexte sur le jugement scolaire. Dans ce cas-ci, nous n'avons pas utilisé les items d'une autre étude pour le construire. Ce questionnaire « enseignant » a été conçu pour récolter des données portant sur :

- la matière concernée ;
- le réseau d'enseignement ;
- l'ancienneté ;
- la taille de la classe ;
- la perception de l'enseignant du nombre d'élèves issus de l'immigration ;
- la perception de l'enseignant du nombre d'élèves faibles ;
- la perception de l'enseignant du nombre d'élèves attentifs ;
- la perception de l'enseignant du nombre d'élèves aux comportements perturbateurs ;
- la perception de l'enseignant du degré de difficulté du CE1D ;
- la perception de l'enseignant de l'adéquation du CE1D au programme ;
- la perception de l'enseignant de l'adéquation du CE1D aux Socles de compétences.

B. Modalités d'administration

Dans le but de recueillir des informations sur les élèves et les enseignants, nous avons transmis les différents questionnaires et le formulaire aux enseignants. Ceux-ci se sont chargés de compléter le formulaire en inscrivant les différentes notes et les scores obtenus à l'épreuve de fin d'année, de soumettre le questionnaire aux élèves et de répondre à celui qui leur était adressé. A travers ces outils, nous avons rappelé l'objet de notre étude et assuré l'anonymat des données que nous récoltions. Les échanges étaient assurés par voie postale et se sont effectués en fin d'année scolaire 2019.

C. Comité d'éthique

Notre recherche a reçu un avis favorable du comité d'éthique de la faculté de Psychologie, Logopédie et Sciences de l'Éducation de l'ULiège. La partie privée des données sera stockée pour la durée nécessaire à la réalisation de l'étude. Après cette période, les données seront détruites soit de manière informatique soit à l'aide d'une déchiqueteuse. Pour rappel, la participation était facultative et n'engendrait aucun risque d'une quelconque nature. Aucun jugement ni préjugé ne sera porté sur les réponses des participants. Ceux-ci, en décidant sur base volontaire de répondre au questionnaire, contribuent à l'avancée de notre recherche. Les participants ont été informés de l'objectif de notre étude et ont accepté d'y prendre part.

3. Traitement des données

A. Variables

Comme mentionné plus haut, les questionnaires ont été construits dans le but d'introduire différentes variables susceptibles d'influencer le jugement scolaire. Dans le but d'exploiter dans nos analyses statistiques les données que nous avons récoltées, nous avons procédé à un codage. Il est important de souligner que toutes les données récoltées n'ont finalement pas été retenues dans le codage des variables. En effet, certaines informations étaient difficilement exploitables et ne nous auraient pas permis de tirer des conclusions.

a) Variables « note » et « score » :

- La variable reprenant les notes des enseignants correspond à notre variable dépendante. Ces notes sont exprimées sur 100.

- La variable reprenant les scores obtenus à l'épreuve externe correspond à la variable que nous devons tenir sous contrôle afin d'interpréter « à performances égales ». Celle-ci est standardisée, sa moyenne est égale à 0 et son écart-type vaut 1.

b) Variables liées aux caractéristiques des élèves :

- La variable concernant **le sexe de l'élève** nous permet d'observer si, à performances égales, les filles(1) et les garçons (2) obtiennent des notes différentes.
- La variable concernant **le retard scolaire** nous permet d'observer si, à performances égales, les élèves « à l'heure » (0) et « en retard » (1) obtiennent des notes différentes. Dans ce cas-ci, nous nous sommes basé sur la question concernant le redoublement en primaire.
- La variable concernant **la langue parlée à la maison** nous permet d'observer si, à performances égales, les élèves déclarant parler français à la maison tout le temps ou presque (1) et ceux qui ne parlent que parfois ou jamais cette langue (2) obtiennent des notes différentes.
- La variable concernant **le niveau socio-économique de l'élève (SES)** nous permet d'observer si, à performances égales, les élèves issus d'un milieu favorisé et ceux d'un milieu défavorisé obtiennent des notes différentes. Pour cette variable, nous avons sommé les réponses de différents items à savoir : le nombre de livres à la maison, la possession de tel ou tel bien matériel et la profession des parents. En ce qui concerne cette dernière information, nous avons catégorisé les différents métiers en les classant dans 12 domaines bien distincts :

(1) sans travail	(5) agriculture et pêche	(9) travail de gestion
(2) indépendant	(6) manuel qualifié	(10) travail intellectuel qualifié
(3) travail de bureau	(7) travail en usine	(11) travail technique qualifié
(4) services et ventes	(8) travail non qualifié	(12) sans objet

Pour ensuite les regrouper en 4 catégories de la manière suivante :

- (0) catégorie 12
- (1) catégories 1, 5, 6, 7 et 8
- (2) catégories 2, 3 et 4
- (3) catégories 9, 10 et 11

Dans les cas où l'élève a répondu à cette question pour ses 2 parents, nous avons décidé de conserver la profession ayant le codage le plus élevé.

c) Variables liées aux caractéristiques de l'enseignant :

- La variable concernant **le sexe de l'enseignant** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes selon que leur professeur soit une femme (1) ou un homme (2).
- La variable concernant **l'ancienneté de l'enseignant** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes selon que leur professeur exerce depuis moins de 10 ans (1) ou au minimum 10 ans (2) dans la fonction.
- La variable concernant **la perception de l'adéquation du CE1D avec les Socles de compétences** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes lorsque leur professeur n'est pas du tout ou plutôt pas d'accord sur cette adéquation (1) comparativement à lorsque l'enseignant se dit tout à fait ou plutôt d'accord avec cette adéquation (2). Comme expliqué précédemment, cette variable n'a pas pu être retenue pour nos analyses statistiques, au vu de l'unanimité des réponses formulées.

d) Variables liées au contexte :

- La variable concernant **la taille de la classe** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes si la taille de leur classe est inférieure (1) ou égale et supérieure (2) à 20 élèves.
- La variable concernant **la moyenne de la classe** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes en fonction de la moyenne de leur classe. Dans ce cas-ci, nous avons calculé la moyenne au CE1D pour chaque classe et nous avons standardisé les scores obtenus avec une moyenne de 0 et un écart-type de 1.
- La variable concernant **le niveau socio-économique de l'établissement** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes en fonction des indices indiquant le niveau socio-économique de l'établissement. Ceux-ci sont recensés dans une liste que nous avons consultée afin d'apparier l'indice adéquat à chaque institution scolaire ayant participé à notre recherche.

B. Analyses statistiques

Dans le but de répondre à nos hypothèses, nous avons réalisé différentes analyses statistiques que nous allons décliner en 4 étapes. Comme expliqué dans la partie consacrée à la question de recherche, nous allons d'abord nous renseigner sur la concordance des notes de l'enseignant avec les scores obtenus au CE1D à un niveau plus global puis nous nous intéresserons aux différentes variables contextuelles, ce qui nécessite d'autres démarches statistiques.

➤ Étape 1 : la corrélation de rang

Il s'agit ici de calculer une corrélation de rang, également appelée de Spearman dans le but de vérifier la concordance entre les notes de la 3^{ème} période et les scores obtenus au CE1D. Cette étape nous permettra de vérifier si le classement suivant l'ordre des notes attribuées par l'enseignant est similaire à celui établi lorsqu'on ne se focalise que sur les scores obtenus à l'épreuve externe. Dans un premier temps, nous avons calculé le coefficient de corrélation pour l'ensemble des données de notre échantillon puis nous avons répété cette opération pour chaque classe. Autrement dit, cette corrélation a d'abord été calculée pour toutes les classes confondues puis pour chaque classe séparément.

➤ Étape 2 : l'écart-type

Afin d'analyser la dispersion des notes attribuées par l'enseignant et celle des scores obtenus à l'épreuve externe, nous avons calculé les différents écart-types pour ces deux types de données. Nous nous sommes d'abord intéressés à la dispersion au sein de chaque classe puis entre les différentes classes en utilisant les moyennes de celles-ci.

➤ Étape 3 : le modèle vide

Cette étape est une condition d'application de la suivante. Pour mener à bien les régressions multiniveaux, nous avons dû créer un modèle vide. Cette opération nous permettra également d'observer la part de variance à l'intérieur et entre les classes au niveau des notes données par les enseignants.

➤ Étape 4 : la régression multiniveaux

Enfin, nous avons réalisé des régressions multiniveaux pour chaque variable étudiée afin de vérifier si le lien entre chacune d'entre elles et les notes de l'enseignant est significatif. Le lien devient significatif au seuil de 0.05 (soit un risque de première espèce inférieur à 5%). Pour ce faire, nous avons d'abord tenu sous contrôle le score obtenu au CE1D considéré comme la

performance de l'élève, cela nous a permis de réaliser les analyses statistiques « à performances égales ». Ensuite, nous avons introduit les différentes variables pour vérifier leur éventuel lien sur la note que les élèves reçoivent par leur enseignant. Nous avons donc réalisé les analyses statistiques sur les variables liées aux caractéristiques de l'élève puis à celles de l'enseignant et enfin, nous avons terminé avec les variables liées au contexte scolaire.

V. Présentation des résultats

1. Concordance entre les notes de l'enseignant et les scores au CE1D

A. Concordance toutes classes confondues

Dans le but de vérifier notre 1^{ère} hypothèse et donc de déterminer si les élèves sont classés de la même façon suivant les notes données par les enseignants pendant l'année et les scores obtenus au CE1D. Après avoir calculé la corrélation de Spearman, également appelée corrélation de rang, nous avons pu vérifier cette hypothèse. Il s'avère que cette corrélation s'élève à 0.66 et est statistiquement significative. Elle peut donc être qualifiée d'assez forte. Cela signifie que les élèves faibles selon les résultats à l'épreuve externe sont également ceux qui reçoivent les notes les moins élevées de la part de leur enseignant, et inversement.

- **Notre hypothèse 1 est donc confirmée au sein de notre échantillon** : en sciences, les élèves sont classés de la même façon suivant les notes données par les enseignants pendant l'année et les scores obtenus au CE1D.

B. Concordance classe par classe

Dans ce cas-ci, nous avons calculé le coefficient de corrélation de Spearman pour chaque classe afin de vérifier la concordance entre les notes de l'enseignant et les scores au CE1D au sein des différentes classes ayant participé à notre recherche. Dans le tableau ci-dessous, les différents coefficients de corrélations sont listés de sorte à vérifier si les enseignants classent les élèves de leur propre classe de la même façon que les épreuves externes.

Classes	Corrélations de Spearman	Prob > r	Classes	Corrélations de Spearman	Prob > r
1	0.45 (min)	0.0310	13	0.61	0.0028
2	0.72	0.0012	14	0.65	0.0014
3	0.45	0.0774	15	0.62	0.0029
4	0.54	0.0072	16	0.60	0.0053
5	0.62	0.0026	17	0.33	0.1191
6	0.60	0.0024	18	0.53	0.0091
7	0.35	0.1897	19	0.59	0.0164
8	0.59	0.0050	20	0.58	0.0615
9	0.73	0.0004	21	0.55	0.0184
10	0.35	0.0981	22	0.88 (max)	< 0.0001
11	0.72	0.0005	23	0.41	0.1156
12	0.67	0.0087			

Tableau 2 Corrélations de Spearman classe par classe entre les notes des enseignants et les scores au CE1D

En observant les données du tableau ci-dessus, nous pouvons constater que 6 coefficients de corrélation de Spearman ne sont pas statistiquement significatifs au seuil de 0.05 (lignes grisées du tableau). En ce qui concerne les corrélations de rang significatives, elles varient de 0.45 à 0.88. Cette variation permet d'affirmer que certains enseignants de notre échantillon classent mieux leurs élèves que d'autres.

En comparant ces résultats à la corrélation de Spearman calculée pour toutes les classes confondues (**0.66**), nous observons que seulement 4 classes ont un coefficient supérieur et 2 ont obtenu une valeur approximativement similaire. En ce qui concerne les autres classes, nous constatons que 11 classes ont une corrélation de rang inférieure à 0.66. Concrètement, cela signifie que les classements des notes des enseignants concordent davantage avec ceux réalisés sur base des scores obtenus à l'épreuve externe pour toutes les classes confondues par rapport au classement au sein de chaque classe.

- **Notre hypothèse 2 n'est pas confirmée au sein de notre échantillon** : le classement réalisé par les enseignants ne concorde pas davantage avec celui obtenu sur base des résultats des élèves au CE1D lorsqu'on l'observe pour chaque enseignant à l'intérieur de sa classe plutôt que pour l'échantillon complet.

C. Dispersion des notes à l'intérieur des classes.

Classes	Moyennes des notes des enseignants	Moyennes des scores au CE1D	Écart-types des notes des enseignants	Écart-types des scores au CE1D
1	79.24	69.24	9.61	6.84
2	61.18	51.47	10.43	9.77
3	70.94	54.06	12.51	9.30
4	68.70	61.52	14.08	9.50
5	54.17	56.67	13.05	10.35
6	50.65	46.09	11.41	9.26
7	61.09	55.16	11.22	10.51
8	64.40	55.48	12.04	10.39
9	56.97	49.74	11.98	11.21
10	77.83	65.54	9.95	9.32
11	72.89	63.16	10.65	9.05
12	58.93	55.18	12.16	9.01
13	73.64	61.36	11.49	13.56
14	64.67	62.24	11.59	13.34
15	54.05	50.67	17.75	8.33
16	61.75	44.25	18.87	13.79
17	54.72	53.26	13.50	12.69
18	74.75	59.26	19.49	13.56
19	62.19	59.71	20.08	11.85
20	50.18	49.18	16.31	8.68
21	67.78	57.94	11.53	10.95
22	62.73	55.23	17.02	13.23
23	55.31	45.25	11.90	10.86
Toutes classes confondues	63.95	56.12	15.65	12.50

Tableau 3 Moyennes et écart-types des notes de l'enseignant et des scores au CE1D

Si nous focalisons notre attention sur la colonne de droite, nous pouvons constater que la quasi-totalité des enseignants constituant notre échantillon ont tendance à disperser davantage

les notes qu'ils attribuent comparativement à la dispersion obtenue sur la base des scores du CE1D. À l'exception des classes 13 et 14, tous les écarts-types sont supérieurs pour les notes des enseignants. Il s'agit donc d'une tendance clairement marquée nous permettant d'interpréter ces résultats. L'écart-type de toutes classes confondues pour les notes de l'enseignant est supérieur de plus de 3 points par rapport à celui pour les scores de l'évaluation externe.

- **Notre hypothèse 3 est confirmée pour notre échantillon** : au sein de chaque classe, la dispersion des notes de l'enseignant est plus importante que celle des résultats obtenus par les élèves au CE1D.

Nous avons également inséré les différentes moyennes au sein de ce tableau puisqu'à nouveau, une tendance assez marquée mérite d'être soulignée. En effet, à l'exception de la classe 5, les élèves semblent être surévalués par leur enseignant comparativement au test externe. Il s'avère que la moyenne des notes est inférieure à la moyenne des scores pour la classe 5. Pour toutes les autres classes, ce constat est inversé. En ce qui concerne la moyenne pour toutes les classes confondues, celle pour les notes des enseignants s'élève à 63.95 et celle des scores du CE1D équivaut à 56.12 marquant ainsi une différence d'environ 8 points sur 100. Les enseignants au sein de l'échantillon ont donc tendance à surestimer la performance moyenne de leurs élèves, mais avec davantage de variabilité entre les élèves, en comparaison aux scores obtenus au CE1D.

2. La décomposition de la variance.

Sur base du modèle vide, nous avons calculé la décomposition de la variance de sorte à confirmer l'utilité des régressions multiniveaux. En ce qui concerne la variance intra-classe, cette dernière équivaut à 74,48%. Par conséquent, la variance inter-classe correspond à 15,52% puisque la somme de ces deux pourcentages doit être égale à 100%. Il est intéressant de préciser que la variance inter-classe se confond avec la variance inter-école puisque chaque classe de notre échantillon appartenait à une école distincte. Quoi qu'il en soit, la majorité des différences se situent à l'intérieur des classes.

Même si la part de la variance inter-classe est plus faible que celle de la variance intra-classe, il est intéressant d'avoir recours aux régressions multiniveaux pour vérifier les différentes influences au niveau « élèves » et au niveau « classe ». Ce dernier comprend les variables relatives à l'enseignant et au contexte.

3. Lien entre les notes des enseignants et les variables étudiées.

Tout d'abord, nous vérifierons le lien entre les notes de l'enseignant et les scores obtenus au CE1D puis nous tenterons de répondre à notre sous-question à savoir, « à performances égales, quelles sont les variables qui interviennent dans les notes que donnent les enseignants à leurs élèves ? ». Pour ce faire, nous allons tenir sous contrôle la variable représentant la performance des élèves c'est-à-dire les scores qu'ils ont obtenus à l'épreuve externe. Cela nous permettra d'introduire les différentes variables tout en continuant à interpréter « à performances égales ». Nous introduirons les variables relatives aux caractéristiques de l'élève puis de l'enseignant et, enfin, du contexte.

A. Lien entre les notes des enseignants et les scores au CE1D

Présenté dans le tableau, ce premier modèle nous permet de vérifier notre 4^e hypothèse et donc de déterminer s'il y a un lien entre les notes des enseignants et les scores obtenus au CE1D.

Variable dépendante :	notes données par l'enseignant	
	<i>Modèle 1</i>	
	β	s.e.
Score CE1D	9.87***	0.57

* $p < 0.05$ ** $p < 0.001$ *** $p < 0.0001$

Tableau 4 Régression du score au CE1D sur les notes des enseignants

Le lien entre les notes de l'enseignant et les scores obtenus au CE1D est significatif. Ce résultat concorde avec celui obtenu lorsque nous avons déterminé la corrélation de rang. Il semble logique que ces 2 variables aient un lien si l'enseignant classe les élèves de la même façon que l'épreuve externe. L'enseignant n'attribue pas ses notes au hasard, elles sont en lien avec les performances de leurs élèves. Plus précisément, nous pouvons dire que pour toute augmentation d'un écart-type du score au CE1D, les notes données par les enseignants augmentent de 5,63 points sur 100.

- **Notre hypothèse 4 est confirmée au sein de notre échantillon** : il y a un lien entre la performance des élèves et les notes qui leur sont données par les enseignants.

B. Lien entre les notes des enseignants et les variables liées aux caractéristiques des élèves

Le tableau ci-dessous reprend les modèles issus des régressions multiniveaux de chaque variable liée aux caractéristiques des élèves sur les notes attribuées par les enseignants. Ces différentes variables correspondent au sexe, au niveau socio-économique, au retard scolaire des élèves et leur pratique du français à la maison. Pour rappel, les scores au CE1D est tenu sous contrôle pour nous permettre de procéder à une interprétation « à performances égales ».

Variable dépendante : notes données par l'enseignant								
Niveau élève	Modèle 2		Modèle 3		Modèle 4		Modèle 5	
	β	s.e.	β	s.e.	β	s.e.	β	s.e.
Score CE1D	10.04***	0.58	9.81***	0.61	9.84***	0.59	10.06***	0.58
Sexe élève	-0.55	1.07						
SES élève			0.45	0.57				
Retard scolaire					-4.13*	1.82		
Français à la maison							-0.84	1.66

* $p < 0.05$ ** $p < 0.001$ *** $p < 0.0001$

Tableau 5 Régressions multiniveaux entre les variables "élèves" et les notes des enseignants

- Lien entre les notes des enseignants et le sexe des élèves.

Notre 2^e modèle nous permet de vérifier le lien entre le sexe des élèves et les notes attribuées par l'enseignant et par conséquent, de confirmer ou infirmer notre hypothèse 5. Les résultats du modèle 2 indiquent que le sexe de l'élève et les notes de l'enseignant n'ont pas de lien statistiquement significatif ($p > 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = -0.55$) ne diffère pas statistiquement de zéro. Autrement dit, que l'élève soit une fille ou un garçon n'influence pas la note qu'il ou elle reçoit de son enseignant.

- **Notre hypothèse 5 est confirmée au sein de notre échantillon** : à performances égales, les notes données par l'enseignant ne diffèrent pas de manière statistiquement significative en fonction du sexe de l'élève pour la discipline des sciences.

- Lien entre les notes des enseignants et le niveau socio-économique de l'élève.

Notre 3^e modèle nous permet de confirmer le lien éventuel entre le SES des élèves et les notes attribuées par l'enseignant et par conséquent, de vérifier notre hypothèse 6. Les résultats du modèle 3 indiquent que le SES de l'élève et les notes de l'enseignant n'ont pas de lien

statistiquement significatif ($p > 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = 0.45$) ne diffère pas statistiquement de zéro. Autrement dit, que l'élève soit issu d'un milieu favorisé ou défavorisé n'influence pas la note qu'il reçoit de son enseignant.

➤ **Notre hypothèse 6 n'est pas confirmée au sein de notre échantillon :** à performances égales, les notes données par l'enseignant ne diffèrent pas de manière statistiquement significative en fonction du milieu socioculturel de l'élève pour la discipline des sciences.

• *Lien entre les notes des enseignants et le retard scolaire de l'élève.*

Notre 4^e modèle s'intéresse au lien entre le redoublement des élèves et les notes attribuées par l'enseignant, ce qui nous permettra de vérifier notre hypothèse 7. Les résultats du modèle 4 indiquent que le retard scolaire de l'élève et les notes de l'enseignant ont un lien statistiquement significatif ($p < 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = -4.13$) diffère statistiquement de zéro. Concrètement, à performances égales, un élève « en retard » obtient de moins bonnes notes (4,13 points de moins sur 100) par les enseignants que les élèves « à l'heure ».

➤ **Notre hypothèse 7 est confirmée au sein de notre échantillon :** à performances égales, les notes données par l'enseignant diffèrent de manière statistiquement significative en fonction du retard scolaire ou non de l'élève pour la discipline des sciences. Les notes sont sous-estimées pour les élèves qui ont connu le redoublement.

• *Lien entre les notes des enseignants et la pratique du français à la maison.*

Notre 5^e modèle nous permet de vérifier le lien entre la pratique du français à la maison et les notes attribuées par l'enseignant. Les résultats du modèle 5 indiquent que la langue parlée à domicile et les notes de l'enseignant n'ont pas de lien statistiquement significatif ($p > 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = -0.84$) ne diffère pas statistiquement de zéro. Autrement dit, à performances égales, que l'élève parle le français ou une autre langue chez lui n'influence pas la note qu'il reçoit de son enseignant. Puisque nous n'avons pas connaissance d'une étude visant à vérifier ce paramètre sur le jugement scolaire, nous n'avons pas émis d'hypothèse. Cette démarche a été organisée de sorte à s'inscrire dans une visée exploratoire.

C. Lien entre les notes des enseignants et les variables liées aux caractéristiques des enseignants

Le tableau ci-dessous reprend les modèles issus des régressions multiniveaux de chaque variable liée aux caractéristiques de l'enseignant sur les notes qu'il attribue à ses élèves. Ces différentes variables correspondent au sexe de l'enseignant et à son ancienneté. Pour rappel, les scores au CE1D est tenu sous contrôle pour nous permettre de procéder à des interprétations « à performances égales ».

Variable dépendante : les notes des enseignants				
Niveau classe (enseignant)	Modèle 6		Modèle 7	
	β	s.e.	β	s.e.
Score CE1D	0.81***	0.05	10.19***	0.58
Sexe enseignant	3.43	2.15		
Ancienneté			4.52*	2.06

* $p < 0.05$ ** $p < 0.001$ *** $p < 0.0001$

Tableau 6 Régressions multiniveaux entre les variables "enseignant" et les notes de l'enseignant

- Lien entre les notes des enseignants et leur sexe.

Notre 6^e modèle nous permet de vérifier le lien entre les notes attribuées par l'enseignant et leur sexe, ce qui nous permettra de confirmer ou infirmer notre hypothèse 8. Les résultats du modèle 6 indiquent que le sexe de l'enseignant et les notes qu'il attribue n'ont pas de lien statistiquement significatif ($p > 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = 3.43$) ne diffère pas statistiquement de zéro. Autrement dit, que le professeur soit une femme ou un homme n'influence pas la note qu'il ou elle attribue à ses élèves.

- **Notre hypothèse 8 est ainsi confirmée au sein de notre échantillon** : à performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction du sexe de l'enseignant pour la discipline des sciences.

- Lien entre les notes des enseignants et leur ancienneté.

Notre 7^e modèle nous permet de vérifier notre hypothèse 9 en s'intéressant au lien entre les notes attribuées par l'enseignant et l'ancienneté de celui-ci. Les résultats du modèle 7 indiquent que l'ancienneté de l'enseignant et les notes qu'il attribue ont un lien

statistiquement significatif ($p < 0.05$). En effet, le coefficient de régression de ce modèle ($\beta = 4,52$) diffère statistiquement de zéro. Autrement dit, que le professeur exerce cette fonction depuis plus ou moins de 10 ans influence les notes que les élèves obtiennent. Les enseignants plus novices (moins de 10 ans) auraient tendance à surévaluer les performances de leurs élèves comparativement à leurs collègues plus anciens.

- **Notre hypothèse 9 n'est dès lors pas confirmée au sein de notre échantillon** : à performances égales, les notes données par l'enseignant diffèrent statistiquement en fonction de l'ancienneté de l'enseignant pour la discipline des sciences.

D. Lien entre les notes des enseignants et les variables liées aux caractéristiques du contexte

Dans ce tableau, nous pouvons retrouver les modèles issus des régressions multiniveaux de chaque variable liée aux caractéristiques du contexte sur les notes que l'enseignant attribue aux élèves. Ces différentes variables correspondent à la performance moyenne (score) de la classe, à la taille de celle-ci ainsi qu'à l'indice socio-économique de l'établissement. Les scores au CE1D sont toujours tenus sous contrôle.

Variable dépendante : les notes de l'enseignant						
Niveau classe (contexte scolaire)	Modèle 8		Modèle 9		Modèle 10	
	B	s.e.	B	s.e.	β	s.e.
Score CE1D	9.64***	0.59	9.84***	0.58	9.84***	0.58
Moyenne classe	1.79	1.12				
ISE école			0.26	0.41		
Taille classe					1.64	2.29

* $p < 0.05$ ** $p < 0.001$ *** $p < 0.0001$

Tableau 7 Régressions multiniveaux entre les variables "contexte" et les notes de l'enseignant

- Lien entre les notes des enseignants et la moyenne de la classe.

Notre 8^e modèle nous permet de vérifier le lien entre les notes attribuées par l'enseignant et la moyenne de la classe et, par conséquent, de vérifier notre hypothèse 10. Les résultats du modèle 8 indiquent que la moyenne de la classe et les notes qu'attribue l'enseignant n'ont pas de lien statistiquement significatif ($p > 0.05$). En effet, le coefficient de régression de ce

modèle ($\beta=1.79$) ne diffère pas statistiquement de zéro. Autrement dit, que la classe soit considérée comme forte ou faible, cela n'a pas d'impact sur la note donnée par l'enseignant.

- **L'hypothèse 10 n'est pas confirmée au sein de notre échantillon** : à performances égales, les notes données par l'enseignant ne diffèrent pas de manière statistiquement significative en fonction du score moyen de la classe de l'élève pour la discipline des sciences.

- Lien entre les notes des enseignants et le niveau socio-économique de l'établissement.

Notre 9^{ème} modèle s'intéresse au lien éventuel entre les notes attribuées par l'enseignant et le niveau socio-économique de l'école, ce qui nous permet de vérifier notre hypothèse 11. Les résultats du modèle 9 indiquent que l'indice socio-économique de l'institution scolaire et les notes que l'enseignant attribue n'ont pas de lien statistiquement significatif ($p>0.05$). En effet, le coefficient de régression de ce modèle ($\beta=0.26$) ne diffère pas statistiquement de zéro. Autrement dit, que l'élève fréquente une école avec un indice socio-économique faible ou élevé n'a pas d'impact sur les notes que fournissent l'enseignant à ses élèves.

- **L'hypothèse 11 est confirmée au sein de notre échantillon** : à performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction de l'indice socio-économique de l'établissement dans lequel se trouve l'élève pour la discipline des sciences.

- Lien entre les notes des enseignants et la taille de la classe.

Notre 10^e et dernier modèle nous permet de vérifier le lien entre les notes attribuées par l'enseignant et la taille de la classe. Les résultats du modèle 10 indiquent que la taille de la classe et les notes que l'enseignant attribue n'ont pas de lien statistiquement significatif ($p>0.05$). En effet, le coefficient de régression de ce modèle ($\beta=1.64$) ne diffère pas statistiquement de zéro. Autrement dit, que l'élève soit dans un groupe classe soit inférieur soit supérieur ou égal à 20 enfants, cela n'a pas d'impact sur la note qu'il reçoit de son enseignant. De nouveau, puisque nous n'avons pas connaissance d'une étude visant à vérifier ce paramètre sur le jugement scolaire, nous n'avons pas émis d'hypothèse. Cette démarche s'inscrit également dans une visée exploratoire.

VI. Interprétation et discussion des résultats.

1. Concordance entre les notes des enseignants et les scores au CE1D

Tout d'abord, nous nous sommes intéressés aux classements des élèves en calculant la corrélation de rang afin de répondre à notre 1^{ère} hypothèse. Après avoir déterminé la corrélation de Spearman, nous avons pu déterminer si les élèves étaient classés de la même façon suivant les notes des enseignants et les scores de l'épreuve externe. En 1989, Hoge et Coladarci (cités par Bressoux et Pansu, 2003) avaient effectué cette démarche statistique et avaient obtenu un coefficient de 0.66. Ces auteurs avaient conclu que les enseignants jugeaient et classaient de manière « valide » leurs élèves. Puisque nous avons obtenu exactement la même valeur dans le cadre de notre recherche, nous pouvons donc tirer les mêmes conclusions. D'après nos résultats, il semble que les jugements des enseignants concordent avec les scores aux épreuves externes. Les élèves étant considérés comme forts par leur enseignant font partie des élèves les plus performants à l'évaluation externe et, inversement, les élèves considérés comme faibles se retrouvent parmi les élèves les moins performants. Par conséquent, les hypothèses 1 et 4 ont été confirmées.

En effet, l'hypothèse 4 avait pour objectif de vérifier le lien entre les notes des enseignants et les scores au CE1D. Statistiquement, nous y avons répondu en construisant notre 1^{er} modèle. Celui-ci indiquait que notre variable dépendante (les notes de l'enseignant) et notre variable indépendante (les scores au CE1D) avaient un lien statistiquement significatif. Nous pouvons donc affirmer que les enseignants ne notent pas les élèves arbitrairement. Plus les élèves sont performants, meilleures sont les notes qu'ils reçoivent et inversement. Nos analyses statistiques nous ont d'ailleurs permis de mettre en évidence que pour toute augmentation d'un écart-type du score au CE1D, les notes données par les enseignants augmentent de 5,63 points sur 100 en sciences au sein de notre échantillon.

Après avoir calculé la corrélation de rang pour l'ensemble des classes ayant participé à notre recherche, nous avons réalisé une démarche statistique supplémentaire dans le but de répondre à notre 2^e hypothèse. Afin de vérifier la concordance entre les notes et les scores pour chaque classe, nous avons calculé les différentes corrélations de Spearman par classe. Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) avaient également réalisé cette étape, ce qui leur a permis de conclure que les classements étaient davantage similaires lorsqu'ils sont observés par classe plutôt qu'au niveau de l'échantillon. Autrement dit, le classement établi par les notes de l'enseignant concorde davantage à celui obtenu sur base des scores de

l'épreuve externe au sein d'une classe. Dans le cadre de leur étude, ils avaient trouvé que la corrélation médiane entre le jugement de l'enseignant et les performances aux épreuves standardisées passait de 0,64 à 0,70 lorsque l'appartenance aux classes était prise en compte.

En ce qui concerne les résultats de notre recherche, les corrélations de Spearman varient de 0.45 à 0.88. Dans ce cas-ci, il est donc difficile de déterminer si le jugement intra-classe est plus précis que le jugement interclasse. Ceci dit, si on s'attarde sur chaque coefficient, nous pouvons constater que seulement 4 classes ont une corrélation de rang supérieure à 0.66 et 2 autres ont obtenu une valeur similaire. Pour rappel, les données n'étaient pas statistiquement significatives pour 6 classes. Si nous globalisons, nous avons obtenu un coefficient intra-classe inférieur à celui inter-classe pour 11 classes sur 17. Ces résultats ne nous permettent pas de confirmer notre hypothèse. D'après nos données, nous ne pouvons pas déclarer que le classement réalisé par les enseignants concorde davantage avec celui obtenu sur base des scores au CE1D lorsqu'on l'observe pour chaque enseignant à l'intérieur de sa classe plutôt que pour l'échantillon complet.

Enfin, en raison de la variation des coefficients (de 0.45 à 0.88) nous pouvons affirmer que certains enseignants rangent mieux leurs élèves en fonction de leur performance que d'autres. Plus la corrélation de rang est élevée, plus leur classement concorde avec celui établi sur base des scores obtenus au CE1D. Ceci dit, il est important de préciser qu'un enseignant ayant un classement similaire à celui de l'épreuve externe ne porte pas assurément un jugement plus exact. En effet, il est probable que ce dernier ait tendance à sous-évaluer ou à surestimer l'ensemble de ses élèves.

2. Dispersion des notes de l'enseignant et des scores au CE1D

En calculant les écart-types, nous avons pu interpréter les différentes dispersions et ainsi, vérifier notre 3^{ème} hypothèse. Celle-ci a été formulée suite à la recherche APER menée par Grisay en 1984 en FW-B (citée par Crahay, 2007). En raison d'une dispersion plus grande selon les notes de l'enseignant, les niveaux de compétence des élèves paraissent plus hétérogènes aux examens conçus par l'enseignant qu'aux tests standardisés (Crahay, 2007). En décomposant la variance, Crahay (2007) avait déterminé que la variation entre élèves au sein des classes passait de 82,8% à 57,9% lorsqu'on se référait aux scores du test standardisé comparativement aux notes qu'attribuait l'enseignant. Nous avons donc tenté de déterminer si les enseignants avaient également tendance à exacerber les différences entre élèves au sein de leur classe en sciences.

En analysant les différents écarts-types, nous avons pu confirmer assez facilement notre hypothèse. D'après nos résultats, les dispersions sont plus importantes pour les notes de l'enseignant que pour les scores obtenus au CE1D. En effet, l'écart-type des notes de l'enseignant toutes classes confondues équivaut à 15.65 alors que celui des scores obtenus à l'épreuve externe est égal à 12.50. En ce qui concerne chaque classe indépendamment, la tendance générale semble clairement évidente. À l'exception de 2 classes sur 23, tous les écarts-types sont supérieurs lorsque l'on s'intéresse à la dispersion des notes de l'enseignant comparativement à celle des scores obtenus au CE1D. De manière générale, les résultats de notre étude tendent à démontrer que les enseignants accentuent les variations interindividuelles à propos des notes qu'ils attribuent. Tout comme le constat établi dans le cadre de la recherche APER, les niveaux de compétence des élèves paraissent plus hétérogènes lorsqu'il revient à l'enseignant d'attribuer une note sur base de ses propres évaluations, dans le cadre du cours de sciences.

Nos résultats nous renvoient à l'effet de Posthumus. En effet, l'enseignant a tendance à concevoir ses interrogations de sorte à ce que la distribution des notes épouse une courbe de Gausse (Crahay, *et al.*, 2019). Selon Crahay (2007), le fait de classer les élèves de façon à retrouver avec une majorité d'élèves moyens et une minorité d'élèves forts et faibles induit une surestimation des écarts entre élèves. Puisque nous avons obtenu un écart-type plus élevé pour les notes que pour les scores, nous pouvons penser que les enseignants de notre échantillon adaptent leurs épreuves aux spécificités du groupe-classe.

De plus, si nous revenons sur la décomposition de la variance, nous pouvons affirmer que la majorité des différences se situe à l'intérieur des classes puisque la variance intra-classe s'élève à 74,48 % et que celle inter-classe équivaut à 25,52 %. Ce constat est en adéquation avec la théorie développée par Crahay (2007) dans le paragraphe précédent. En effet, l'enseignant a tendance à surestimer les différences entre élèves au sein de leur classe.

Si nous nous référons aux moyennes, nous avons pu constater qu'elles sont globalement plus élevées pour les notes des enseignants par rapport aux scores du CE1D. La moyenne toutes classes confondues des notes est supérieure d'approximativement 8 points sur 100. En ce qui concerne la moyenne de chaque classe, celle obtenue à l'épreuve externe n'est supérieure que pour une seule classe sur les 23 participantes. Pour toutes les autres classes, la moyenne des notes de l'enseignant est supérieure à celle des scores obtenus au CE1D. D'après ce constat, nous pourrions déclarer que l'enseignant a tendance à surévaluer ses élèves en sciences. Ceci

dit, il est important de rappeler que nous avons sélectionné la note que les élèves avaient obtenue à la 3^{ème} période dans leur bulletin. Nous pouvons nuancer nos propos en émettant l'hypothèse que les élèves n'ont pas été exclusivement évalués sur un test reprenant l'ensemble de la matière pour établir cette note. À raison d'une quantité de matière plus importante pour un seul test et de la longueur de celui-ci, il est probable que les élèves obtiennent des résultats moins élevés à l'épreuve externe plutôt qu'aux différentes interrogations administrées par l'enseignant tout au long de l'année.

3. Lien entre les variables étudiées et les notes de l'enseignant

En analysant la concordance et la dispersion des notes de l'enseignant et des scores au CE1D, nous venons de répondre à notre principale question de recherche. Dès à présent, en introduisant les différentes variables, nous allons y apporter des éléments de réponse supplémentaires en répondant à notre sous-question à savoir, « à performances égales, quelles sont les variables qui interviennent dans les notes que donnent les enseignants à leurs élèves ? ».

A. Variables liées aux caractéristiques de l'élève

Comme mentionné au 3^{ème} chapitre de la partie théorique, plusieurs études ont été menées dans le but de vérifier l'influence du genre de l'élève sur les notes des enseignants. En fonction des matières, les constats diffèrent. En ce qui concerne le français, Felouzis (1997) a observé une différence en faveur des filles alors que pour les mathématiques, il semblerait que les copies bonnes ou moyennes des garçons soient surestimées (Lafontaine et Monseur, 2009). Notre recherche s'inscrivant dans le cadre du cours de sciences, nous n'avons pas connaissance de l'existence d'une étude menée dans un contexte similaire au nôtre. Pour rappel, même si elle a été menée auprès d'étudiants à l'université, l'étude de Kacprzyk et al. (2019) n'avait pas mis en évidence une différence significative selon le genre des apprenants sauf lorsque le retrait de point par mauvaise réponse était d'application. Nous avons donc émis l'hypothèse que **le sexe de l'élève** ne biaisait pas les notes de l'enseignant.

En fonction de nos résultats, nous avons pu confirmer notre hypothèse. En effet, il n'y aurait pas de lien entre le sexe de l'élève et les jugements scolaires en sciences. Autrement dit, à performances égales, les notes attribuées par l'enseignant ne devraient pas être différentes que l'élève soit une fille ou un garçon. Nous pouvons penser que cela est lié à la matière d'enseignement. Notons qu'au cours de sciences, les compétences sont parfois plutôt relatives

au français et dans certains cas aux mathématiques en fonction des différents thèmes travaillés tout au long de l'année. Ceci dit, il est important de rappeler que notre échantillon n'est pas représentatif de la population et qu'il serait intéressant d'approfondir cette question dans un contexte similaire à celui de ce présent travail, mais avec une méthode d'échantillonnage plus valide.

Ensuite, nous avons tenté de vérifier si **le niveau socio-économique de l'élève** influençait les notes de l'enseignant à performances égales. Pour rappel, les études francophones (Duru-Bellat et Mingat, 1993, Jarlegan, 1990, cités par Bressou et Pansu, 2003; Oget, 1999, cité par Merle, 2018) et anglo-saxonnes (Hamilton *et al.*, 1990, Jussim, 1986, 1991, Ready et Wright, 2011, Rist, 1970, cités par Meissel et al., 2017) s'accordent pour dire que les élèves issus de milieux défavorisés seraient victimes d'un jugement plus sévère de la part de l'enseignant. Il nous a donc paru évident de formuler notre hypothèse dans ce sens.

En analysant nos données, il s'avère qu'à performances égales, les notes données par l'enseignant ne diffèrent pas de manière statistiquement significative en fonction du milieu socioculturel de l'élève pour la discipline des sciences, du moins au sein de notre échantillon. Effectivement, nous n'avons pas trouvé de lien significatif entre la note de l'enseignant et le contexte social de l'élève. En d'autres termes, les enseignants de notre échantillon ne discriminent pas leurs élèves en fonction de leur classe sociale. Ces résultats ne confirmant pas ce qui a été mis en évidence par les différentes études, il est important de rappeler à nouveau que notre échantillon n'est pas représentatif de la population.

Notre 7^e hypothèse avait pour but de vérifier l'impact du **retard scolaire de l'élève** sur les notes qu'il reçoit de la part de son enseignant. D'après les recherches menées sur cette caractéristique, les enseignants auraient tendance à juger plus sévèrement les élèves ayant connu le redoublement. Ce constat a été mis en évidence à plusieurs reprises notamment par Duru-Bellat et Mingat (1993), Jarlegan (1999, cités par Bressoux & Pansu, 2003) ainsi que Murat (1998) et Oget (1999, cités par Merle, 2018).

Dans ce cas-ci, notre hypothèse a été confirmée. Puisque nous avons trouvé un lien statistiquement significatif entre les notes de l'enseignant et le retard scolaire, nous avons mis en évidence qu'un élève « en retard » reçoit des notes sous-estimées de la part de leur enseignant. D'après nos résultats, les élèves ayant connu le redoublement reçoivent 4,13 points de moins sur 100 que leurs condisciples dits « à l'heure ». Pour cette hypothèse, notre conclusion va dans le même sens que les études menées par les différents chercheurs.

Si nous prenons en considération le contexte de la FW-B, notre système a massivement recouru au redoublement (Lafontaine et al., 2019), ce qui signifie qu'un certain pourcentage d'élèves seraient victimes de cette sous-notation de la part de leur enseignant. Pour rappel, quasiment 1 élève sur 2 a connu le redoublement à l'âge de 15 ans. En plus des différentes conséquences que le redoublement peut avoir sur un enfant (Crahay, 2007), celui-ci sera victime de cette injustice pour les années suivantes dans son parcours scolaire.

La dernière caractéristique liée à l'élève était relative à **la pratique du français à la maison**. Pour celle-ci, nous n'avons pas émis d'hypothèse puisque cette démarche s'inscrivait dans une visée exploratoire en raison du manque d'étude réalisée sur le sujet. Nous n'avons pas trouvé de lien statistiquement significatif entre les notes de l'enseignant et la langue parlée à domicile. Par conséquent, la fréquence à laquelle le français est pratiqué à la maison ne biaiserait pas les notes que l'enseignant attribue à ses élèves. Ceci dit, il pourrait être intéressant de mener une recherche sur ce sujet avec un échantillon représentatif de la population.

B. Variables liées aux caractéristiques de l'enseignant

Sur la base des études auxquelles Chouinard (2002, cité par Lafontaine et Monseur, 2009) a fait référence à savoir, celles menées par Brophy (1985), Duru-Bellat (1990) ainsi que celle de Parker-Price et Claxton (1996), nous avons formulé notre 8^e hypothèse de sorte à ce qu'elle corresponde à ce qui a été mis en évidence dans ces recherches. Nous voulions vérifier qu'à performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction du **sexe de l'enseignant** pour la discipline des sciences.

Nos résultats nous ont permis de confirmer notre hypothèse. Effectivement, nous n'avons pas trouvé de lien statistiquement significatif entre le sexe de l'enseignant et les notes qu'il attribue. Autrement dit, les élèves ne devraient pas recevoir une note différente selon que leur enseignant soit un garçon ou une fille. Ce constat est rassurant puisque les notes des enseignants ne sont pas biaisées en fonction de leur sexe.

À travers notre 9^e hypothèse, nous nous sommes intéressés à **l'ancienneté de l'enseignant**. Puisque Impara et Plake (1998, cités par Sudkamp et al., 2012) affirmaient qu'il n'y a pas encore d'étude qui ont permis de mettre en évidence l'effet de l'ancienneté du professeur sur la précision du jugement scolaire, nous avons supposé que l'ancienneté de l'enseignant n'avait pas d'impact sur les notes qu'ils attribuent. De plus, même si les enseignants plus anciens seraient davantage capables d'établir un jugement plus précis (Krolack-Schwerdt, *et*

al., 2009, 2012, cités par Artelt et Rausch, 2014), Artelt et Rausch (2014) avaient recommandé d'être prudent dans l'interprétation de cette étude en remettant en cause les conditions d'application.

Contre toute attente, notre hypothèse n'a pas été vérifiée. Nous avons trouvé un lien statistiquement significatif entre l'ancienneté de l'enseignant et les notes qu'il attribue à ses élèves. D'après nos résultats, il semblerait que les enseignants plus novices auraient tendance à surévaluer les performances de leurs élèves comparativement à leurs collègues plus anciens. Au sein de notre échantillon et à performances égales, un enseignant exerçant sa fonction depuis moins de 10 ans attribue, en moyenne, 4,52 points de plus sur 100, comparativement à un enseignant en fonction depuis 10 ans ou plus.

Il serait intéressant de mener plusieurs recherches d'après ce constat. Nous n'avons pas beaucoup d'informations théoriques nous permettant de justifier le biais introduit par l'ancienneté de l'enseignant. Ceci dit, nous pouvons formuler différentes hypothèses qui apporteraient des éclaircissements. En nous référant à notre revue de la littérature, nous pouvons supposer que l'ancienneté de l'enseignant soit liée à la personnalité professionnelle et au niveau d'exigence. Concrètement, en raison de l'évolution du métier et de la société, les plus anciens seraient peut-être plus « dogmatiques ». Pour rappel, les professeurs dogmatiques seraient plus rigides et conventionnels (Merle, 2018) et ils biaiseraient davantage l'information en accentuant les différences entre les élèves (Babad, *et al.*, 1982, cités par Bressoux et Pansu, 2003). De plus, il est possible que le niveau d'exigence soit plus élevé pour les anciens et qu'ils attribuent moins facilement les points que leurs collègues plus novices. Quoi qu'il en soit, il ne s'agit ici que de suppositions mais il serait intéressant d'apporter des éléments de réponse à ce constat qui peut sembler surprenant dans notre étude.

C. Variables liées aux caractéristiques du contexte

En ce qui concerne l'influence de **la moyenne de la classe** sur les notes des enseignants, notre revue de littérature indiquait qu'un élève moyen reçoit une note diminuée si le niveau de sa classe est élevé (Bressoux et Pansu, 2001) et qu'à contrario, les élèves d'une classe faible avaient tendance à être surévalués par leur enseignant (Morlaix, 2016, citée par Merle, 2018). Nous avons donc émis l'hypothèse qu'à performances égales, les notes données par l'enseignant diffèrent de manière statistiquement significative en fonction de la moyenne de la classe de l'élève pour la discipline des sciences. En nous référant aux études menées, les notes seraient sous-estimées dans les classes fortes et surestimées dans les classes faibles.

D'après nos résultats, nous n'avons pas trouvé de lien significatif entre la moyenne de la classe et les notes des enseignants. Même si ce constat ne va pas dans le sens des études menées sur le sujet, il serait toutefois positif. En effet, cela signifie qu'un même élève au sein de différentes classes serait jugé de façon assez similaire même si le niveau de chacune d'entre elles varie. Même si Crahay (2007) avait mis en évidence la surestimation de la variance intra-classe et que cette dernière soit plus élevée dans notre recherche que la variance inter-classe, les élèves ayant participé à notre étude ne reçoivent pas une note biaisée en fonction du niveau de leur classe.

Des études récentes ont permis de mettre en évidence que **le niveau socio-économique de l'établissement** dans lequel l'élève poursuit sa scolarité n'aurait pas d'effet significatif sur les notes qu'attribuent les enseignants (Meissel *et al.*, 2017). Du coup, nous avons formulé notre hypothèse de sorte à ce qu'à performances égales, les notes données par l'enseignant ne diffèrent pas statistiquement en fonction de l'indice socio-économique de l'établissement.

Dans ce cas-ci, nos résultats concordent avec la littérature. En effet, nous n'avons pas trouvé de lien significatif entre le SES de l'institution et les notes de l'enseignant. En d'autres termes, que l'élève soit inscrit dans une école avec un indice socio-économique faible ou élevé n'influence pas les notes que celui-ci reçoit de la part de son enseignant. Il s'agit d'un constat rassurant puisque si le lien entre ces 2 variables s'était avéré significatif, cela aurait pu contribuer à la logique du marché scolaire. Dans un système où les enfants sont ségrégués selon leurs caractéristiques sociales, il est important de ne pas marquer davantage les inégalités déjà présentes en FW-B.

Ceci dit, il est important d'être prudent dans notre interprétation en considérant l'échantillon. En effet, seulement 3 classes ont un ISE inférieur à 10/20 sur les 23 classes participantes. En ce qui concerne les écoles avec un milieu socio-économique élevé, à peine 2 classes ont un ISE supérieur à 15/20. Ce manque de variabilité a peut-être influencé nos résultats. Dès lors, il serait intéressant de reproduire la présente recherche sur un échantillon aléatoire et simple afin de vérifier si, avec davantage de variabilité, l'ISE de l'école peut avoir un impact statistiquement significatif.

Pour finir, tout comme pour la pratique du français à la maison, nous avons tenté de vérifier l'effet de la taille dans la classe sur le jugement scolaire dans une visée exploratoire. Le manque d'étude menée sur ce sujet ne nous a pas permis de formuler une hypothèse.

Il en résulte que nous n'avons pas trouvé de lien statistiquement significatif entre la taille de la classe et les notes des enseignants. Autrement dit, que l'élève soit intégré dans un groupe classe supérieur ou inférieur à 20 élèves n'influence pas le jugement scolaire porté par l'enseignant. Ce constat est une bonne nouvelle en termes d'équité mais ne concerne que notre échantillon. Il serait judicieux de vérifier cette hypothèse dans le cadre d'une étude où l'échantillon est représentatif de la population.

4. Biais et limites de notre recherche

Comme nous l'avons mentionné à plusieurs reprises au cours de cette discussion, l'échantillon n'est pas représentatif de la population globale. Nous ne pouvons donc pas élargir nos interprétations à l'ensemble des élèves de la FW-B. Il est important de préciser que nos observations et nos conclusions ne sont applicables que pour notre échantillon. Pour rappel, celui-ci n'est pas aléatoire et simple puisque les participants ont dû nous répondre et marquer leur accord avant d'intégrer l'étude. En plus de cela, il ne comprend que 23 enseignants et 441 élèves, ce qui semble peu pour pouvoir tirer des conclusions à l'échelle du système. Parmi les professeurs participants, une grande majorité travaille dans le réseau libre confessionnel et l'indice socio-économique de leur établissement se situait principalement entre 11 et 14/20. Idéalement, il aurait fallu une plus grande diversité au sein de notre échantillon.

Ensuite, il est important de prendre du recul par rapport aux données récoltées, que ce soient les scores au CE1D ou les notes de l'enseignant. Pour rappel, celles-ci sont représentées par la note obtenue à la 3^{ème} période. Par conséquent, la matière sur laquelle porte cette notation est probablement limitée à ce que les élèves ont appris en fin de 2^{ème} année secondaire contrairement au CE1D qui reprend l'ensemble des savoirs et savoir-faire travaillés tout au long du 1^{er} degré en sciences. La matière sur laquelle ont été établis les notes des enseignants et les scores à l'épreuve externe étant différente, cela constitue une limite de notre recherche.

En ce qui concerne les scores obtenus au CE1D, il paraît important de rappeler que ce sont sans doute les enseignants chargés du cours de sciences qui ont corrigé. Même s'ils ont reçu un guide précis reprenant les différents critères et indiquant la pondération des points, il est possible que des biais de correction soient intervenus à ce moment précis. Le score à cette épreuve représentant la performance effective des élèves, cet élément correspond également à une limite de notre recherche.

Et enfin, par rapport aux questionnaires que nous avons soumis aux enseignants (annexe 3), il est possible que des biais liés à la désirabilité sociale, à l'acquiescement ou encore à la tendance aux extrêmes soient intervenus (Lafontaine, 2017). En effet, ceux-ci devaient indiquer leur perception par rapport aux élèves de leur classe et à l'évaluation externe certificative. En ce qui concerne les questionnaires complétés par les élèves (annexe 2), l'introduction de ces biais serait moins évidente. En effet, à part pour l'estimation du nombre de livres à la maison, les questions étaient assez précises de sorte à récolter des informations sur le contexte de vie de l'élève en toute objectivité.

VII. Conclusion et perspectives.

Dans un premier temps, le but de notre recherche était de vérifier la concordance entre les notes de l'enseignant et les performances effectives des élèves. À travers la revue de littérature, nous avons vu que le jugement des enseignants pouvait être biaisé selon différentes caractéristiques relatives à l'élève, à l'enseignant et au contexte. Même si l'enseignant est considéré comme une personne légitimement habilitée à produire des jugements scolaires (Felouzis, 1997), il s'avère que ce dernier ne soit pas exclusivement basé sur les performances de l'élève.

Toutefois, d'après les résultats de notre recherche, les notes qu'attribue l'enseignant sont en lien avec les scores qu'ils obtiennent au CE1D, cela nous a permis d'affirmer que même si les jugements que porte l'enseignant peuvent être biaisés, ceux-ci ne sont pas arbitraires. De plus, selon nos données, les enseignants classent les élèves de manière similaire aux résultats obtenus à l'épreuve externe. Même si Hoge et Coladarci (1989, cités par Bressoux et Pansu, 2003) avaient également mis en évidence ce constat, il nous paraissait important de le vérifier dans notre contexte.

Toujours par rapport à la concordance entre ces deux variables, il a été montré dans le cadre de ce travail que les enseignants avaient tendance à disperser davantage le niveau de compétence des élèves. Cette observation étant en lien avec les théories avancées par Crahay (2007) sur la surestimation des différences intra-classe, nos résultats semblent cohérents. Que ce soit au sein de chaque classe ou toutes classes confondues, nous avons déterminé des écart-types plus élevés pour les notes des enseignants représentant le jugement scolaire comparativement aux scores obtenus au CE1D reflétant la performance effective de chaque élève.

Puisque la recherche APER (1984, Grisay, citée par Crahay, 2007) a été menée en FW-B et que son objectif était similaire au nôtre puisqu'elle a montré que les enseignants ne jugeaient pas toujours les élèves de la même manière alors que ces derniers avaient des performances identiques lorsqu'ils étaient évalués sur base d'une évaluation externe, il aurait été intéressant de reprendre partiellement sa méthodologie. Nous aurions pu aller plus loin en sélectionnant 3 classes similaires du point de vue des moyennes au CE1D dans le but de comparer les notes avec les scores de chaque élève, en n'oubliant pas de représenter la moyenne et l'écart-type.

Dans un second temps, en considérant les études menées sur les différents biais du jugement scolaire, nous avons introduit différentes variables susceptibles d'impacter l'exactitude des notes attribuées par l'enseignant.

En ce qui concerne les caractéristiques propres à l'élève, il apparaît que le retard scolaire est un paramètre qui influence le jugement scolaire. Concrètement, un enfant ayant déjà connu le redoublement est susceptible de recevoir une note plus basse que ses condisciples « à l'heure » à performances égales. Rappelons que les différents constats ne peuvent s'interpréter que pour les élèves de notre échantillon mais ce résultat concorde avec les différentes études menées sur le sujet (Bressoux & Pansu, 2003).

Par rapport aux autres caractéristiques relatives à l'élève, celles-ci ne semblent pas biaiser le jugement que portent les enseignants. Même si le sexe de l'élève peut avoir une influence sur les notes qu'il reçoit dans certaines disciplines (Felouzis, 1997), cela ne se vérifie pas en sciences. Il en est de même pour la pratique du français à la maison et le niveau socio-économique. Concernant ce dernier paramètre, nos résultats ne vont pas dans le même sens que les études citées dans la revue de la littérature. Il nous paraît donc intéressant de réaliser une étude similaire à la nôtre en considérant un échantillon représentatif.

Pour les variables liées aux caractéristiques de l'enseignant, il a été montré que le sexe de l'enseignant n'intervient pas sur l'exactitude du jugement qu'il porte. Ce résultat concorde avec les études menées qui avaient pour but de vérifier l'influence de ce paramètre sur les notes qu'attribue l'enseignant (Lafontaine et Monseur, 2009). Par contre, nos observations concernant l'ancienneté de l'enseignant ont été surprenantes. En effet, il semble que les enseignants plus novices ont tendance à surévaluer les élèves comparativement à leurs collègues plus chevronnés. Nous avons tenté d'expliquer ce constat en émettant des suppositions mais il serait intéressant de poursuivre ces recherches dans un contexte similaire au nôtre.

Enfin, d'après nos résultats, les caractéristiques liées au contexte n'interviennent pas dans le jugement scolaire des enseignants. Bien que cela soit cohérent pour l'indice socio-économique de l'établissement par rapports aux études menées sur le sujet (Meissel et al., 2017), nous avons trouvé un résultat contradictoire avec notre revue de littérature pour la moyenne de la classe. D'après nos analyses statistiques, ce dernier paramètre n'influence pas la note de l'enseignant alors que Crahay (2007) et plusieurs chercheurs (Bressoux et Pansu, 2001) avaient mis en évidence que les notes que recevaient un élève dépendaient des

performances du reste de la classe. De nouveau, il serait intéressant de vérifier l'influence de ce paramètre sur l'exactitude du jugement scolaire dans le contexte de la FW-B.

Pour terminer, il serait intéressant que l'enseignant ait connaissance de ces différents biais dans le souci d'établir un jugement le moins biaisé possible. Idéalement, la note que reçoit un élève devrait être exclusivement basée sur ses performances. En termes de justice scolaire, il n'est pas normal qu'un élève reçoive une note plus basse s'il a connu le redoublement et si son enseignant exerce dans la fonction depuis plus de 10 ans. Pour cette raison, il est appréciable que les élèves soient soumis à une épreuve commune et externe avec un guide de correction précis.

VIII. Bibliographie.

Artelt, C., & Rausch, T. (2014). Accuracy of Teacher Judgments. When and for what reasons? In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teachers' professional development: Assessment, training and learning* (pp. 27-43). Sense Publishers.

Auwarter, A., & Aruguete, M. (2008). Effects of Student Gender and Socioeconomic Status on Teacher Perceptions. *The Journal of Educational Research*, 101(4), 243-246. <https://doi.org/10.3200/JOER.101.4.243-246>

Bressoux, P., & Pansu, P. (2001). Effet de contexte, valeur d'internalité et jugement scolaire. *L'orientation scolaire et professionnelle*, 30, 353-371. <https://doi.org/10.4000/osp.5133>

Bressoux, P., & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Presses universitaires de France.

Chouinard, R. (2002). Chapitre 9. Différences d'attitudes et de comportement en classe selon l'appartenance sexuelle. In T. Nault, & J. Fijalkow (Eds.), *La gestion de la classe* (pp.185-198). De Boeck Supérieur. <https://doi.org/10.3917/dbu.fijal.2002.01.0185>

Circulaire n° 7972 de la communauté française du 16 février 2021 concernant les directives relatives à l'organisation des épreuves externes certificatives « CE1D » et « CESS » de l'année scolaire 2020-2021. (2021).

Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire?* (3rd ed.). De Boeck Supérieur. doi:10.3917/dbu.craha.2007.01.

Crahay, M., Marcoux, G., Lafontaine, D., Baye, A., Monseur, C., Dachet, D., Fresson, M., Dardenne, B., Boraita, F., & Mottier Lopez, L. (2019). *Peut-on lutter contre l'échec scolaire?* (4th ed.) De Boeck Supérieur.

Décret de la Communauté française du 2 juin 2006 relatif à l'évaluation externe des acquis des élèves de l'enseignement obligatoire et au certificat d'études de base au terme de l'enseignement primaire. (2006). *Moniteur belge*, 23 août, p.41954.

Dépret, E., & Filisetti, L. (2001). Juger et estimer la valeur d'autrui : des biais de jugement aux compétences sociales. *L'orientation scolaire et professionnelle*. <https://doi.org/10.4000/osp.5090>

Fédération Wallonie-Bruxelles. (2019a) *Épreuve externe certificative – CE1D*. <http://www.enseignement.be/index.php?page=26247&navi=30570>.

- Fédération Wallonie-Bruxelles. (2020). *Évaluations et outils d'évaluation*.
<http://www.enseignement.be/index.php?page=24755&navi=335>
- Felouzis, G. (1997) *L'efficacité des enseignants*. Presses Universitaires de France.
- Felouzis, G. (2015) *Les inégalités scolaires*. Presses Universitaire de France.
- Kacprzyk, J., Parsons, M., Maguire, P., & Stewart, G. (2019). Examining gender effects in different types of undergraduate science assessment. *Irish Educational Studies*, 38(4), 467-480.
<https://doi.org/10.1080/03323315.2019.1645721>
- Lafontaine, D. (2017). Évaluations à large échelle : Prendre la juste mesure des effets de contexte. In Detroz, P., Crahay, M. & Fagnant, A. (Eds.), *L'évaluation à la lumière des contextes et des disciplines* (pp. 21-51). Belgique, Louvain-La-Neuve : De Boeck
- Lafontaine, D. (2001). Le dispositif d'évaluation externe en Communauté française de Belgique. *Cahier du service de pédagogie expérimentale* (7), 195-218.
- Lafontaine, D., Bricteux, S., Hindryckx, G., Matoul, A., & Quittre, V. (2019). *Performance des jeunes de 15 ans en lecture, mathématiques et sciences. Premiers résultats de PISA 2018 en Fédération Wallonie-Bruxelles*. Service d'analyse des systèmes et des pratiques d'enseignement. Faculté de Psychologie et des Sciences de l'Éducation. Évaluation et Qualité de l'Enseignement. Université de Liège.
- Lafontaine, D., & Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe des élèves? *Mesure et évaluation en éducation*, 32 (2), 71-98.
<https://doi.org/10.7202/1024955ar>
- Marchand, F. (1996). *Évaluation des élèves et conseil de classe*. (2nd ed.). Desclée de Brouwer.
- Maroy, C., Mons, N., Resnick, J., Berényi, E., Bajomi, I., Neumann, E., Dembélé M., Goulet, S., Lapointe, P., Beniger, M.-A., Giannas, V., Tchimou, M., Mangez, C., Cattonar, B., Dumay, X., Brassard, A., Lusignan, J., Pelletier, G., Huguenin, J.-M., ... , Cuculou, S. *L'école à l'épreuve de la performance*. (1st ed.). De Boeck.
- Maroy, C., Zachary, M.-D., Vandenberghe, V., Dupriez, V., Cattonar, B., Casalfiore, S., & Ketele, J.-M. (2003). *L'enseignement secondaire et ses enseignants*. (1st ed.). De Boeck.
- Meissel, K., Meyer, F., Yao, E.S., & Rubie-Davies, C.M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48-60. <https://doi.org/10.1016/j.tate.2017.02.021>

Merle, P. (1996). *L'évaluation des élèves : enquête sur le jugement professoral*. (1st ed.) Presses Universitaires de France.

Merle, P. (2018). *Les pratiques d'évaluation scolaire*. (1st ed.). Presses Universitaires de France.

Mottier Lopez, L. (2008). Discussion notionnelle : les relations entre jugement et évaluation. *Revue suisse des sciences de l'éducation*, 30(3), 455-463. Academic Press Fribourg.

Mottier Lopez, L., & Allal, L. (2008). Le jugement professionnel en évaluation: un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30(3), 465-482. Academic Press Fribourg.

Mottier Lopez, L., Crahay, M., Allal, L., Duru-Bellat, M., Lafontaine, D., Soussi, A., Nidegger, C., Maroy, C., Meunier, M., Hurteau, M., Perréard Vité, A., Minten, P., Loeffler, A., Carette, V., Vanhulle, S., Van Nieuwenhoven, C., Pouget, M., Figari, G., & Jorro, A. (2009). *Évaluations en tension*. (1st ed.). De Boeck.

Nicaise, J., (2002). De la non-objectivité du jugement professoral en matière d'évaluation des performances des élèves. *Cahiers du service de Pédagogie expérimentale*, 11-12/2002, 133-148. Université de Liège.

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

Schépens, L. (2017). *L'effet de l'ethnicité des élèves sur le jugement scolaire des enseignants de français en première année de l'enseignement secondaire de la Fédération Wallonie-Bruxelles*. Université de Liège.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104(3), pp. 743-762. <http://dx.doi.org/10.1037/a0027627>

Tarquinia, C., & Tarquinia, P. (2001). Norme d'internalité, attractivité physique et évaluation scolaire. *L'orientation scolaire et professionnelle*. <https://doi.org/10.4000/osp.5141>

IX. Annexes.

1. Annexe 1 : Formulaire pour la retranscription des points

Établissement :

Classe :

Matière concernée :



N'oubliez pas de préciser sur combien est la cote de la dernière période

Numéro d'ordre	Note de la dernière période /	Note obtenue au CE1D / 100
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

2. Annexe 2 : Questionnaire « élèves »



QUESTIONNAIRE DESTINÉ AUX ÉLÈVES

Nous commençons par te rappeler que tu peux décider à tout moment de ne plus participer à cette recherche et que cela ne te causera pas de problème. Néanmoins, nous te remercions d'avance pour ta participation qui nous sera d'une grande aide.

*Toutes les questions qui te sont posées ci-dessous vont nous permettre de mener à bien notre recherche dans le cadre de notre mémoire à l'Université de Liège. Nous avons besoin du nom de ton école et de ton numéro d'ordre pour pouvoir traiter les résultats sans commettre d'erreur mais ceux-ci n'apparaîtront nulle part, tout sera complètement **anonyme**.*

Nous te rappelons également que la direction et les professeurs ne sont pas à l'initiative de cette recherche et que cela n'aura pas d'impact sur ta scolarité.

SECTION 1 : À PROPOS DE TOI

- Q1** École :
Classe :
Numéro d'ordre :

Pour ceux qui n'ont pas de numéro d'ordre : Nom

- Q2** Quelle est ton année de naissance ? (noircis une seule case)

2002..... <input type="checkbox"/>	2005..... <input type="checkbox"/>	2008..... <input type="checkbox"/>
2003..... <input type="checkbox"/>	2006..... <input type="checkbox"/>	2009..... <input type="checkbox"/>
2004..... <input type="checkbox"/>	2007..... <input type="checkbox"/>	2010..... <input type="checkbox"/>

- Q3** Es-tu une fille ou un garçon ? (noircis 1 seule case)

Une fille. ₁

Un garçon. ₂

- Q4** As-tu fréquenté l'enseignement maternel ? (noircis 1 seule case)

Non. ₁

Oui, pendant une année scolaire, ou moins. ₂

Oui, pendant plus d'une année scolaire. ₃

- Q5** En primaire, t'est-il arrivé de redoubler une année ? (noircis 1 seule case)

Non, jamais. ₁

Oui, une fois. ₂

Oui, deux fois ou plus. ₃

Q6 En secondaire, as-tu fréquenté une année différenciée ? (Noircis 1 seule case)

Non. ₁

Oui, la 1^{ère} année différenciée seulement. ₂

Oui, la 2^{ème} année différenciée seulement. ₃

Oui, la 1^{ère} et la 2^{ème} année différenciée. ₄

SECTION 2 : TA FAMILLE ET TON ENVIRONNEMENT À LA MAISON

Les questions de cette section portent sur ta famille et sur ton environnement à la maison.

Certaines des questions qui suivent concernent tes parents (ou bien la personne qui tient pour toi le rôle de maman ou de papa dans votre famille : par exemple, ta tutrice ou tuteur, ta belle-mère ou ton beau-père, ta mère ou ton père adoptif, etc.) Nous parlons de ton **parent 1 et parent 2**, il n'y a pas d'ordre. S'il n'y a qu'une personne, tu peux barrer la question 10 concernant le parent 2.

Si tu partages ton temps entre deux familles différentes, réponds aux questions suivantes pour les parents (ou les beaux-parents, ou les tuteurs...) avec lesquels tu passes le plus de temps.

Q7 A quelle fréquence parles-tu français à la maison ? (Noircis 1 seule case)

Je parle tout le temps français à la maison. ₁

Je parle presque tout le temps français à la maison. ₂

Je parle parfois français et parfois une autre langue à la maison. ₃

Je ne parle jamais français à la maison. ₄

Q8 Si tu N'ES PAS né(e) en Belgique, quel âge avais-tu quand tu es arrivé(e) ?

Si tu avais moins de 12 mois à ce moment, indique « zéro » (0).

Si tu es né(e) en Belgique, ignore cette question et passer à la question 9.

.....ans

Q9 Quelle est l'activité professionnelle principale de ton « parent 1 » ?

(Par exemple, enseignante, aide-cuisinière, directrice de ventes.)

(Si elle ne travaille pas actuellement, indique sa dernière activité professionnelle principale.)

Écris le nom de son travail :

Q10 Quelle est l'activité professionnelle principale de ton « parent 2 » ?

(Par exemple, enseignant, aide-cuisinier, directeur de ventes.)

(S'il ne travaille pas actuellement, indique sa dernière activité professionnelle principale.)

Écris le nom de son travail :

Q11 Environ combien de livres y a-t-il chez toi ?

Ne compte pas les revues, les journaux ni les livres d'école. (Noircis 1 seule case)

Aucun ou très peu
(entre 0 et 10 livres) 1

Ceci représente 10 livres



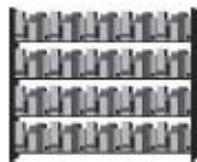
Assez pour remplir une étagère
(entre 11 et 25 livres) 2

Ceci représente 25 livres



Assez pour remplir
une bibliothèque
(entre 26 et 100 livres) 3

Ceci représente 100 livres



Assez pour remplir
deux bibliothèques
(entre 101 et 200 livres) 4

Ceci représente 200 livres



Assez pour remplir
trois bibliothèques ou plus
(plus de 200 livres) 5

Ceci représente plus de 200 livres



Q12 Y a-t-il les choses suivantes chez toi ? (Noircis 1 seule case par ligne)

- | | Oui | Non |
|---|---------------------------------------|---------------------------------------|
| a) Un ordinateur ou une tablette. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| b) Une table ou un bureau à ta disposition pour travailler. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| c) Une chambre rien qu'à toi. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| d) Une connexion à internet. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| e) Un ordinateur ou une tablette rien qu'à toi. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| f) Un GSM rien qu'à toi. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| g) Une console de jeux (par exemple, PlayStation®, Wii®, Xbox®). | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |
| h) Un système d'alarme pour ta maison. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ |

3. Annexe 3 : Questionnaire « enseignant »



Questionnaire destiné aux enseignants

Nous commençons par vous rappeler que votre direction n'est pas à la demande de cette recherche et que même si celle-ci a accepté que votre école participe, vous pouvez décider à tout moment de ne plus participer à cette recherche sans que cela ne vous porte préjudice. Néanmoins, nous vous remercions d'avance pour votre participation qui nous sera d'une aide précieuse.

Toutes les questions qui vous sont posées ci-dessous vont nous permettre de mener à bien notre recherche dans le cadre de notre mémoire à l'Université de Liège. Nous avons besoin du nom de votre établissement pour pouvoir traiter les résultats sans commettre d'erreur mais ceux-ci n'apparaîtront nulle part, tout sera complètement anonyme.

SECTION 1 : A PROPOS DE VOUS

Q1. Dans quel établissement enseignez-vous ?

.....

Q2. Êtes-vous un homme ou une femme ?

₁ Une femme

₂ Un homme

Q3. Pour quelle(s) classe(s) remplissez-vous ce questionnaire ?

.....

Q4. A quel réseau appartient votre établissement ?

₁ Enseignement organisé par la Fédération Wallonie-Bruxelles

₂ Enseignement officiel subventionné (CECP et CPEONS)

₃ Enseignement libre subventionné

Q5. Quelle matière enseignez-vous ?

₁ Mathématiques

₂ Français

₃ Sciences

Q6. Depuis combien de temps enseignez-vous ?

₁ Moins de 5 ans

₂ Entre 5 et 10 ans

₃ Entre 10 et 15 ans

₄ Plus de 15 ans

SECTION 2 : À PROPOS DE VOTRE CLASSE

Q7. Combien d'élèves avez-vous dans votre classe ?

Q8. Indiquez votre degré d'accord avec les propositions suivantes en cochant la case qui convient.

Au sein de votre classe, combien d'élèves ...	Moins de 5. <input type="checkbox"/> ₁	Entre 5 et 10. <input type="checkbox"/> ₂	Entre 10 et 15 <input type="checkbox"/> ₃	Entre 15 et 20. <input type="checkbox"/> ₄	Plus de 20. <input type="checkbox"/> ₅
... sont issus de l'immigration ?					
... considérez-vous comme faibles ?					
... considérez-vous comme attentifs ?					
... présentent des comportements perturbateurs au moins une fois par semaine ?					
... réussiront le CE1D d'après vous ?					

SECTION 3 : À PROPOS DES EVALUATIONS EXTERNES CERTIFICATIVES

Q9. Indiquez votre degré d'accord avec les propositions suivantes en cochant la case qui convient.

Selon vous, le CE1D ...	Pas du tout d'accord. <input type="checkbox"/> ₁	Plutôt pas d'accord. <input type="checkbox"/> ₂	Plutôt d'accord. <input type="checkbox"/> ₃	Tout à fait d'accord. <input type="checkbox"/> ₄
... est trop facile pour les élèves de ma classe				
... est trop difficile pour les élèves de ma classe				
... est en adéquation avec le programme.				
... est en adéquation avec le socle de compétences.				