
Les notes attribuées par les enseignants en français en Fédération Wallonie-Bruxelles concordent-elles avec les performances des élèves au CE1D ?

Auteur : Delsalle, Pauline

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 : 2019-2020

URI/URL : <http://hdl.handle.net/2268.2/9227>

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.



LIÈGE université

**Psychologie, Logopédie
& Sciences de l'Éducation**

Les notes attribuées par les enseignants en français
en Fédération Wallonie-Bruxelles concordent-elles avec
les performances des élèves au CE1D ?

Pauline DELSALLE

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

Promotrice : Dominique LAFONTAINE

Année académique 2019-2020

Remerciements

Je tiens tout d'abord à remercier Madame Lafontaine, ma promotrice, pour ses conseils, ses remises sur le droit chemin ainsi que son soutien tout au long de la réalisation de ce mémoire. Je remercie également Mesdames Virginie Dupont et Doriane Jaegers pour la réalisation des analyses statistiques ainsi que leur aide quant à l'interprétation des résultats. Je les remercie toutes trois pour leur grande disponibilité, surtout en cette période particulière.

Je remercie également d'avance Mesdames Noémie Joris et Charlotte Dejaeher d'avoir accepté d'être lectrices de ce mémoire ainsi que pour le temps qu'elles y consacreront.

Je tiens à remercier mes camarades de l'université qui ont mené leur mémoire en parallèle avec le mien, les directions ainsi que les enseignants et les élèves qui ont accepté de participer à cette recherche en me consacrant un peu de leur temps précieux en fin d'année scolaire. Ce travail n'aurait pas été réalisable sans leur implication.

Au terme de ces quatre années d'études universitaires, je tiens à remercier Sophie Delvaux et Julie Godinas, sans qui je ne me serais pas lancée dans cette aventure enrichissante. Je voudrais enfin remercier toutes les personnes, famille, amis et collègues qui m'ont permis d'arriver à la mener de concert avec une vie professionnelle et de tenir bon du début de ce parcours jusqu'à la finalisation de ce mémoire. Votre intérêt, votre soutien et vos encouragements m'ont permis de toujours m'accrocher, merci infiniment.

Table des matières

I.	INTRODUCTION	6
II.	REVUE DE LITTÉRATURE.....	8
1.	LE CONTEXTE EN FEDERATION WALLONIE-BRUXELLES	8
1.1.	<i>Les inégalités et le retard scolaire</i>	8
1.2.	<i>Le contexte d'évaluation en Fédération Wallonie-Bruxelles.....</i>	9
2.	LE JUGEMENT SCOLAIRE DES ENSEIGNANTS	11
2.1.	<i>Définition du jugement scolaire.....</i>	11
2.2.	<i>Objectivité du jugement scolaire</i>	12
2.3.	<i>Les enjeux et biais sociaux de l'évaluation scolaire</i>	13
2.4.	<i>Importance et impact du jugement des enseignants sur les élèves.....</i>	14
	a) Impact sur les apprentissages.....	14
	b) Les prophéties autoréalisatrices.....	14
	c) L'échec scolaire et les inégalités sociales	15
2.5.	<i>Exactitude du jugement scolaire par les enseignants.....</i>	17
	a) Corrélation entre les notes données par les enseignants et les résultats aux épreuves standardisées.....	17
	b) Corrélation de rang.....	19
	c) Conclusions sur l'exactitude du jugement.....	19
	d) L'effet Posthumus.....	19
2.6.	<i>Jugement des enseignants : les biais</i>	20
2.6.1.	Les caractéristiques liées aux élèves.....	21
	a) Le sexe de l'élève	21
	b) Le comportement	22
	c) Difficultés, troubles d'apprentissage, du comportement et attention.....	23
	d) La compétence des élèves	24
	e) Le retard scolaire	25
	f) Le niveau socio-économique et culturel de l'élève	26
	g) La situation familiale.....	27
	h) L'origine ethnique.....	28
	i) L'attractivité physique	29
2.6.2.	Les caractéristiques liées à l'enseignant	30
	a) Le sexe de l'enseignant.....	30
	b) L'expérience de l'enseignant	32
2.6.3.	Les caractéristiques liées au jugement	32
	a) Les informations sur l'étude	32
	b) Les échelles de mesure.....	33
2.6.4.	Les caractéristiques liées au test	34
	a) La matière.....	34

b)	Évaluation spécifique dans une matière.....	34
2.6.5.	Les caractéristiques contextuelles	35
a)	La moyenne de la classe	35
b)	L'indice socio-économique de l'établissement.....	36
c)	La moyenne de l'établissement.....	36
III.	QUESTION DE RECHERCHE ET HYPOTHESES.....	38
IV.	CONCEPTION DE LA RECHERCHE	43
1.	PRESENTATION DE LA RECHERCHE	43
2.	DESCRIPTION DE L'ÉCHANTILLON	43
3.	METHODES ET INSTRUMENTS.....	44
3.1.	<i>Présentation des questionnaires</i>	44
3.1.1.	Questions concernant les élèves.....	45
3.1.2.	Questions concernant les enseignants	45
4.	TRAITEMENT DES DONNEES	46
4.1.	<i>Présentation des variables</i>	46
4.2.	<i>Analyses statistiques</i>	48
V.	PRESENTATION DES RESULTATS.....	49
1.	CARACTERISTIQUES DE L'ÉCHANTILLON	49
1.1.	<i>Variables concernant les élèves</i>	49
1.2.	<i>Variables concernant la classe</i>	50
1.3.	<i>Analyse des fréquences</i>	52
2.	VARIANCE DES NOTES SCOLAIRES DONNEES PAR LES ENSEIGNANTS.....	52
3.	CONCORDANCE ENTRE LES NOTES DONNEES PAR LES ENSEIGNANTS ET LES RESULTATS AU CE1D	53
3.1	<i>Concordance toutes classes confondues</i>	53
3.2	<i>Concordance classe par classe</i>	53
3.3	<i>Dispersion des notes à l'intérieur des classes</i>	54
4.	LIEN ENTRE LES NOTES DES ENSEIGNANTS ET LES VARIABLES ETUDIEES.....	55
4.1.	<i>Lien entre les notes des enseignants et les résultats au CE1D</i>	55
4.2.	<i>Lien entre les notes des enseignants et les variables de niveau « élève »</i>	56
4.3.	<i>Lien entre les notes des enseignants et les variables de niveau « classe »</i>	58
VI.	INTERPRETATIONS ET DISCUSSION DES RESULTATS.....	60
1.	CONCORDANCE DES NOTES DONNEES PAR LES ENSEIGNANTS AVEC LES RESULTATS AU CE1D.....	60
2.	LES BIAIS DU JUGEMENT DE L'ENSEIGNANT.....	62
2.1.	<i>Le sexe de l'élève</i>	62
2.2.	<i>Le retard scolaire</i>	64
2.3.	<i>Le niveau socio-économique et culturel de l'élève</i>	65

2.4.	<i>La langue parlée à la maison</i>	66
2.5.	<i>Le sexe de l'enseignant</i>	67
2.6.	<i>L'expérience de l'enseignant</i>	67
2.7.	<i>La moyenne de la classe</i>	68
2.8.	<i>La taille de la classe</i>	69
2.9.	<i>L'indice socio-économique de l'établissement</i>	69
3.	BIAIS ET LIMITES DE LA RECHERCHE.....	70
VII.	CONCLUSION ET PERSPECTIVES	72
IX.	BIBLIOGRAPHIE	74
X.	TABLE DES TABLEAUX ET FIGURES	81
XI.	ANNEXES	82
1.	ANNEXE 1 : DOCUMENT DESTINE A LA RECOLTE DES NOTES.....	82
2.	ANNEXE 2 : QUESTIONNAIRE DESTINE AUX ENSEIGNANTS	83
3.	ANNEXE 3 : QUESTIONNAIRE DESTINE AUX ELEVES	85
4.	ANNEXE 4 : DONNEES BRUTES POUR LA DECOMPOSITION DE LA VARIANCE DES NOTES SCOLAIRES	89

I. Introduction

Au début des années 90, le Rapport sur l'état du système éducatif rédigé par des experts de l'Ocdé (Ocdé, 1991) met en avant deux caractéristiques inquiétantes du système éducatif en Communauté française de Belgique : l'importance des taux d'échecs et la disparité de réussite en fonction de l'établissement fréquenté. (Lafontaine, 2001, p. 195)

Dans les années 1980, en Communauté française de Belgique, une recherche a été menée par Grisay (1992) dans le but de mettre un dispositif en place afin de réduire les taux d'échec scolaire. Celle-ci a pour origine le constat selon lequel les enseignants pouvaient donner des notes différentes à des élèves qui avaient pourtant un même niveau de performance, mesuré par un test standardisé. Des élèves qui avaient donc des performances égales pouvaient soit être autorisés par leur enseignant à passer à l'année suivante soit être obligés de recommencer leur année.

Ces deux constats nous ont fait réfléchir au système scolaire en Fédération Wallonie-Bruxelles ainsi qu'à la part de responsabilité qu'ont les enseignants au niveau de la réussite des élèves. Aujourd'hui, la Fédération Wallonie-Bruxelles présente toujours des taux de retard scolaire élevés ainsi que des inégalités d'origine sociale importantes par rapport à d'autres pays (PISA, 2019). En 2017-2018, selon les Indicateurs de l'enseignement (2019), 33,9 % des élèves de deuxième secondaire ont déjà redoublé. De plus, les différences liées à l'origine sociale expliquent 60 % des différences entre les écoles. (PIRLS, 2016)

Toutefois, depuis 1999, un référentiel commun a été mis en place pour toutes les écoles de la Fédération Wallonie-Bruxelles (Décret Missions, 1997) et des évaluations externes sont obligatoires pour tous en deuxième secondaire depuis 2013 (Décret relatif à l'évaluation externe, 2006). Le dispositif mis en place par Grisay (1992) pour contrer les taux importants d'échec scolaire consistait à élaborer un recueil de questions « types » adaptées au niveau demandé qui pouvait être utilisé par les enseignants dans leurs cours et/ou leurs évaluations. Cela avait permis de réduire les taux d'échec scolaire. Il avait donc eu un impact sur la façon d'évaluer et de noter des enseignants. Les enseignants ayant maintenant accès à l'ensemble des évaluations externes depuis

qu'elles ont été mises en place, nous nous questionnons sur la façon dont ils notent leurs élèves.

Ce travail s'intéresse dès lors à la concordance entre les notes que les enseignants en français donnent à leurs élèves de deuxième secondaire à la dernière période de l'année et les performances de ceux-ci telles qu'elles sont mesurées au CE1D. De plus, nous tenterons d'observer, pour notre échantillon, quelles sont les variables selon lesquelles deux élèves qui ont des performances égales pourraient obtenir des notes différentes par leurs enseignants.

Dans ce mémoire, nous nous intéresserons tout d'abord de manière théorique à certaines caractéristiques de la Fédération Wallonie-Bruxelles telles que le taux d'échec scolaire et les inégalités sociales et nous présenterons plus spécifiquement la littérature existante concernant la concordance du jugement des enseignants avec la performance des élèves et les variables qui pourraient également être en lien avec celui-ci. Puis, nous présenterons notre question de recherche et les hypothèses posées sur base de la littérature ainsi que le détail de notre méthodologie. Ensuite, les résultats obtenus au regard de nos hypothèses seront présentés. La partie discussion et interprétations nous permettra de mettre nos résultats en lien avec la littérature ainsi que d'attirer l'attention sur les limites et les biais que peut comporter notre recherche. Enfin, nous présenterons les conclusions de ce travail ainsi que les perspectives possibles.

II. Revue de littérature

La première partie de cette revue de littérature met brièvement en lumière le contexte de ce mémoire. Nous présenterons les concepts de retard scolaire et d'inégalités en Fédération Wallonie-Bruxelles (FW-B) ainsi que l'évolution du contexte scolaire et d'évaluation en FW-B depuis les années 1990 jusqu'à aujourd'hui.

Dans la deuxième partie, nous aborderons de manière plus spécifique le thème du jugement scolaire de l'enseignant. Nous définirons ce concept et nous intéresserons à son importance, ses impacts et ses enjeux sociaux. Ensuite, nous examinerons l'exactitude du jugement des enseignants telle qu'elle est présentée dans la littérature et les variables qui peuvent avoir un lien avec celui-ci.

1. Le contexte en Fédération Wallonie-Bruxelles

1.1. Les inégalités et le retard scolaire

Tout d'abord, la FW-B de Belgique est caractérisée par d'importants taux de **retard scolaire** par rapport à d'autres pays. Les premiers résultats de PISA 2018 (2019) indiquent que 41,1 % des élèves de 15 ans sont en retard scolaire alors qu'au niveau de l'OCDE, seuls 10,6 % le sont. Contrairement à la FW-B, près de la moitié des pays de l'OCDE qui participent à PISA pratiquent la promotion automatique. Concernant plus particulièrement les élèves qui sont soumis à l'épreuve du CE1D, les Indicateurs de l'enseignement 2019 recensent, en deuxième année du secondaire, un taux de retard de 33,9 % en 2017-2018.

Ensuite, toujours selon les premiers résultats de PISA 2018 en Fédération Wallonie-Bruxelles, il apparaît que celle-ci est classée parmi les pays dans lesquels les **inégalités liées à l'origine sociale** sont les plus marquées. Concernant les **différences entre écoles**, ce rapport montre que la variance entre écoles au niveau de la performance en lecture est en diminution. En 2009, cette part de variance était de 56 % alors qu'en 2018, elle n'était plus que de 35 %. « Ceci signifie que les résultats des élèves diffèrent moins qu'auparavant d'une école à l'autre. » (p.58) Toutefois, l'enquête PIRLS (2016) qui concerne l'année dans laquelle les élèves ont reçu quatre années d'enseignement conclue que : « Si un quart de la variance des résultats à l'enquête PIRLS s'explique par des **différences entre écoles**, 60 % de ces différences sont liées à l'**origine sociale et culturelle** des élèves. » (PIRLS 2016 : rapport final, p.114)

1.2.Le contexte d'évaluation en Fédération Wallonie-Bruxelles

Dans les années 1970-80, la recherche APER menée par Grisay (1992) montre que les programmes contiennent des « contenus à voir » ainsi que des « méthodes à utiliser ». Les objectifs officiels sont assez flous. Selon Lafontaine (2001), il n'existait pas, à ce moment-là, de référentiel commun à tous les réseaux d'enseignement. Chaque réseau avait alors pour responsabilité de définir ses programmes d'enseignement qui devaient néanmoins être validés par la Communauté française. Chaque réseau pouvait également, s'il le voulait, organiser des épreuves externes, mais cela n'était pas obligatoire.

Aujourd'hui, en FW-B, les décisions concernant les matières d'enseignement, les objectifs généraux et l'importance horaire des matières se font à un niveau central. Celles-ci sont prises pour l'ensemble de la Fédération alors que les méthodes ainsi que les manuels utilisés résultent d'une décision prise par les établissements (Crahay, 2007).

Selon Lafontaine (2001), un rapport rédigé par l'Ocdé au début des années 1990 a mené la Communauté française de Belgique à mettre en place un **dispositif d'évaluation externe** et à créer un **référentiel commun**, les « socles de compétences ». Sa version définitive sera arrêtée en 1999 et sera d'application pour tous les élèves. Néanmoins, s'il existe maintenant un référentiel commun à tous les réseaux, des programmes différents sont tout de même d'application dans chacun d'eux. Ceux-ci doivent être basés sur les socles de compétences et être approuvés par le ministre mais les enseignants ne partent donc pas toujours forcément des mêmes références pour créer leurs évaluations. Cela peut entraîner des différences entre les écoles et les élèves (Fédération Wallonie-Bruxelles, 2020).

Toujours selon Lafontaine (2001), la communication des résultats des épreuves externes aux enseignants pourrait leur permettre de comparer la moyenne de leur classe avec celle de l'ensemble des élèves qui ont été évalués par le dispositif. « En entamant l'isolement du maître face à sa classe, on peut espérer “modérer” des “débordements” (dépassements de programme, surestimation de l'hétérogénéité de la classe, exigences trop ou trop peu élevées...) » (Lafontaine, 2001, pp.203-204)

À l'heure actuelle, l'évaluation des élèves en FW-B se fait de plusieurs manières et à des moments différents. D'une part, chaque établissement établit les modalités d'**évaluation interne** sur base des socles de compétences, des compétences terminales

et des profils de formation. (Fédération Wallonie-Bruxelles, 2019b) D'autre part, il existe aussi **deux types d'évaluations externes**. Selon l'article 2 du titre 1^{er} du décret relatif à l'évaluation externe des acquis des élèves de l'enseignement obligatoire (2006), une évaluation externe est « une évaluation dont la conception et la mise en œuvre sont confiées à des personnes extérieures à l'équipe éducative d'un établissement scolaire. »(p.1)

Les **épreuves externes non certificatives** n'affectent pas le parcours scolaire des élèves, mais permettent d'informer les équipes éducatives de l'Administration des acquis des élèves. Les **évaluations externes certificatives** permettent quant à elles d'obtenir un certificat comme le CEB, le CE1D et le CESS (Fédération Wallonie-Bruxelles, 2019b).

Les évaluations externes certificatives qui sont communes pour tous sont obligatoires. En ce qui concerne le CE1D, tous les élèves de 2^e commune et 2^e complémentaire de l'enseignement ordinaire ou de l'enseignement spécialisé de forme 4 sont obligés de présenter les épreuves (Fédération Wallonie-Bruxelles, 2019a). Le décret relatif à l'évaluation externe des acquis des élèves de l'enseignement obligatoire (2006) donne plusieurs précisions sur cette évaluation. Un groupe de travail est constitué afin de concevoir les épreuves, élaborer les consignes de passation et de correction et définir les critères de correction et de réussite de l'épreuve. Tous les élèves doivent passer les épreuves au même moment et la direction doit veiller au respect des consignes et des modalités de passation. Concernant la correction, ce sont les pouvoirs organisateurs (qui peuvent déléguer aux directions) qui sont responsables du respect des consignes et des modalités de correction. « Les modalités de correction sont définies de manière à garantir l'**impartialité** et la **confidentialité**. » (p. 11) De plus, « les corrections des épreuves de plusieurs établissements peuvent être regroupées dans un même centre de correction » (p.21), mais cela reste à l'initiative des pouvoirs organisateurs.

Il apparait donc que si les épreuves, les consignes ainsi que les critères de correction sont les mêmes pour tous les élèves, la correction peut être réalisée par les enseignants eux-mêmes. Le regroupement des copies n'est pas obligatoire pour l'épreuve du CE1D contrairement aux modalités de correction du CEB en fin de 6^e primaire. Pour cette épreuve, les inspecteurs doivent regrouper un certain nombre d'enseignants pour

organiser la correction. Celui-ci doit veiller à ce que les enseignants ne corrigent pas les copies de leurs élèves (Circulaire concernant les dispositions relatives à l'octroi du Certificat d'études de base, 2019).

2. Le jugement scolaire des enseignants

2.1. Définition du jugement scolaire

Mottier Lopez (2008) parle de **jugement évaluatif**. Elle considère celui-ci comme étant une **opération constitutive de l'évaluation pédagogique**, mais précise qu'ils ne se confondent pas. Mottier Lopez (2008) et Roegiers (2010) ont tous les deux donné une définition en plusieurs étapes de l'évaluation sur base desquelles nous pouvons obtenir le schéma ci-dessous. Le jugement des enseignants aurait lieu entre la phase d'interprétation et celle de prise de décision (Lecoite, 1997, cité par Mottier Lopez, 2008).



Figure 1 Schéma des étapes de l'évaluation sur base de Mottier Lopez (2008) et Roegiers (2010)

Bressoux (2018) définit le **jugement scolaire des enseignants** comme étant une **activité de construction d'une image de la réalité qui est faite pour agir** plutôt que pour la connaissance. Les enseignants s'en servent entre autres pour décider de ce qu'ils vont enseigner et comment, pour **certifier** des compétences ou encore pour **orienter** l'élève dans son cursus scolaire. Ce jugement est influencé par différentes informations contextuelles ou individuelles qui peuvent entraver son exactitude.

« Les jugements se forment toujours en relation avec un « tout contextuel » et non pas à partir d'objets ou d'événements isolés. » (Mottier Lopez, 2008, p.459) Ils sont également **liés à des situations sociales** et toujours **attachés à une intention ou une finalité** (Mottier Lopez, 2008). Toutefois, ces jugements demandent de la **transparence** et de la **rigueur** (Lafortune, 2006, citée par Allal & Lafortune, 2008).

2.2.Objectivité du jugement scolaire

Lorsque les enseignants jugent leurs élèves sur le plan scolaire, ce n'est pas en sommant différentes informations distinctes sans prêter attention à l'élève lui-même. Il s'agit pour les enseignants de juger la valeur d'autrui et ce jugement prend en compte une valeur sociale. L'évaluation n'est pas neutre, car elle est réalisée en vue de remplir des fonctions dans l'institution scolaire comme le suivi des progrès, la détection de difficultés ou la décision d'orientation. Les élèves qui sont évalués sont des personnes à part entière pour les enseignants et pas uniquement des sujets inconnus à qui il faut attribuer une note. L'enseignant connaît ses élèves, le contexte de la classe ainsi que les autres élèves qu'il doit également évaluer. Le jugement au sein de la classe apparaît surtout comme un **mécanisme comparatif** des élèves entre eux plutôt qu'un jugement individuel sorti de tout contexte (Bressoux & Pansu, 2003).

Selon Vial (2012), rechercher une note juste, une valeur vraie d'une copie est une chimère, la notation n'est pas considérée comme fiable. On pourrait définir l'évaluation comme :

Une comparaison entre un produit réel et un modèle de référence, comme un jugement comparatif dont il serait absurde de se demander s'il est « objectif » ou s'il n'est pas « objectif » puisque « des examinateurs différents n'examinent pas de la même façon les mêmes objets ; des examinateurs différents n'utilisent pas de la même façon les échelles de notes ; les jugements des examinateurs ne sont donc pas stables. (Hadji, 1992 cité par Vial, 2012 pp. 25-26)

Cardinet (1987) a abordé la question de l'objectivité dans l'évaluation. Il lui a donné quatre caractéristiques. Pour lui, l'objectivité de l'évaluation est :

1. **Nécessaire**, car les élèves comme les enseignants ont besoin de savoir exactement ce qui est évalué et ce qui a été appris.
2. **Souhaitée**, car on observe dans diverses études que les notes attribuées aux élèves tiennent souvent d'un « tirage de loterie ». La réussite ou l'échec des élèves ne sont donc pas toujours justifiés par ses compétences.

3. **Impossible**, car les objectifs, les performances attendues, les conditions d'observation, les exigences et la signification de l'évaluation sont différents pour chaque enseignant.
4. **À rejeter**, car pour y arriver, il faudrait définir clairement de manière exhaustive les objectifs des évaluations et constituer des moyens d'observations qui pourraient être comparables. Cette standardisation de l'évaluation serait possible uniquement grâce un ordinateur qui ne pourrait alors pas s'adapter aux individualités des élèves.

Gerard (2002) précise quant à lui que ce n'est pas tant la subjectivité de l'évaluation qu'il faut éviter, mais l'arbitraire. "Une évaluation arbitraire serait celle qui procéderait « n'importe comment », sans se fonder sur une démarche consciente, rigoureuse et critique." (p.2)

2.3. Les enjeux et biais sociaux de l'évaluation scolaire

D'après Crahay (2007), l'enjeu de l'évaluation pour un enseignant est important pour lui-même, car il en va de sa crédibilité autant par rapport à ses élèves qu'au regard de ses collègues.

Premièrement, l'enseignant, lorsqu'il est face à ses élèves, doit « négocier ». C'est-à-dire qu'il doit faire adhérer ses élèves à ce qu'il a planifié au niveau didactique, et les faire progresser même si ceux-ci tentent de le pousser à revoir ses exigences vers le bas. Si l'enseignant a des exigences trop basses, les élèves peuvent voir cela comme du laxisme et à l'inverse, des exigences trop importantes risquent de mener au découragement des élèves et à terme, au conflit.

Deuxièmement, l'évaluation doit également être prise en compte au niveau de la relation entre les différents membres d'une équipe éducative. Le fait de faire réussir un élève ou de le faire recommencer son année peut impacter la relation d'un enseignant avec ses collègues. En effet, si un enseignant fait échouer un élève qui est tout juste à la limite des performances demandées, cela ne lui sera pas souvent reproché par ses collègues même s'il prend le risque de retenir l'élève dans sa scolarité. À l'inverse, s'il le fait réussir et que cet élève éprouve de grandes difficultés l'année suivante, les collègues de l'enseignant qui auront cet élève en charge risquent de le lui reprocher. Cela influencera donc sa réputation, il pourra être étiqueté comme « mauvais prof » ou

« prof trop facile » alors que la représentation sociale d'un « bon prof » va de pair avec un « prof exigeant » et avec qui tous les élèves ne réussissent pas.

2.4. Importance et impact du jugement des enseignants sur les élèves

Donner un jugement adéquat est un élément central dans la profession des enseignants. Ils évaluent très souvent pour diverses raisons et cela peut **impacter la suite de la scolarité et la vie des élèves** (Artelt & Raush, 2014). Le jugement réalisé par les enseignants est très souvent utilisé pour **prendre des décisions** (Meissel et al., 2017).

a) Impact sur les apprentissages

Premièrement, selon Beswick et al. (2005), le jugement que les enseignants portent sur les élèves peut permettre **d'identifier ceux qui présentent des difficultés** à l'école notamment en lecture et ainsi prendre des décisions pour les aider.

Deuxièmement, le jugement des enseignants **influence la perception des élèves de leurs compétences cognitives**. En effet, les élèves tiennent compte des notes qu'ils obtiennent en les comparant avec celles des autres pour se juger eux-mêmes (Möller et al., 2009). Comme le suggérait Marsh en 2007 (cité par Möller et al., 2009), cette perception de soi au niveau des performances académiques a son importance, car elle a un **impact sur l'apprentissage des élèves**. Les élèves qui ont une meilleure perception d'eux-mêmes sur un plan scolaire réussissent mieux.

b) Les prophéties autoréalisatrices

Bressoux et Pansu (2003) se sont intéressés au lien qui pouvait exister entre le jugement des enseignants et les prophéties autoréalisatrices. La prophétie autoréalisatrice est définie comme une « définition fautive de la situation, provoquant un comportement nouveau qui rend vraie la conception, fautive à l'origine » (Merton, 1997, cité par Bressoux & Pansu, 2003, p.80). Certaines études présentées par Bressoux et Pansu montrent que les attentes des enseignants peuvent influencer les progrès réalisés par les élèves même si elles sont erronées au départ. Jussim et Eccles (1992) soulignent que les revues des travaux réalisés dans le milieu scolaire montrent qu'il existe bien des effets des prophéties autoréalisatrices, même si leurs tailles sont en moyenne faibles (de 0,10 à 0,30). Toutefois, Good et Thompson (cités par Bressoux et Pansu, 2003) précisent en 1998 que lorsque les attentes portent sur des groupes d'élèves

appartenant à des minorités, par exemple par rapport à leur origine ethnique ou socio-économique, l'ampleur des effets peut être accentuée. En 1997, Madon, Jussim et Eccles (cités par Bressoux et Panus, 2003) ont montré que lorsque les enseignants surestimaient les notes des élèves plus faibles, l'effet de la prophétie autoréalisatrice était assez fort. Elle permettrait alors de favoriser les élèves plutôt que de les pénaliser.

C'est ce qu'observe également Felouzis en 1997 (cité par Bressoux & Pansu, 2003). Plus les notes attribuées à un élève sont indulgentes, plus il aura tendance à évoluer positivement durant l'année, ce qui confirmera l'effet de la prophétie autoréalisatrice. Il ajoute que les élèves qui reçoivent des notes plus sévères (pour un même niveau de performance) progressent moins. C'est aussi ce que met avant une étude réalisée en 2002 par Trouilloud, Sarrazin, Martinek et Guillet (cités par Bressoux & Pansu, 2003) en éducation physique. Les élèves qui reçoivent des jugements favorables de la part de leurs enseignants se sentent plus compétents. Leur motivation est alors plus autodéterminée et, à terme, leur performance est améliorée. Le fait d'attribuer aux élèves faibles des notes plus indulgentes n'est pas considéré par Felouzis comme du laxisme. Il précise d'ailleurs que ce n'est pas la note donnée en tant que telle qui est primordiale pour expliquer l'effet des prophéties, mais plutôt l'effet produit par celle-ci sur l'élève.

c) L'échec scolaire et les inégalités sociales

Crahay (2007) attire l'attention sur le fait que le jugement de l'enseignant va influencer la réussite ou l'échec de l'élève et peut donc mener à un redoublement.

Réussite et échec (...) résultent en dernier ressort d'un jugement opéré par des acteurs scolaires et notamment par des maîtres, des jugements réglés par des critères et qui ne peuvent être totalement arbitraires, qui peuvent même être critiqués, attaqués, contestés par les intéressés, dont la légitimité est donc toujours relativement problématique. (Hutmacher, 1993, p.33)

Crahay (2007) ajoute que les taux de retard scolaire sont surtout un reflet du jugement des enseignants des compétences de leurs élèves et pas forcément le reflet des compétences en elles-mêmes.

La décision de promotion ou de redoublement ne se fonde guère sur un bilan des acquisitions réalisées durant l'année écoulée. Les enseignants

s'efforcent de pronostiquer dans quelle mesure chaque élève est apte à suivre les enseignements de l'année à venir. (p.147)

Il conclut que les enseignants expliquent leur décision concernant le redoublement pour un élève principalement par des causes non scolaires. « Au moment de décider un redoublement, les enseignants troquent leur rôle d'activateur des apprentissages pour celui de psychologue. » (p.147)

Dans la réalité scolaire, si le redoublement sanctionne bien les difficultés de certains élèves, ce sont pourtant les enseignants (et les inspecteurs) qui repèrent ces difficultés et qui en dernière analyse décident d'un éventuel redoublement. On pourrait par conséquent, dans cette perspective, considérer tout aussi bien que la fréquence des redoublements n'est pas fonction du nombre d'élèves, mais du nombre d'enseignants qui en décident. Il faut dans ce cas s'intéresser moins aux élèves et à leurs caractéristiques qu'aux décideurs, c'est-à-dire aux enseignants, à leurs caractéristiques et au champ de rapports sociaux dans lequel ils décident de la promotion-relégation. (Hutmacher, 1993, pp. 93-94)

La recherche APER (Grisay, 1992) part du constat du nombre d'échecs élevé en Fédération Wallonie-Bruxelles dans les années 1970-80. Les chercheurs ont recueilli différentes informations dans 25 classes de primaire en région liégeoise : les performances des élèves à des épreuves externes en mathématiques et en français, les notes des élèves données par les enseignants à la fin de l'année, les questions qui composaient les évaluations des enseignants ainsi que la décision de réussite ou d'échec pour chaque élève. Ces informations ont d'abord été récoltées pour analyser les différences entre les notes données par les enseignants et celles obtenues par les élèves aux épreuves externes. Il apparaît que pour des élèves ayant un même niveau de performances mesuré par les épreuves standardisées, certains obtiennent des notes supérieures à la moyenne par leur enseignant et peuvent donc passer à l'année suivante alors que d'autres sont en échec et doivent donc recommencer leur année. Cette recherche montre donc que le redoublement est en partie lié au jugement de l'enseignant. Deux élèves qui ont pourtant des compétences égales n'ont pas les mêmes chances de réussite en étant évalués par leur enseignant. Grisay se pose alors la question

de l'impact sur la motivation et sur le rendement des élèves si, ce qui n'est qu'une « erreur » d'appréciation, est ancré dans le réel.

On ne luttera pas efficacement contre l'échec scolaire tant qu'il n'y aura pas une entente claire sur ce qu'est la réussite ; et on ne peut pas laisser à chaque enseignant le soin d'en donner sa propre définition. (Grisay, 1992, p. 228)

2.5.Exactitude du jugement scolaire par les enseignants

Les études présentées dans cette section s'appuient sur deux types d'analyses. La première est une **corrélation entre les notes données par les enseignants et les résultats obtenus à l'épreuve standardisée**. Celle-ci rapporte dans quelle mesure les notes des enseignants, reçues par les élèves, sont liées à la performance de ceux-ci, mesurée par un test standardisé. La deuxième est une **corrélation de rang** qui compare la façon dont les élèves sont classés sur base de notes données par l'enseignant d'une part et suivant les résultats à une épreuve externe d'autre part.

a) Corrélation entre les notes données par les enseignants et les résultats aux épreuves standardisées

La méta-analyse de Hoge et Coladarci (1989) se base sur 16 études qui contenaient toutes une première variable représentant le jugement des enseignants sur les performances de leurs élèves ainsi qu'une deuxième qui concernait les performances des élèves mesurées par une épreuve standardisée. Les corrélations entre le jugement de l'enseignant et les résultats aux épreuves standardisées varient, selon les études, de 0.28 à 0.92 avec une corrélation médiane de 0.66.

- Ces corrélations montrent une correspondance moyenne à forte entre les deux mesures et donc un jugement plutôt correct des enseignants par rapport aux performances des élèves.
- Les auteurs observent toutefois des corrélations plus élevées lorsqu'elles sont calculées pour **chaque classe séparément**. Ces corrélations vont de 0.48 à 0.86 et la corrélation médiane est alors de 0.70. Le jugement scolaire d'un enseignant envers ses élèves paraîtrait donc plus exact au sein de sa classe qu'au sein de l'échantillon complet.

Plus tard, Südkamp et al. (2012) ont réalisé une méta-analyse qui se base sur 75 études réalisées entre 1989 et 2009. Dans celle-ci, la corrélation entre le jugement des enseignants et les notes obtenues aux tests standardisés est de 0.63.

- Cela va dans le sens des résultats qu’avaient obtenus Hoge et Coladarci, les enseignants jugent plutôt bien leurs élèves compte tenu de leur performance, mais cela n’est pas parfait et peut être amélioré.

Sanrey et al. (2018) ont quant à eux réalisé une étude qui concernait 180 élèves en CP au niveau du français et des mathématiques. Dans celle-ci, la performance des élèves était mesurée grâce à des tests standardisés et le jugement des enseignants concernait la réussite attendue des élèves à chaque item. De plus, les enseignants désignaient parmi leurs élèves ceux qui selon eux avaient un niveau scolaire faible, moyen ou élevé. Les chercheurs ont calculé la corrélation entre ces jugements et les performances des élèves. Celle-ci apparaît comme plutôt forte (0.68).

- Comme pour les études précédentes, le jugement des enseignants concorde assez bien, mais pas parfaitement avec les performances des élèves telles qu’elles sont mesurées par l’évaluation standardisée.

Jussim et Eccles (1992) ont réalisé une étude dont les analyses sont basées sur une partie des données d’une étude longitudinale concernant les mathématiques, réalisée au Michigan. Ils ont constitué deux échantillons pour des classes concernant des élèves du grade 6 (11 à 12 ans). Le premier comprend 27 enseignants et 443 élèves et le deuxième, 71 enseignants et 1288 élèves. Dans cette étude, les enseignants devaient juger leurs élèves sur base de questionnaires qui concernaient le talent, l’effort et le rendement en mathématiques de ceux-ci. Les chercheurs ont récolté les notes finales des élèves attribuées par leurs enseignants ainsi que leurs résultats à un test standardisé. Les notes récoltées étaient celles de la fin de l’année précédente ainsi que celles de la fin de l’année en cours.

- Suite à diverses analyses réalisées sur base de ces données, Jussim et Eccles concluent notamment que si les jugements des enseignants concordent assez bien avec les performances des élèves mesurées grâce à des tests standardisés, c’est surtout parce qu’ils sont exacts.

b) Corrélation de rang

Bressoux et Pansu (2003) ont réalisé une étude au niveau du CE2 qui concernait 18 classes, soit 18 enseignants et 342 élèves. Chaque enseignant devait porter un jugement sur la valeur scolaire de ses élèves suivant une échelle qui lui était donnée, allant de 0 pour un élève qu'il considérait comme faible, à 10 s'il le considérait comme fort. Ils ont également recueilli des informations contextuelles ainsi que scolaires dont notamment les résultats en français et en mathématiques à une épreuve standardisée.

- Les résultats de cette étude montrent que les enseignants classent leurs élèves dans un ordre qui se rapproche de celui déterminé par les résultats aux épreuves standardisées.

Hoge et Coladarci (1989), dans leur méta-analyse, trouvent quant à eux une corrélation de rang de 0.76. Les études qu'ils ont analysées montrent que les enseignants jugent assez bien leurs élèves lorsqu'ils doivent les classer les uns par rapport aux autres.

c) Conclusions sur l'exactitude du jugement

De manière générale, les études présentées s'accordent entre elles. Le jugement des enseignants concorde assez bien avec les résultats obtenus aux épreuves standardisées et les enseignants classent leurs élèves relativement bien dans le même ordre que celui dans lequel ils sont placés sur base des résultats des évaluations standardisées. Si Jussim et Eccles (1992) concluent que la raison principale pour laquelle les jugements des enseignants prédisent assez bien les résultats futurs des élèves est qu'ils sont assez corrects, Hoge et Coladarci observent tout de même des variations importantes des corrélations calculées pour chaque enseignant séparément. Il apparaît donc qu'en plus de la performance des élèves, d'autres variables sont à prendre en considération dans la formation du jugement de l'enseignant.

d) L'effet Posthumus

Pour expliquer ces corrélations assez élevées, Crahay (2007) s'est intéressé à l'effet Posthumus. Selon lui, cet effet influencerait les jugements des enseignants, car il les pousserait à ajuster les cotations au sein de leurs classes de façon à approcher une distribution gaussienne. En 1984, Grisay (citée par Crahay, 2007) met en évidence le mécanisme de gonflement des variations interindividuelles. L'enseignant a tendance à

étaler la dispersion des résultats des élèves de sa classe. L'écart entre la note la plus faible et la plus élevée est alors plus important au niveau des notes que donne chaque enseignant à ses élèves que celui qui peut être observé entre les notes à l'épreuve externe. C'est pourquoi le classement des élèves apparaît comme correct au vu des corrélations énoncées préalablement, mais il y aurait une surestimation des écarts entre les élèves au sein des classes qui ne se retrouverait pas dans la distribution interclasses.

Aussi équitable et démocratique qu'il s'efforce d'être dans son évaluation, le maître peut donc fort bien contribuer inconsciemment à accroître l'inégalité, simplement parce que l'outil qu'il emploie tend - par nature, pourrait-on dire - à étaler la dispersion. (Grisay, 1992, p. 61)

2.6. Jugement des enseignants : les biais

Même s'il existe une relation entre le jugement porté par les enseignants et les scores obtenus par les élèves aux épreuves standardisées, les performances des élèves semblent n'expliquer qu'une partie de la formation du jugement, car lorsque le pourcentage de variance expliquée est calculé, 50 % de celle-ci restent inexpliqués (Hoge & Coladarci, 1989). D'autres variables seraient donc à prendre en considération pour expliquer la formation du jugement des enseignants.

Sur base de leur méta-analyse ainsi que de leurs recherches théoriques, Südkamp et al. (2012), ont proposé un modèle qui comprend plusieurs variables qui pourraient, selon eux, avoir un lien avec le jugement d'un enseignant.

Nous pouvons observer sur ce modèle que si ces variables ont un lien direct avec la précision du jugement donné, elles peuvent également avoir un lien de manière indirecte avec celle-ci. Dans la suite, nous nous intéresserons aux caractéristiques présentes sur ce modèle, à savoir les caractéristiques de l'élève, de l'enseignant, du jugement ainsi que du test et nous ajouterons à celles-ci certaines caractéristiques contextuelles comme la moyenne de la classe ou l'indice socio-économique de l'établissement.

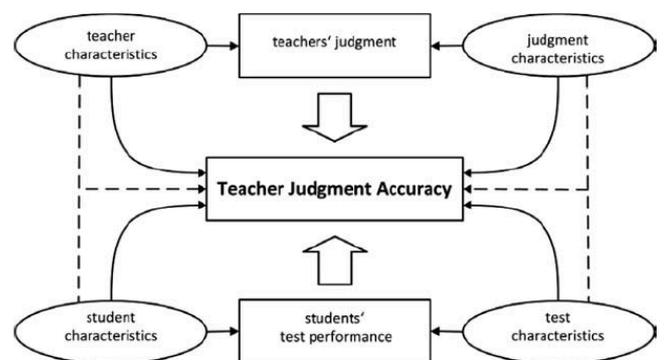


Figure 2 A model of teacher-based judgments of students' academic achievement. (Südkamp et al., 2012, p.756)

2.6.1. Les caractéristiques liées aux élèves

a) Le sexe de l'élève

Hoge et Coladarci (1989), dans leur méta-analyse, relatent les résultats obtenus concernant le lien entre le sexe de l'élève et le jugement des enseignants dans 3 des 16 études qu'ils ont analysées. Aucune de ces études ne trouve d'effet significatif du sexe de l'élève sur les notes données par les enseignants.

Sanrey et al. (2018), ont analysé le lien entre le sexe des élèves et les notes données par les enseignants en tenant la performance sous contrôle dans leur étude¹ qui concernait 180 élèves en CP au niveau du français et des mathématiques

- Ils observent que les garçons sont surévalués par rapport aux filles. À performances égales, les garçons obtiennent de meilleures notes que les filles quand on considère les domaines du français et des mathématiques en même temps.
- De plus, ils observent que les filles de faible niveau scolaire sont souvent sous-évaluées. À performances égales, pour deux élèves « faibles », les filles seront moins bien notées que les garçons.

Bressoux et Pansu² (2003) ont observé les effets du sexe de l'élève sur le jugement des enseignants, toutes choses égales par ailleurs. Dans le modèle présenté, ils tiennent notamment sous contrôle la performance des élèves, la profession des parents et le retard scolaire.

- En français, ils ne relèvent pas d'effet significatif du sexe de l'élève sur ce jugement.
- En mathématiques, ils relèvent un jugement significativement meilleur pour les garçons. À performances égales, les garçons obtiennent de meilleures notes par leurs enseignants que les filles.

Bennett et al. (1993) ont réalisé une étude sur base d'un échantillon de 797 élèves de maternelle et de début du primaire dans l'Ohio et dans le Bronx. Ils ont récolté les performances des élèves à un test standardisé qui avait été conçu pour détecter ceux qui éprouvaient des difficultés d'apprentissage. Celui-ci comprenait divers domaines

¹ Étude présentée à la page 18 (2.5-a)

² Étude présentée à la page 19 (2.5-b)

comme l'arithmétique, le langage et la cognition, la reconnaissance des lettres, la reconnaissance des mots, la lecture de manière orale ainsi que la compréhension à la lecture. Les jugements des enseignants comportaient notamment des notes données à leurs élèves en mars et en avril.

- Ils n'observent pas d'effet significatif du sexe de l'élève sur les notes données par les enseignants mis à part pour les filles au grade 1 (6-7 ans). Celles-ci recevaient de meilleures notes que les garçons, à performances pourtant égales.

Duru-Bellat et Mingat (1993) ont quant à eux réalisé une recherche dont l'échantillon était composé d'élèves de 5e (2^e secondaire). La performance de ceux-ci a été mesurée grâce à une épreuve commune de connaissances établie sur base des programmes de ce niveau et le jugement de l'enseignant était représenté par la note moyenne obtenue par ces élèves cette année-là.

- À performances égales, les filles obtiennent de meilleures notes que les garçons.

En 1997, Felouzis (cité par Bressoux, 1998) présente les résultats d'une recherche empirique qui analysait les résultats en français pour 685 élèves répartis dans 25 classes, elles-mêmes faisant partie de 5 lycées différents. Elle portait également sur les résultats en mathématiques de 927 élèves issus de 36 classes dans 5 autres lycées. Pour l'ensemble de ces élèves, il a notamment recueilli leurs résultats à des épreuves standardisées à leur entrée en 5^e et à leur sortie, les notes qu'ils ont obtenues de la part de leurs enseignants ainsi que des entretiens réalisés avec ces professeurs.

- Felouzis observe que pour les élèves de seconde (4^e secondaire) les filles obtiennent de meilleures notes en français que les garçons.

b) Le comportement

L'étude de Bennett et al. (1993) concernait 797 élèves de maternelle et de début du primaire dans l'Ohio et dans le Bronx. Les performances des étudiants étaient mesurées grâce à un test standardisé qui comportait entre autres des questions de langue, de mémoire et de mathématiques. Les jugements recueillis étaient notamment des notes données par les enseignants à leurs élèves en mars et en avril. À tous les niveaux scolaires et dans toutes les régions concernées par leur étude, il existe un lien statistiquement significatif entre la perception du comportement des élèves par les

enseignants et le jugement scolaire de ces derniers lorsqu'on tient sous contrôle des variables telles que le sexe et les performances académiques antérieures. De plus, ils observent que les garçons sont systématiquement considérés par les enseignants comme ayant de moins bons comportements que les filles.

- Les élèves qui se conduisent moins bien en classe sont jugés plus sévèrement que les autres, ils obtiennent de moins bonnes notes de la part de leurs enseignants que des élèves qui se conduisent de manière appropriée selon les enseignants alors qu'ils ont un même niveau de performance.
- Les garçons étant considérés par les enseignants comme ayant un moins bon comportement que les filles, ils sont donc également perçus comme ayant de moins bonnes performances scolaires.

La méta-analyse de Dusek et Joseph (1983) rapporte également que le comportement des élèves influence le jugement des enseignants. Dans celle-ci, ils présentent notamment l'étude de Purgess réalisée en 1979 (citée par Dusek & Joseph, 1983), dans laquelle 511 enseignants en primaire devaient lire un dossier concernant des élèves fictifs. Certains élèves étaient présentés comme perturbateurs, d'autres pas. Suite à cette lecture, les enseignants devaient évaluer la probable réussite de ces élèves, leur performance actuelle ainsi que le risque de redoublement. Leurs résultats montrent que le comportement des élèves a un effet significatif concernant toutes les prédictions des enseignants.

- Les élèves qui sont présentés comme ayant une moins bonne conduite en classe, comme perturbateurs, sont jugés comme ayant moins de chance d'avoir de bonnes performances scolaires à ce moment-là, mais aussi pour les années suivantes. Ils sont également vus comme ayant plus de chances de doubler.

c) Difficultés, troubles d'apprentissage, du comportement et attention

En 2003, Tournaki a réalisé une recherche concernant 384 enseignants qui travaillaient avec des élèves de 6 à 14 ans à New York. Chaque enseignant a reçu la description d'un élève de 9-10 ans qui vivait près de l'école avec son petit frère et un de ses parents ou les deux. Dans la suite de la description, les caractéristiques concernant le sexe de l'élève, sa performance en lecture, son niveau d'attention ainsi que son comportement, diffèrent selon les versions. La description concernant la performance en lecture précise si l'élève présente des difficultés en lecture et si celles-

ci sont diagnostiquées par un spécialiste. Les élèves ont été catégorisés selon quatre conditions, le niveau de lecture de chaque élève étant défini comme : requis, 2 ans en retard, 2 ans en retard à cause de difficultés en lecture ou 2 ans en retard à cause de troubles d'apprentissages en lecture. Après avoir lu le cas proposé, les enseignants devaient marquer leur accord ou non avec différents items. Les résultats montrent que les enseignants pronostiquent une meilleure chance de réussite pour les élèves qui ont le niveau requis en lecture que pour ceux qui ont un niveau plus bas de 2 ans, mais pour lesquels aucune difficulté ou aucun trouble n'a été précisé. De plus, les prédictions de réussite les plus faibles sont attribuées aux élèves qui ont 2 ans de retard en lecture et pour lesquelles des difficultés en lecture ou des troubles d'apprentissages ont été diagnostiqués.

- À un même niveau de performance, c'est-à-dire, un retard de deux ans en lecture, les élèves pour qui des difficultés ou des troubles d'apprentissage sont diagnostiqués ont moins de chance de réussite selon les enseignants.

Tournaki observe également que les enseignants jugent les filles de manière plus extrême si elles sont diagnostiquées comme ayant un trouble d'apprentissage. Elle suppose que cela pourrait être lié au fait que les garçons étant plus souvent diagnostiqués, si une fille l'est, elle aussi, les enseignants ont l'impression que ses difficultés sont très importantes.

- Pour deux élèves qui ont des performances égales et qui présentent un trouble d'apprentissage avec un diagnostic établi, une fille sera moins bien notée par son enseignant qu'un garçon.

d) La compétence des élèves

Dans la méta-analyse réalisée par Hoge et Coladarci (1989), il ressort que le jugement de certains enseignants peut être en lien avec le niveau de performance des élèves.

- Les élèves ayant un haut niveau de performance seraient surestimés et les moins performants sous-estimés par les enseignants. Par rapport aux résultats qu'ils obtiennent aux épreuves standardisées, qui mesurent leurs performances, un élève « fort » va recevoir de meilleures notes de la part de son enseignant et un élève « faible » en recevra de moins bonnes.

Néanmoins, Coladarci (1986) a réalisé une étude qui concernait des élèves de niveau primaire dans le Montana. L'échantillon était composé de sept enseignantes qui avaient entre 12 et 15 ans d'ancienneté et d'un seul enseignant de sexe masculin qui avait 8 ans d'ancienneté. Pour leur recherche, chacun des élèves a répondu à un questionnaire standardisé correspondant à son niveau et chaque enseignant a été interviewé une ou deux semaines après cette épreuve. Lors de l'interview, il était demandé à l'enseignant de prédire pour 6 de ses élèves s'ils avaient répondu correctement à certains items du test standardisé. Les résultats de cette étude montrent que les élèves les plus performants étaient jugés avec plus de précision.

- Il y a moins de différence entre les notes données par les enseignants et celles obtenues au test standardisé pour les élèves les plus performants que pour les plus faibles.

En 2018, dans leur étude³ qui concernait 180 élèves de CP (1^{re} primaire) au niveau de leurs compétences en français et en mathématiques, Sanrey et al. observent que les élèves moyens et forts sont surévalués.

- Les élèves les plus performants reçoivent des notes plus élevées de la part de leur enseignant que celles qu'ils obtiennent à l'épreuve standardisée.

Par contre, contrairement à Coladarci (1986), ils ont observé que les élèves plus faibles, étaient évalués plus justement. C'est-à-dire qu'il y avait moins de différences entre les notes données par l'enseignant et celles obtenues à l'épreuve standardisée pour les élèves les moins performants que pour les plus performants.

e) Le retard scolaire

Bressoux et Pansu (2003), relèvent dans leur étude⁴ que le jugement scolaire est plus sévère pour les élèves qui sont en retard scolaire, toutes choses égales par ailleurs.

- Pour deux élèves ayant notamment la même performance, la connaissance du retard scolaire d'un de ces élèves amène l'enseignant à lui donner de moins bonnes notes. En français, les élèves en retard scolaire ont en moyenne 0,8 point en moins sur 10 que les autres et en mathématiques, la différence est

³ Étude présentée à la page 18 (2.5-a)

⁴ Étude présentée à la page 19 (2.5-b)

de 0,6 point pour des élèves qui obtiennent pourtant les mêmes résultats aux épreuves standardisées.

Bressoux et Pansu (2003) ont également rapporté les résultats d'autres études réalisées par Jarlegan en 1999 et par Felouzis en 1997 qui ont observé le lien entre le retard scolaire et le jugement des enseignants. Si Jarlegan trouve, au niveau du CE1 et du CM2, un effet négatif du retard scolaire, Felouzis quant à lui, ne trouve pas d'effet significatif de cette variable au lycée.

- Pour ces échantillons, au niveau de l'école élémentaire et à performances égales, un élève en retard scolaire obtient de moins bonnes notes qu'un élève « à l'heure » alors qu'au lycée, ils sont notés de la même façon.

f) Le niveau socio-économique et culturel de l'élève

Dusek et Joseph (1983), dans leur méta-analyse, ont observé les résultats de vingt études concernant le lien entre le jugement des enseignants et le niveau socio-économique de l'élève. Pour dix d'entre elles, individuellement, aucun effet significatif n'a été observé concernant le lien entre cette variable et les attentes des enseignants concernant les élèves. Toutefois, la méta-analyse de ces études rapporte, elle, un effet significatif. Pour 64 % des élèves issus des classes moyennes, les enseignants s'attendent à ce qu'ils réussissent mieux que ceux qui sont issus d'un milieu plus défavorisé.

L'étude⁵ de Duru-Bellat et Mingat (1993) rapporte les mêmes résultats pour des élèves de 5^e (2^e secondaire). À performances égales, les élèves qui ont des parents cadres supérieurs obtiennent de meilleures notes que les élèves issus d'autres milieux sociaux.

Bressoux et Pansu⁶ (2003) observent également qu'en CE2, les élèves socio-économiquement favorisés (enfants de cadres supérieurs) sont avantagés par rapport à ceux qui sont plus défavorisés à ce niveau (surtout les enfants d'ouvriers) alors qu'ils ont des performances égales. Sur une échelle de 1 à 10, les élèves issus de familles socio-économiquement favorisées récoltent entre un demi et presque un point de plus par leurs enseignants alors qu'ils obtiennent les mêmes résultats aux évaluations externes.

⁵ Étude présentée à la page 22 (2.6.1-a)

⁶ Étude présentée à la page 19 (2.5-b)

Plus tard, dans une étude qui concernait les effets du statut socio-économique des élèves sur les perceptions des enseignants, Auwarter et Arugete (2008) se sont basées sur un échantillon de 106 enseignants du Missouri aux États-Unis. Parmi ceux-ci, 41 % enseignaient à l'école élémentaire (5-11 ans), 18 % en « middle school » (11-14 ans) et 41 % en « high school » (14-18 ans). Chaque enseignant a reçu une description fictive d'un élève qui avait des difficultés scolaires ainsi qu'au niveau de son comportement. De plus, à chaque élève étaient attribués le sexe féminin ou masculin et un statut socio-économique faible ou élevé. Le niveau socio-économique (SES) était attribué en précisant la profession des parents. Suite à cette lecture, les enseignants devaient répondre à un questionnaire qui comportait notamment des questions concernant les attentes futures des enseignants concernant ces élèves. Les résultats de leur étude montrent que les enseignants prédisent un avenir moins prometteur pour les élèves à qui un SES plus faible était attribué que pour ceux qui étaient décrits comme ayant un SES plus élevé.

Au contraire des études présentées ci-dessus, Felouzis (cité par Bressoux & Pansu, 2003) ne trouve pas d'effet significatif au niveau du lycée dans sa recherche empirique⁷ réalisée en 1997. Pour son échantillon, à ce niveau d'étude, des élèves ayant le même niveau de performance obtiennent les mêmes notes par leurs enseignants qu'ils soient issus d'un milieu socio-économique favorisé ou non.

g) La situation familiale

Levine (1981) a réalisé une recherche aux États-Unis avec un échantillon de 100 enseignants. Parmi ceux-ci 72 % enseignaient à un niveau primaire, 9 % à des élèves de 11 à 14 ans et 19 % à des élèves du deuxième cycle du secondaire. Leur étude avait pour but de déterminer si les enseignants avaient des attentes plus négatives concernant les élèves issus de familles monoparentales que pour les autres. Pour cela, un questionnaire a été créé par les chercheurs pour récolter des informations concernant les espérances des enseignants par rapport aux difficultés psychosociales, mesurées par 13 items ainsi que les performances académiques des élèves mesurées par 7 items. Chacun des enseignants devait répondre à chaque item par « beaucoup plus de chances pour les élèves issus de familles monoparentales », « un peu plus de chance pour les élèves issus de familles monoparentales », « aussi susceptibles », « un peu plus de

⁷ Étude présentée à la page 22 (2.6.1-a)

chance pour les élèves issus de familles avec les deux parents» ou « beaucoup plus de chance pour les élèves issus de familles avec les deux parents».

- Les résultats montrent que les enseignants s'attendent à ce que les enfants issus de familles monoparentales soient moins performants au niveau scolaire, mais surtout qu'ils aient plus de difficultés sur le plan psychologique et social que les enfants qui vivent avec leurs deux parents.

Santrok et Tracy (1978) se sont également intéressés aux stéréotypes que pouvaient développer les enseignants par rapport aux statuts des structures familiales de leurs élèves. Les 30 participants étaient soit des enseignants soit des futurs enseignants issus de l'université du Texas à Dallas. Il était demandé à chaque participant de regarder une vidéo de 20 minutes qui présentait un élève de 8 ans, David, qui faisait partie soit d'une famille dont les parents étaient divorcés soit d'une famille dans laquelle les parents étaient encore ensemble. Suite à cela, chaque enseignant devait noter David sur une échelle de 1 à 9 par rapport à des traits de personnalité et répondre à des questions son comportement en indiquant si les propositions étaient de « très peu probable » à « très probable ».

- Les analyses montrent que les enseignants évaluent moins favorablement les élèves dont les parents sont divorcés au niveau de leur personnalité par rapport à ceux dont les parents sont toujours ensemble.

h) L'origine ethnique

Glock et Krolak-Schwerd (2013) ont réalisé une étude en Allemagne. Dans ce pays, la plupart des élèves issus de minorités ethniques sont d'origine turque. L'objectif de cette recherche était de déterminer si, pour ces élèves, le fait d'être immigré influençait le jugement des enseignants. Quarante futurs enseignants ont dû juger des élèves fictifs pour lesquels ils avaient reçu des informations écrites. Pour une même copie, l'élève était présenté soit comme allemand soit comme turc. Ensuite, chaque futur enseignant devait évaluer ces élèves en mathématiques et en allemand.

- Les chercheurs ne trouvent pas d'effet significatif de l'origine ethnique des élèves sur le jugement des enseignants **en mathématiques**.
- En allemand et plus particulièrement **en compréhension à la lecture** ainsi qu'en **orthographe**, ils observent que les élèves turcs les plus faibles obtiennent de moins bonnes notes que les élèves allemands, à performances

égales. Par contre, les élèves « premiers de classe » obtiennent les mêmes notes, qu'ils soient turcs ou allemands.

Toujours concernant l'origine ethnique des élèves, la méta-analyse de Dusek et Joseph (1983) présente les résultats de 24 études qui se sont intéressées aux jugements des enseignants pour des élèves « noirs » et des élèves « blancs ». Si pour onze de ces études, les résultats ne sont pas significatifs, pour les quatorze autres, il ressort que le jugement et les attentes des enseignants envers leurs élèves ont un lien avec leur origine ethnique. Environ 54 % des élèves « blancs » sont jugés comme plus performants que les élèves « noirs ». Nous attirons l'attention sur le fait que ces résultats concernent uniquement la perception des enseignants, la performance n'est pas tenue sous contrôle.

i) L'attractivité physique

Dans la méta-analyse de Dusek et Joseph (1983), les résultats de 14 études réalisées entre 1973 et 1980 sont rapportés. Parmi celles-ci, 11 prenaient en compte les attentes des enseignants au niveau des performances académiques des élèves et 13 au niveau de leur développement social et de leur personnalité. Dans la plupart des études, les enseignants ont reçu des informations concernant des élèves fictifs comme par exemple leur performance à un test standardisé ainsi qu'une photo de ceux-ci. La moitié des enseignants a reçu des photos de filles et de garçons qui avaient été jugés auparavant comme « attractifs » et d'autres comme « inattractifs ». Les attentes des enseignants concernant ces élèves ont été mesurées de différentes manières comme par exemple en leur demandant un avis sur la réussite future des élèves, les relations de ceux-ci avec leurs camarades ou encore leur intérêt envers l'école.

- De manière générale, les études montrent que l'attractivité physique affecte les attentes des enseignants concernant les performances scolaires, les habiletés sociales et la personnalité des élèves.
- Concernant les études qui s'intéressaient aux performances académiques, quatre ne relevaient pas d'effet significatif tandis que les sept autres montraient un effet significatif. Dans ces dernières, les enseignants avaient des attentes supérieures pour les élèves qui étaient considérés comme « plus attractifs ».

La méta-analyse de Ritts et al. (1992) s'intéresse aux résultats de 17 études qui observent l'effet de l'attractivité physique sur des variables telles que l'intelligence, le

potentiel scolaire futur, les notes et d'autres, regroupées dans la catégorie « académique ». La seconde catégorie, quant à elle, concerne les habiletés sociales. La relation entre l'« attractivité » des élèves et le jugement des enseignants concernant les mesures académiques a été étudiée dans 12 de ces études tandis que 4 l'étudiaient par rapport aux habiletés sociales. Les résultats montrent que l'effet de l'attractivité physique sur le jugement des enseignants est faible à modéré, mais qu'il est un peu plus important pour les jugements concernant les habiletés sociales plutôt que les jugements au niveau académique.

En 1972, Willis (cité par Dusek & Joseph, 1983) observe également un lien entre les notes données par les enseignants concernant l'attractivité physique de leurs élèves et leurs attentes concernant leur réussite. Néanmoins, si cet effet est présent au début de l'année scolaire, il s'estompe au fur et à mesure et les corrélations ne sont alors plus significatives pour le reste de l'année entre l'« attractivité physique » des élèves et les attentes des enseignants. En 1974, Adams et LaVoie (cités par Dusek & Joseph, 1983) soulignent également que plus les enseignants connaissent leurs élèves, moins cette caractéristique a d'importance.

Pour expliquer cela, Leclercq et al. (2004), mentionnent l'effet de halo qui est défini comme suit par Rosenzweig (2009) : « L'effet de halo consiste à inférer d'une impression générale une caractéristique spécifique. » (p.9) Pour Leclercq et al., la note que l'enseignant donne à ses élèves serait donc, entre autres, influencée par certaines caractéristiques physiques de l'élève via cet effet de halo. Les élèves dont le physique est jugé comme plus « attractif » obtiennent des notes plus élevées que les autres par leurs enseignants même s'ils ont des performances égales.

2.6.2. Les caractéristiques liées à l'enseignant

a) Le sexe de l'enseignant

En 1996, Chatel (cité par Merle, 1998) a réalisé une recherche sur base d'un échantillon de plus de 100 enseignants en sciences économiques et sociales en année terminale. Les résultats montrent que les enseignants masculins notent leurs élèves plus sévèrement que les femmes. Autrement dit, un même élève reçoit une moins bonne note par un enseignant de sexe masculin que par une de ses collègues féminines.

Toujours en 1996, Parker et Claxton ont réalisé une recherche qui portait sur un échantillon de 38 enseignants de sexe masculin et 155 de sexe féminin au Kansas. Parmi ceux-ci, 121 enseignaient à des élèves de 6 à 11 ans, 33 à des élèves de 11 à 14 ans et 34 à des élèves de 14 à 18 ans. Chacun d'entre eux a répondu à un questionnaire qui contenait des items qui concernaient des caractéristiques attribuées au sexe de l'élève ainsi que des caractéristiques démographiques. Sur base des résultats obtenus, les auteurs concluent notamment que les enseignants considèrent que les élèves sont différents selon qu'ils soient des filles ou des garçons, mais leur étude ne permet pas de savoir si ces perceptions différentes influencent leur comportement. Il leur semble néanmoins que les enseignants sont plus critiques envers les élèves de même sexe.

Lafontaine et Monseur (2009) ont quant à eux réalisé une étude qui visait notamment à étudier si le sexe des enseignants influençait leur jugement en mathématiques. Cette étude était basée sur un échantillon de 48 enseignants du secondaire supérieur en Communauté française de Belgique dont 28 femmes et 20 hommes. Deux groupes assez comparables au niveau du sexe, de l'ancienneté et de la formation initiale ont été constitués. Les enseignants de ces deux groupes devaient alors évaluer quatre copies fictives pour des élèves qui de 5^e secondaire. Chacune de ces copies a été construite par un enseignant expérimenté et validée par deux autres enseignants. Celles-ci étaient évaluées comme « très faible », « assez faible », « assez bonne » ou « bonne » par ces enseignants. Deux copies étaient attribuées à un garçon pour le premier groupe, à une fille pour le deuxième et inversement pour les deux autres copies. Suite à leurs analyses, Lafontaine et Monseur observent que les enseignants, qu'ils soient de sexe féminin ou masculin, jugent de manière plus sévère les copies qui sont attribuées à des élèves de leur propre sexe. Néanmoins, en réalisant deux analyses de variance qui portaient sur deux copies à la fois, ils n'observent cet effet que pour une paire de copies. Pour la deuxième, la différence n'est pas significative. Ils invitent donc à être prudent quant à ces résultats.

En 1974, Good et al. (cités par Brophy, 1985) rapportent les résultats d'une étude qui impliquait 16 enseignants en « junior high school » (11-14 ans) dans deux domaines, les mathématiques et les sciences sociales. Pour chaque matière, l'échantillon était composé de 4 enseignants de sexe masculin et 4 de sexe féminin. Leurs résultats ne montrent pas que les enseignants traitent les élèves du même sexe que le leur différemment de ceux du sexe opposé.

En 2002, Chouinard indique que les recherches qui concernent les différences qu'il peut y avoir entre les enseignants selon leur sexe montrent qu'il y a peu de différences entre les hommes et les femmes au niveau de leur comportement avec les élèves si ce n'est par rapport à leur façon de communiquer et de gérer l'ordre et la discipline. Il précise également qu'il faut être prudent dans l'interprétation des résultats des études réalisées qui concernent le sexe de l'enseignant, car il y a généralement peu d'enseignants masculins dans les échantillons.

b) L'expérience de l'enseignant

En 2009, Krolak-Schwerdt et al. (cités par Artelt & Rausch, 2014) ont réalisé une étude dont l'échantillon était composé d'enseignants et d'étudiants universitaires en sciences humaines. Tous étaient amenés à lire des rapports de cas concernant des élèves fictifs. Sur base de ceux-ci, ils devaient se faire une idée des élèves décrits et prédire leur future performance (Krolak-Schwerdt et al. 2009). Les résultats montrent que l'expérience des enseignants a un effet sur leur jugement. Dans cet échantillon, les enseignants expérimentés sont plus justes. Autrement dit, à performances égales, les élèves reçoivent des notes plus proches de leur performance par des enseignants plus expérimentés (Krolak-Schwerdt et al., 2009, cités par Artelt & Rausch, 2014).

2.6.3. Les caractéristiques liées au jugement

a) Les informations sur l'étude

Südkamp et al. (2012) distinguent deux types de jugement au sein de leur méta-analyse. Le premier, qualifié de « direct », est réalisé par les enseignants en estimant par exemple le nombre d'items que leurs élèves vont réussir. Pour cela, les enseignants ont connaissance du test. Le deuxième type est « indirect » lorsque les enseignants n'ont aucune information concernant le test standardisé en fonction duquel leur jugement va être analysé. Les auteurs ont comparé ces types de jugement dans leur méta-analyse sur base de 74 mesures issues d'études antérieures. Dans celles-ci, 86,8 % des enseignants n'avaient pas connaissance du test de performance par rapport auquel leur jugement allait être comparé et 13,2 % connaissaient la mesure. Les auteurs relèvent une corrélation de 0.76 pour les jugements directs et de 0.61 pour les jugements indirects.

- Les jugements des enseignants concordent davantage avec les performances des élèves lorsque les enseignants ont connaissance de l'épreuve standardisée et jugent leurs élèves sur base de celle-ci, de manière « directe ».

b) Les échelles de mesure

Südkamp et al. (2012) observent dans les études qu'ils décrivent que le jugement des enseignants est plus juste lorsqu'il est mesuré par des tests qui comprennent plusieurs catégories plutôt que lorsqu'il est mesuré de façon globale. Les enseignants font donc plus de différences entre des élèves qui ont les mêmes performances lorsqu'ils donnent un jugement global. Machts et al. (2016) ont observé la même chose dans leur méta-analyse qui se basait sur les résultats de 33 études, en comparant les jugements donnés sur une échelle dichotomique et ceux exprimés grâce à plus de deux items. Ils observent que le jugement des enseignants concorde davantage avec les performances des élèves lorsqu'il est exprimé sur une échelle qui comprend plus de 2 catégories.

Néanmoins, si Südkamp et al. s'attendaient à observer les mêmes résultats dans leur méta-analyse sur base de 64 mesures comportant des échelles qui varient de 2 à 100 points, ils ne trouvent pas d'effet significatif du nombre de points que comprend l'échelle de mesure sur le jugement scolaire des enseignants.

- Pour deux élèves qui ont des performances égales, ils ne peuvent pas affirmer que la taille de l'échelle que l'enseignant utilise amène celui-ci à juger ces élèves différemment.

Hoge et Coladarci⁸ (1989) ont quant à eux observé les différentes manières d'exprimer un jugement. Ils ont déterminé 5 catégories :

- Les enseignants expriment un jugement sur une échelle donnée comme par exemple de faible à excellent.
- Les enseignants classent les élèves de leur classe les uns par rapport aux autres.
- Les enseignants estiment les notes que les élèves vont obtenir aux épreuves standardisées.
- Les enseignants estiment le nombre de réponses correctes que chaque élève donnera à l'épreuve standardisée.

⁸ Étude présentée à la page 17 (2.5-a)

- Les enseignants prédisent les réponses des élèves pour chaque item du test standardisé.

Des corrélations différentes sont observées dans les études en fonction de la méthodologie adoptée. Pour les études dans lesquelles les enseignants doivent estimer la performance de leurs élèves selon une échelle donnée, la corrélation médiane est de 0.61. Celle-ci est moins importante que pour toutes les autres formes de jugement. Dans l'ordre croissant des corrélations, nous retrouvons donc ensuite l'estimation du nombre de réponses correctes (0.67), la prédiction des réponses à chaque item et l'estimation de la note (0.70) et enfin le classement des élèves les uns par rapport aux autres (0.76).

En 1985, Hopkins et al. (cités par Hoge et Coladarci, 1989) expliquent que les corrélations moins élevées qui sont associées aux jugements donnés selon une échelle donnée reflètent probablement le fait que les enseignants ne veulent pas utiliser l'entièreté de ces échelles. Cela réduit donc également les variabilités entre les enseignants. Hoge et Coladarci précisent tout de même que, quelle que soit la façon dont le jugement est mesuré, il y a un lien important entre le jugement de l'enseignant et la performance des élèves.

2.6.4. Les caractéristiques liées au test

a) La matière

Dans leur méta-analyse, Südkamp et al. (2012) ont analysé des études qui concernaient des matières différentes afin d'observer si le jugement des enseignants avait un lien avec la matière évaluée. Parmi celles-ci, 63,1 % concernaient le domaine du langage et 39,9 % le domaine des mathématiques. Ils ne trouvent pas d'effet significatif de la matière sur la concordance des jugements des enseignants avec la performance des élèves.

- Les jugements réalisés par les enseignants concordent de la même façon avec les performances des élèves que cela soit dans le domaine du langage ou des mathématiques.

b) Évaluation spécifique dans une matière

Südkamp et al. (2012) font remarquer qu'il existe des différences entre les tests standardisés qui tentent de mesurer la performance des élèves. Certains sont conçus pour mesurer les performances par rapport à un seul point de matière spécifique tandis

que d'autres le sont pour mesurer différents aspects de la matière concernée ou même différentes matières. Dans leur méta-analyse, ils ont tenu compte de 103 mesures dont 12,2 % mesuraient différentes matières, 25,5 % une seule matière et 62,3 % un point spécifique dans une matière donnée. Ils n'observent pas d'effet significatif de la spécificité ou non des notions évaluées sur le jugement des enseignants.

- Les jugements réalisés par les enseignants concordent de la même façon avec les performances des élèves selon que les épreuves standardisées traitent d'un point spécifique, d'une matière ou de plusieurs matières différentes.

2.6.5. Les caractéristiques contextuelles

a) La moyenne de la classe

Bressoux et Pansu (2003) expliquent que la moyenne de la classe est liée à la fonction de l'évaluation qui consiste à discriminer les élèves. Une moyenne, ni trop élevée, ni trop basse permettrait d'éviter les effets plafond et plancher afin de mieux différencier les élèves selon leurs résultats. C'est donc pour obtenir une moyenne « socialement acceptable » et pouvoir discriminer les élèves d'une classe entre eux que les enseignants jugent plus sévèrement les élèves qui sont dans une classe dont la moyenne est plus élevée. Cela peut être en partie expliqué par l'effet Posthumus dont nous avons parlé précédemment. « Un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver d'année en année, approximativement la même distribution (gaussienne) de notes. » (Crahay, 2007, p.79) Des élèves qui ont les mêmes performances peuvent donc être « classés » différemment au sein de leur classe selon la moyenne de celle-ci. Si un élève est plus performant que la plupart de ses camarades, ses notes seront plus élevées que s'il se trouve dans une classe où la plupart des élèves ont de meilleures performances que lui.

Dans la recherche qu'ils ont réalisée, Bressoux et Pansu (2003) observent que le jugement des enseignants en mathématiques et en français au niveau CE2 est influencé par le niveau moyen de la classe. Les enseignants ne jugent pas tellement les élèves individuellement, mais les uns par rapport aux autres. Dans leur étude, plus la moyenne de la classe est élevée, plus l'enseignant est sévère en jugeant individuellement les élèves de cette classe.

Meissel et al. (2017) ont également réalisé une étude en Nouvelle-Zélande afin d'observer les liens possibles entre la moyenne de la classe et les jugements des enseignants concernant la lecture et l'écriture. L'échantillon était composé de 194 classes soit 4771 élèves pour les performances en lecture ainsi que de 561 classes soit 11 765 élèves concernant les performances en écriture. Les élèves étaient âgés de 8 à 13 ans. Les données récoltées pour chaque élève sont les résultats obtenus à une épreuve standardisée en lecture ou en écriture ainsi que le jugement global de l'enseignant concernant ces domaines à la fin de l'année pour chacun d'eux. Ces auteurs observent les mêmes résultats que Bressoux et Pansu (2003) en ce qui concerne plus précisément le jugement de tâches d'écriture.

- À performances égales, un élève qui se trouve dans une classe dont la moyenne est élevée obtient de moins bonnes notes qu'un autre se trouvant dans une classe dont la moyenne est plus faible.

Néanmoins, en 1987, Kaufmann (cité par Bressoux & Pansu, 2003) n'a quant à lui pas relevé de relation significative au collège entre le niveau de la classe et la sévérité des enseignants en français. Selon Bressoux et Pansu, cette différence pourrait être attribuée aux caractéristiques de son échantillon, les élèves et enseignants étant issus d'un seul district.

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

Meissel et al. (2017) ne trouvent pas d'effet significatif de l'indice socio-économique de l'établissement sur le jugement des enseignants dans leur étude⁹ réalisée en Nouvelle-Zélande. Au sein de leur échantillon, deux élèves qui ont des performances égales et qui se trouvent dans des établissements avec des indices socio-économiques différents obtiennent les mêmes notes de la part de leurs enseignants.

c) La moyenne de l'établissement

Duru-Bellat et Mingat¹⁰ (1993), dans une étude qui concernait des élèves de 5^e (2^e secondaire), relèvent que plus la moyenne de l'établissement mesurée grâce aux résultats des élèves au test de connaissances standardisé est élevée, plus les jugements des enseignants vis-à-vis de leurs élèves sont sévères et inversement. Meissel et al.

⁹ Étude présentée à la page 35 (2.6.5-a)

¹⁰ Étude présentée à la page 22 (2.6.1-a)

(2017)¹¹, en Nouvelle-Zélande, observent le même effet pour les élèves de 8 à 13 ans de leur échantillon au niveau de la lecture et de l'écriture.

- Dans ces deux études, il apparaît donc qu'à performances égales, deux élèves qui se trouvent dans des établissements qui ont des moyennes différentes n'obtiennent pas les mêmes notes par leurs enseignants. L'élève qui se trouve dans l'établissement dont la moyenne est la plus faible obtiendra une meilleure note.

¹¹ Étude présentée à la page 35 (2.6.5-a)

III. Question de recherche et hypothèses

Les diverses recherches consultées dans la littérature montrent que si, de manière générale, les notes des enseignants concordent assez bien avec les performances des élèves, ce n'est pas de manière parfaite.

Nous nous basons notamment sur la recherche APER (Grisay, 1992) qui a été réalisée en Fédération Wallonie-Bruxelles dans les années 1970-80 et qui 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. Ainsi, à performances égales, deux élèves pouvaient soit être jugés aptes à passer à l'année suivante ou obligés de recommencer leur année en fonction de l'enseignant qui les évaluait. Cette recherche montrait que la difficulté des évaluations données par les enseignants pouvait varier de l'un à l'autre, pour un même niveau et de mêmes objectifs à évaluer. Au moment de cette recherche, il n'existait pas d'évaluations externes ni de questions « types » sur lesquelles les enseignants pouvaient baser leur enseignement ainsi que leurs propres évaluations. Chaque enseignant était alors le seul acteur dans l'évaluation de ses élèves. La recherche APER a mis en place une évaluation externe ainsi qu'un recueil de questions « types » adaptées à la difficulté dans certaines écoles. Cela a permis de réduire le taux d'échec scolaire. Aujourd'hui, des évaluations externes certificatives sont obligatoires en FW-B à certains niveaux d'étude. Se pose dès lors la question de la concordance des notes des enseignants avec les résultats des élèves à ces évaluations après quelques années de mise en place de ce système.

Dans la littérature consultée, nous avons pu observer que les méthodologies des recherches différaient surtout dans le choix de ce qui représente le jugement d'une part et la performance des élèves de l'autre. S'il est régulièrement demandé aux enseignants de prédire la réussite ou les notes que les élèves obtiendront ultérieurement afin de les comparer avec celles qu'ils obtiennent à une épreuve donnée, les notes réellement données par l'enseignant au cours de l'année, celles qui se retrouvent dans son bulletin, sont rarement utilisées dans les recherches que nous avons pu consulter. Dès lors, dans notre recherche, les notes données par l'enseignant aux élèves lors de la dernière période de l'année représenteront le jugement et la performance sera représentée par les résultats des élèves au CE1D.

Notre principale question de recherche est donc : Les notes des enseignants en français concordent-elles avec les résultats que les élèves obtiennent au CE1D ?

Bressoux et Pansu (2003) observent que les enseignants classent relativement bien leurs élèves dans le même ordre que celui dans lequel ils sont placés selon les résultats obtenus à une évaluation standardisée. Notre première hypothèse est donc la suivante :

- Les élèves sont classés de la même façon suivant les notes données par les enseignants pendant l'année et celles obtenues au CE1D. (H₁)

Toutefois, la méta-analyse de Hoge et Coladarci (1989) a montré que le lien entre le jugement des enseignants et la performance des élèves était plus important lorsqu'on calculait la corrélation (Pearson) entre ces variables classe par classe plutôt qu'au niveau de l'ensemble de l'échantillon. Nous posons donc l'hypothèse que si ce lien est plus fort en calculant une corrélation de Pearson intraclasse, la corrélation de rang (Spearman) sera également plus importante à l'intérieur des classes que pour l'échantillon complet.

- Le classement réalisé par les enseignants concorde 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. (H₂)

De plus, les recherches de Hoge et Coladarci (1989), Südkamp, Kaiser et Möller (2012) ainsi que celles de Sanrey, Bressoux et Pansu (2018) montrent que les enseignants jugent assez bien leurs élèves par rapport à leur performance. Ils trouvent des corrélations assez élevées entre le jugement des enseignants et la performance des élèves aux épreuves externes. Nous émettons donc l'hypothèse qu'il y a un lien entre ces deux variables.

- Il y a un lien entre la performance des élèves et les notes qui leur sont données par les enseignants. (H₃)

Pour expliquer ces corrélations élevées, Crahay (2007) évoquait l'effet Posthumus. En 1984, Grisay (citée par Crahay, 2007) mettait en évidence que l'écart entre la note la plus faible et la plus élevée était plus important pour les notes données par l'enseignant que pour les notes de l'épreuve externe lorsqu'on le calculait pour chaque classe séparément.

- 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. (H₄)

Au vu de la littérature et des recherches qui ont déjà été menées, nous observons que le jugement des enseignants est assez exact au regard des performances des élèves. Néanmoins, la part de variance expliquée par la performance (environ 50 %) permet de dire qu'il y a d'autres variables qui sont en lien avec les notes que les enseignants attribuent à leurs élèves.

Nous posons donc une sous-question : À performances égales, quelles sont les variables qui sont en lien avec les notes que donnent les enseignants à leurs élèves ?

En 1997, Felouzis (cité par Bressoux et Pansu, 2003) a observé un lien entre le jugement des enseignants et le sexe des élèves. Selon son étude, à performances égales, une fille obtient de meilleures notes qu'un garçon en français. Pour notre recherche qui concerne le français, nous nous attendons donc aux mêmes résultats.

- **À performances égales**, il y a un lien significatif entre les notes données par l'enseignant et le **sexe de l'élève**. (H₅)

⇒ À performances égales, les filles obtiennent de meilleures notes que garçons et inversement.

Concernant le retard scolaire de l'élève, la recherche de Bressoux et Pansu (2003) ainsi que celles qu'ils citent (Jarlegan, 1999 ; Duru-Bellat & Mingat, 1993 ; Felouzis, 1997) s'accordent à dire que le retard scolaire a un lien avec le jugement des enseignants en défaveur des élèves qui ont dû recommencer une ou plusieurs années. Nous nous attendons donc à observer le même effet pour notre échantillon.

- **À performances égales**, il y a un lien significatif entre les notes données par l'enseignant et le **retard scolaire de l'élève**. (H₆)

⇒ À performances égales, les élèves qui sont en retard scolaire obtiennent de moins bonnes notes que les élèves « à l'heure » et inversement.

Concernant le niveau socio-économique et culturel de l'élève, Duru-Bellat et Mingat en 1993 et Jarlegan en 1999 (cités par Bressoux et Pansu, 2003) ainsi que Bressoux et Pansu (2003) concluent qu'à performances égales, les élèves issus de milieux socio-économiques favorisés obtiennent de meilleures notes par leurs enseignants que ceux qui sont issus de milieux défavorisés. Nous émettons donc l'hypothèse suivante :

- **À performances égales**, il y a un lien significatif entre les notes données par l'enseignant et le **niveau socio-économique et culturel de l'élève**. (H₇)
 ⇒ À performances égales, les élèves dont le milieu socio-économique et culturel est favorisé obtiennent de meilleures notes que les élèves défavorisés et inversement.

Les auteurs qui se sont intéressés au sexe de l'enseignant tirent comme conclusions que cette variable n'a soit pas de lien avec les notes données par les enseignants ou qu'il faut être prudent quant aux résultats obtenus même s'ils sont significatifs (Chouinard, 2002 ; Lafontaine & Monseur, 2009 ; Good et al., 1974, cités par Brophy, 1995). Notre hypothèse est donc la suivante :

- **À performances égales**, il n'y a pas de lien significatif entre les notes données par l'enseignant et le **sexe de l'enseignant**. (H₈)

Concernant l'expérience des enseignants, Krolak et al. (cités par Artelt & Rausch, 2014) ont observé que le jugement des enseignants pouvait varier selon leur expérience. Selon eux, un enseignant plus expérimenté serait plus juste dans ses notes par rapport à la performance des élèves.

- **À performances égales**, il y a un lien significatif entre notes données par l'enseignant et l'**expérience de l'enseignant**. (H₉)

Bressoux et Pansu (2003) et Meissel et al. (2017) observent un effet significatif de la moyenne de la classe sur le jugement des enseignants en faveur des élèves qui se trouvent dans une classe dont la moyenne est plus faible par rapport à ceux qui se trouvent dans une classe dont la moyenne est élevée. Nous émettons donc l'hypothèse suivante :

- **À performances égales**, il y a un lien significatif entre les notes données par l'enseignant et la **moyenne de la classe de l'élève**. (H₁₀)
 ⇒ À performances égales, les élèves qui se trouvent dans une classe pour laquelle la moyenne est élevée obtiennent de moins bonnes notes que les élèves qui se trouvent dans une classe dont la moyenne est plus faible et inversement.

Concernant l'indice socio-économique de l'établissement, une seule étude a été consultée, celle de Meissel et al. (2017). Dans celle-ci, il n'y a pas d'effet significatif

de cette variable sur le jugement de l'enseignant. Nous nous attendons donc aux mêmes résultats pour notre échantillon.

- À **performances égales**, il n'y a pas de lien significatif entre les notes données par l'enseignant et **l'indice socio-économique de l'établissement** dans lequel se trouve l'élève. (H₁₁)

IV. Conception de la recherche

1. Présentation de la recherche

Notre recherche consiste à observer si les notes données par les enseignants en français lors de la dernière période de l'année concordent avec celles que les élèves de deuxième secondaire obtiennent au CE1D. Pour comparer ces deux variables, nous avons récolté ces deux notes pour chaque élève auprès de plusieurs enseignants de français en 2^e secondaire. De plus, afin de répondre à notre sous-question de recherche, à savoir, « À performances égales, quelles sont les variables qui sont en lien avec les notes que donnent les enseignants à leurs élèves ? », nous avons eu recours à deux questionnaires, l'un destiné aux enseignants, l'autre aux élèves. Ces questionnaires ont pour but de récolter des informations concernant les caractéristiques des individus de notre échantillon, afin de pouvoir les analyser et ainsi observer si elles ont un lien avec les notes que donnent les enseignants à leurs élèves en tenant sous contrôle la performance de ceux-ci.

2. Description de l'échantillon

L'échantillon de base de notre recherche est constitué de 22 enseignants en français en 2^e année commune du secondaire travaillant tous dans des écoles différentes ainsi que de 446 élèves. Néanmoins, en raison de certaines données manquantes, nous avons travaillé sur un échantillon de 21 enseignants et de 434 élèves issus de 22 classes. Nous attirons l'attention sur le fait que notre échantillon n'est pas aléatoire et simple. En effet, les enseignants qui le composent sont ceux qui ont accepté de participer. Nous ne pourrions donc pas élargir nos résultats et interprétations à l'ensemble de la population. Ceux-ci seront uniquement valables pour les enseignants qui constituent notre échantillon.

Les 21 enseignants ayant participé sont engagés dans des écoles de la Fédération Wallonie-Bruxelles, mais appartiennent à des réseaux différents. La plupart des enseignants qui composent notre échantillon travaillent dans l'enseignement libre subventionné.

Les autres caractéristiques des enseignants ainsi que celles des élèves, faisant partie des variables étudiées, seront présentées au début de la section consacrée aux résultats.

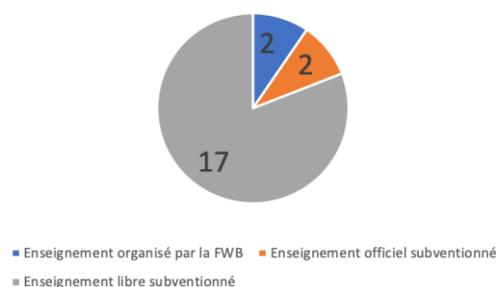


Figure 3 Répartition des enseignants selon le réseau

3. Méthodes et instruments

Pour analyser le jugement des enseignants en français, nous avons décidé de mener une recherche quantitative afin de comparer les notes qu'ils attribuent aux élèves à la dernière période de l'année avec les notes obtenues par ceux-ci à l'épreuve externe certificative dans la même matière, en fin de deuxième année du secondaire, le CE1D. Les données ont été récoltées durant les mois de mai et juin 2019.

Nous pensons qu'il y a quelques avantages à prendre comme étalon la note obtenue par les élèves au CE1D. En effet, les questions ainsi que les consignes de passation et de correction sont les mêmes pour tous les élèves. Cela supprime déjà une partie des biais d'évaluation liés au test en lui-même et aux critères de correction. Néanmoins, nous ferons tout de même preuve de vigilance, car il est fort possible que les copies des élèves de l'épreuve du CE1D aient été corrigées par les enseignants eux-mêmes. De plus, aucune épreuve ne représente la performance « vraie » d'un élève, il y a toujours des erreurs de mesure et d'interprétation.

Le choix d'utiliser les notes de la dernière période attribuées par les enseignants est justifié par le fait que ce sont les épreuves et les notes attribuées par le professeur qui sont les plus proches dans le temps de l'évaluation externe. Cela peut être contesté par le fait que ces épreuves ne reprennent pas l'ensemble de la matière comme pour l'épreuve externe, mais nous pouvons justifier cela grâce à la stéréotypie des notes, c'est-à-dire une tendance de l'évaluateur à attribuer aux élèves, des notes comparables à celles qu'ils ont reçues auparavant (Crahay, 2007). Nous estimons que les notes attribuées à la dernière période de l'année sont proches de celles qui auraient été attribuées lors d'une autre épreuve, reprenant l'ensemble de la matière. Nous pensons également que les enseignants, à ce moment de l'année, ne voient pas exclusivement de la nouvelle matière et que l'évaluation par compétence permet d'éviter que les évaluations soient faites uniquement sur un point de matière ciblé.

3.1. Présentation des questionnaires

Pour mener à bien notre recherche, nous avons récolté les notes données par les enseignants lors de la dernière période de l'année ainsi que celles obtenues au CE1D via un document¹² remis aux enseignants eux-mêmes. De plus, les enseignants¹³ et les

¹² Document présenté en annexe 1

¹³ Document présenté en annexe 2

élèves¹⁴ ont répondu à un questionnaire différent pour chacun d'eux afin de nous fournir les informations nécessaires.

3.1.1. Questions concernant les élèves

Les questions concernant les élèves sont basées sur les questionnaires contextuels des études PIRLS 2016 et PISA 2009. Le choix et l'adaptation des questions ont été réalisés en fonction de nos hypothèses de recherche ainsi qu'en tenant compte de l'âge des élèves. Les questions de ce questionnaire concernent :

- l'année de naissance de l'élève,
- le sexe de l'élève,
- la fréquentation ou non de l'enseignement maternel,
- le redoublement en primaire ou en secondaire en vue de déterminer si l'élève est en retard au niveau scolaire,
- la langue parlée à la maison,
- l'âge d'arrivée en Belgique, le cas échéant,
- l'activité professionnelle des parents, le nombre de livres au domicile ainsi que les biens qu'ils possèdent, en vue de déterminer le niveau socio-économique de l'élève.

3.1.2. Questions concernant les enseignants

Les questions concernant les enseignants ne se basent sur aucun questionnaire existant. Elles ont été créées en fonction de notre questionnement et nos hypothèses de recherche. Ce questionnaire ayant été créé pour 3 mémoires menés en parallèle concernant des matières différentes, un item « matière » est présent dans le questionnaire. Il est évident que celui-ci n'est pas utile pour notre recherche, nous n'avons tenu compte que des questionnaires remplis par les enseignants en français. Les questions de ce questionnaire concernent :

- le réseau d'enseignement,
- l'ancienneté,
- la taille de la classe,
- certaines perceptions des enseignants par rapport aux élèves de leur classe, à savoir le nombre d'élèves issus de l'immigration, le nombre d'élèves considérés

¹⁴ Document présenté en annexe 3

comme faibles, attentifs, qui présentent des comportements perturbateurs ainsi que leur estimation du nombre de réussites,

- la perception des enseignants du degré de difficulté du CE1D et de son adéquation avec le programme et les socles de compétences.

4. Traitement des données

4.1. Présentation des variables

Les données récoltées ont tout d'abord été codées suivant le codage présent sur les questionnaires. Ensuite, certaines d'entre elles ont été recodées afin de pouvoir effectuer, plus facilement, un traitement statistique. Nous présentons ici les différentes variables recodées ou non sur lesquelles nous nous sommes basée pour nos analyses statistiques. Toutes les données recueillies ont été codées de manière à les rendre anonymes.

Variables « élève » :

- 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 par les enseignants.
- La variable concernant le **retard scolaire** a été créée sur base de la question concernant le redoublement de l'élève en primaire. La plupart des élèves déclarant ne pas avoir fréquenté l'enseignement différencié, nous n'avons pas tenu compte de cette question pour établir cette variable. Celle-ci nous permet d'observer si, à performances égales les élèves obtiennent des notes différentes par les enseignants selon qu'ils soient « à l'heure » (0) ou « en retard » (1).
- La variable concernant l'utilisation du **français à la maison** a été créée sur base de la question concernant la langue parlée à la maison. Cette variable nous permet d'observer si, à performances égales, les notes données par l'enseignant aux élèves déclarant parler français à la maison tout le temps ou presque (1), diffèrent de celles données à ceux qui ne le parlent que parfois ou jamais (2).
- Les variables concernant la profession des parents, le nombre de livres chez l'élève ainsi que la possession d'un ordinateur personnel et d'un système d'alarme à la maison ont été recodées et sommées pour créer une nouvelle variable qui représente le **niveau socio-économique et culturel de l'élève**. Cette variable nous permet d'observer si, à performances égales, les notes

données par l'enseignant diffèrent en fonction du niveau socio-économique et culturel de l'élève.

Variables « classe » :

- La variable concernant **l'ancienneté de l'enseignant** a été créée sur base de la question prévue à cet effet. Nous faisons la différence entre les enseignants qui exercent depuis moins de 10 ans (1) et les autres (2). Cette variable nous permet d'observer si, à performances égales, les élèves reçoivent des notes différentes en fonction de l'ancienneté de l'enseignant.
- Le nombre d'élèves dans chaque classe a permis de créer la variable qui concerne **la taille de la classe**. Nous faisons la différence entre les classes qui comportent moins de 20 élèves (1) et celles dans lesquelles il y a 20 élèves ou plus (2). Cette variable nous permet d'observer si, à performances égales, les élèves reçoivent des notes différentes en fonction de la taille de leur classe.
- La variable concernant **la moyenne de la classe** a été créée en calculant la moyenne de la classe au CE1D. Cette variable nous permet d'observer si, à performances égales, les élèves reçoivent des notes différentes en fonction de la moyenne de leur classe.
- La variable concernant **le sexe de l'enseignant** nous permet d'observer si, à performances égales, les élèves reçoivent des notes différentes selon que l'enseignant soit une femme (1) ou un homme (2).
- La variable concernant **le niveau socio-économique de l'établissement** nous permet d'observer si, à performances égales, les élèves obtiennent des notes différentes de la part de leur enseignant selon que cet indice est plus faible ou plus élevé.
- La variable concernant **la perception de l'adéquation du CE1D avec les socles de compétences** permet d'observer si, à performances égales, les enseignants donnent des notes différentes à leurs élèves selon qu'ils soient pas du tout d'accord ou plutôt pas d'accord (1) ou plutôt ou totalement d'accord (2).
- Les variables concernant le **nombre d'élèves perçus comme faibles, attentifs** et le **nombre estimé de réussite** dans chaque classe n'ont pas pu être recodées de manière à pouvoir effectuer des analyses statistiques. En effet, les données récoltées ne permettant pas de calculer un pourcentage, les analyses n'auraient pas pu permettre de tirer des conclusions.

4.2. Analyses statistiques

Grâce à ces données, nous pourrions effectuer plusieurs analyses statistiques. Nous vérifierons tout d'abord la concordance des notes des enseignants avec les résultats que les élèves obtiennent au CE1D. Ensuite, nous nous attarderons sur les variables qui peuvent avoir un lien avec le jugement de l'enseignant.

Premièrement, nous calculerons une **corrélation de rang** (coefficient r de Pearson) entre les notes des élèves attribuées par les enseignants à la dernière période de l'année et celles qu'ils ont obtenues au CE1D. Cette analyse permettra de vérifier si les enseignants classent leurs élèves de la même façon que le classement qui ressort des résultats au CE1D. Nous avons réalisé cette corrélation tout d'abord pour l'ensemble des données et ensuite pour chacune des classes séparément.

Ensuite, nous créerons un **modèle vide** afin d'observer la part de variance à l'intérieur et entre les classes au niveau des notes données par les enseignants. Cela permettra de confirmer l'utilité de réaliser des analyses multiniveaux si une part de la variance se situe effectivement entre les différentes classes.

Dans un troisième temps, nous analyserons les écarts-types des notes des enseignants et des notes obtenues par les élèves au CE1D afin d'observer si, au sein de chaque classe, la dispersion des notes est plus importante pour celles qui sont données par les enseignants que pour celles obtenues par les élèves au CE1D.

Enfin, nous réaliserons une **analyse de régression multiniveaux** pour chaque variable étudiée. De plus, un modèle de régression multiniveaux sera également réalisé en introduisant en même temps les variables significatives afin d'observer si elles le sont toujours tenues sous contrôle l'une de l'autre. Toutes ces analyses seront réalisées en tenant sous contrôle le score obtenu au CE1D considéré comme la performance de l'élève. Les analyses seront donc toutes réalisées « **à performances égales** ». Pour chaque variable, nous observerons si le coefficient obtenu est significatif. Cela permettra d'observer si ces variables ont un lien avec les notes de l'enseignant. Dans cette partie, certaines analyses seront réalisées afin de valider les hypothèses émises sur base de la littérature. D'autres par contre, seront réalisées de manière exploratoire. Ces analyses multiniveaux seront présentées tout d'abord pour les variables de niveau élève puis de niveau classe. Les effets seront considérés comme significatifs au seuil de .05.

V. Présentation des résultats

1. Caractéristiques de l'échantillon

L'analyse des fréquences obtenues sert à décrire les caractéristiques de notre échantillon, mais également à déterminer les variables que nous pouvons étudier. Seules celles qui présentent une certaine variabilité au niveau des réponses obtenues sont utiles. En effet, si les variables étudiées ne présentent que peu de variabilité, celles-ci ne pourront pas expliquer notre variable dépendante, à savoir les notes attribuées par l'enseignant.

1.1. Variables concernant les élèves

- Répartition des élèves selon le sexe

Selon les Indicateurs de l'enseignement (2019), en Fédération Wallonie-Bruxelles (FW-B) dans le premier degré de l'enseignement secondaire, il y a 51,4 % de garçons dont 46,6 % au degré commun et 48,6 % de filles dont 45,2 % au degré commun.

Dans notre échantillon, les garçons sont un peu surreprésentés (56,1 %) et les filles un peu sous-représentées (39 %).

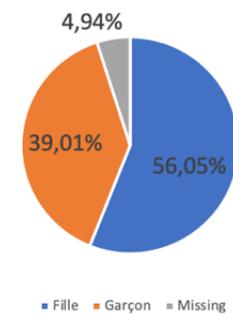


Figure 4 Répartition des élèves en fonction du sexe

- Répartition des élèves en fonction du retard scolaire

Selon les Indicateurs de l'enseignement (2019), en 2017-2018, le taux de retard scolaire en FW-B était de 33,9 % en deuxième année du secondaire.

Dans notre échantillon, les élèves en retard scolaire sont sous-représentés (10,1 %).

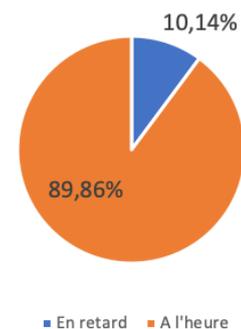
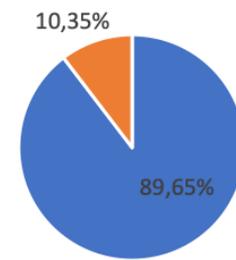


Figure 5 Répartition des élèves en fonction du retard scolaire

- Répartition des élèves en fonction de l'utilisation du français à la maison

Selon les premiers résultats de PISA 2018 en FW-B (2019), pour les élèves de 15 ans, le français est la langue le plus souvent parlée à la maison pour 82,6 % d'entre eux. Même si notre échantillon est composé d'élèves plus jeunes, nous retrouvons un pourcentage assez proche d'élèves qui parlent français à la maison (89,7 %).

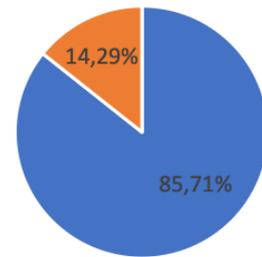


■ Tout le temps ou presque ■ Parfois ou jamais
Figure 6 Répartition des élèves en fonction de l'utilisation du français à la maison

1.2. Variables concernant la classe

- Répartition des enseignants selon le sexe

Selon les Indicateurs de l'enseignement 2019, en janvier 2018, la population d'enseignants en FW-B était composée de 63 % de femmes dans l'enseignement secondaire. De plus, en 2016-2017, le nombre d'enseignants diplômés en français fle masculins était de 18,8 %, en français morale de 16,7 % et en français religion de 10,8 %.

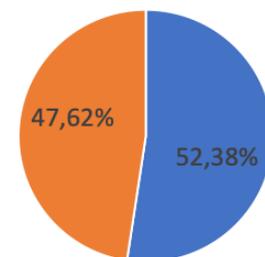


■ Femme ■ Homme
Figure 7 Répartition des enseignants selon le sexe

Dans notre échantillon, les enseignants de sexe masculin sont représentés à hauteur d'environ 14 %. Ce pourcentage semble être assez proche de la population en FW-B.

- Répartition des enseignants selon l'ancienneté

Selon l'enquête TALIS 2018, les enseignants au premier degré du secondaire en FW-B ont en moyenne 18 ans d'expérience professionnelle dont 15 ans comme enseignant. Près de la moitié des enseignants de notre échantillon ont moins de 15 ans d'ancienneté. Ne connaissant pas le nombre exact d'années d'expérience dans l'enseignement pour chaque enseignant, nous ne pouvons calculer la moyenne pour notre échantillon.



■ Moins de 15 ans ■ 15 ans ou plus

Figure 8 Répartition des enseignants selon l'ancienneté

- Répartition des classes selon leur taille

En FW-B, le nombre d'élèves par classe au premier degré du secondaire est déterminé par le décret relatif à l'enseignement fondamental et à l'enseignement secondaire ordinaires, notamment en matière de taille des classes (2012). Sauf dérogation, les classes ne peuvent rassembler plus de 24 élèves. Nous observons dans notre échantillon qu'environ 30 % des classes comportent moins de 20 élèves, les autres étant composées de 20 à 24 élèves.

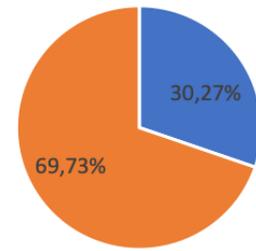


Figure 9 Répartition des classes selon leur taille

- Répartition des enseignants selon leur estimation du nombre d'élèves faibles, attentifs et qui réussiront le CE1D dans leur classe.

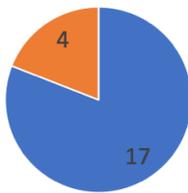


Figure 10 Répartition des enseignants selon leur perception du nombre d'élèves faibles dans leur classe.

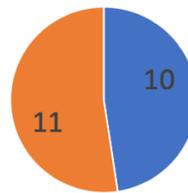


Figure 11 Répartition des enseignants selon leur perception du nombre d'élèves attentifs dans leur classe.

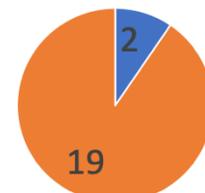


Figure 12 Répartition des enseignants selon leur estimation du nombre de réussites au CE1D dans leur classe.

Les questions concernant la perception des élèves considérés comme attentifs ou comme faibles ainsi que celle concernant l'estimation du nombre de réussites pour chaque classe ne permettent pas de calculer un pourcentage d'élèves par classe et donc de réaliser des analyses statistiques étant donné que les tailles des classes sont différentes ; 10 élèves pouvant en effet représenter moins ou plus de la moitié d'une classe. Nous n'analyserons donc pas ces données de manière statistique, mais nous pouvons tout de même observer les fréquences obtenues. Nous observons donc que 4 enseignants estiment qu'il y a plus de 10 élèves faibles dans leur classe et 10 enseignants répondent que moins de 10 élèves de leur classe sont attentifs. Néanmoins, seuls 2 d'entre eux estiment que moins de 10 élèves ne réussiront pas leur CE1D. Il semble donc, sans confirmation statistique et au seul regard de ces résultats que les enseignants ne prédisent pas toujours un échec au CE1D pour des élèves qu'ils ne considèrent pas comme attentifs ou qu'ils estiment faibles.

- Répartition des enseignants selon leur perception de l'adéquation de l'épreuve du CE1D avec les socles de compétences

À cette question, tous les enseignants ont répondu que l'épreuve du CE1D était en adéquation avec ce qui est prescrit dans les socles de compétences. Il n'y a donc aucune variation entre les enseignants.

1.3. Analyse des fréquences

Suite à l'observation des différentes fréquences, certaines variables ne seront pas utilisées pour réaliser les analyses statistiques.

- Les variables concernant la perception des enseignants du nombre d'élèves considérés comme faibles, attentifs et qui réussiront le CE1D sont supprimées pour la suite de la recherche. Les données récoltées ne permettant pas de calculer un pourcentage d'élèves pour chaque classe, il ne nous semble pas intéressant de les utiliser telles quelles.
- La variable qui concerne l'adéquation avec les socles de compétences est également supprimée pour les analyses, car tous les enseignants y ont répondu positivement. Il n'y a donc aucune variation entre les enseignants.

2. Variance des notes scolaires données par les enseignants

La décomposition de la variance réalisée grâce au modèle vide¹⁵ sert à observer la variance des notes scolaires à l'intérieur des classes, c'est-à-dire entre les élèves d'une même classe et entre les classes ou les écoles, car dans notre échantillon, chaque école n'étant représentée que par une seule classe, leurs variances ne peuvent être distinguées.

Nous trouvons, pour notre échantillon une variance **intra-classe** de 83,3 % et une variance **inter-classes** ou **inter-écoles** de 16,7 %. 83,3 % de la variance des notes données par les enseignants se situent donc à l'intérieur des classes et 16,7 % de cette variance se situent entre les différentes classes et donc pour notre échantillon, également entre les écoles. Un certain pourcentage de variance se situant tout de même entre les classes, nous pouvons avoir recours à une analyse multiniveaux afin d'étudier le phénomène, non pas uniquement à l'intérieur de chaque classe, mais également entre elles.

¹⁵ Les données brutes pour la décomposition de la variance se trouvent en annexe 4.

3. Concordance entre les notes données par les enseignants et les résultats au CE1D

3.1 Concordance toutes classes confondues

Pour vérifier notre première hypothèse, et ainsi observer si les élèves sont classés de la même façon suivant les notes données par les enseignants pendant l'année que suivant celles obtenues au CE1D, nous avons effectué une corrélation de Spearman, c'est-à-dire une corrélation de rang. Nous observons que cette corrélation est de **0.6** et qu'elle est significative.

- De manière générale, les enseignants classent leurs élèves dans un ordre assez semblable à celui dans lequel ils sont placés lors du CE1D. Les notes données par les enseignants concordent donc plutôt bien avec les résultats du CE1D lorsqu'on observe la façon dont elles sont « rangées ».

3.2 Concordance classe par classe

Nous avons réalisé cette même corrélation de rang pour chaque classe individuellement dans le but de répondre à notre deuxième hypothèse : « Le classement réalisé par les enseignants concorde davantage avec celui dans lequel sont placés les élèves sur base de leurs résultats au CE1D lorsqu'on l'observe pour chaque enseignant à l'intérieur de sa classe plutôt que pour l'échantillon complet. »

Tableau 1 Corrélation de Spearman entre les notes de la 3e période et les notes du CE1D par classe

Classes	Corrélation de Spearman	Pr > r	Classes	Corrélation de Spearman	Pr > r
1	0.7*	0.0002	12	0.4	0.1353
2	0.5*	0.0351	13	0.6*	0.0046
3	0.6*	0.0216	14	0.7*	0.0008
4	0.4	0.0590	15	0.7*	0.0004
5	0.5*	0.0319	16	0.7*	0.0005
6	0.6*	0.0049	17	0.8*	<0.0001
7	0.7*	0.0038	18	0.8*	<0.0001
8	0.7*	0.0011	19	0.6*	0.0097
9	0.6*	0.0042	20	0.7*	0.0158
10	0.4*	0.0472	21	0.6*	0.0067
11	0.4	0.0929	22	0.7*	0.0002

*Note. *p < .05.*

Dans le tableau 1, nous observons que les corrélations de rang pour chaque classe individuellement sont toutes significatives sauf trois. Concernant ces corrélations significatives, elles s'étendent de 0.4 à 0.8. Dix d'entre elles sont supérieures à 0.6 qui était la corrélation observée pour l'échantillon complet.

- Le classement des notes données par les enseignants concorde davantage avec le classement des résultats au CE1D pour près de la moitié des enseignants (10 sur 22) composant notre échantillon lorsqu'on observe celui-ci au sein de chaque classe.

3.3 Dispersion des notes à l'intérieur des classes

Dans le tableau 2, les écarts-types des notes données par les enseignants ainsi que des résultats obtenus par les élèves au CE1D sont présentés classe par classe afin de répondre à notre quatrième hypothèse : « Au sein de chaque classe, la dispersion des notes de l'enseignant est plus importante que la dispersion des résultats obtenus au CE1D. »

Tableau 2 Écarts-types des notes données par les enseignants et des résultats obtenus au CE1D

Classes	Écarts-types des notes données par les enseignants	Écarts-types des notes obtenues au CE1D	Classes	Écarts-types des notes données par les enseignants	Écarts-types des notes obtenues au CE1D
1	11.7	9.1	12	13.2	11.2
2	11.9	12.5	13	20.2	13.4
3	14.26	13.8	14	15.8	13.3
4	8.47	9.9	15	13.6	15.0
5	17.0	11.9	16	12.2	14.3
6	14.8	14.2	17	13.9	13.9
7	8.3	8.5	18	14.7	14.1
8	17.3	9.6	19	18.9	15.1
9	11.1	12.2	20	18.6	16.3
10	10.0	10.1	21	16.6	14.1
11	12.4	11.7	22	13.9	14.5

Les écarts-types des notes données par les enseignants sont plus élevés que ceux obtenus pour les notes obtenues au CE1D pour 13 classes sur les 22¹⁶. Nous observons donc qu'un peu plus de la moitié des enseignants de notre échantillon étalent davantage les notes qu'ils donnent à leurs élèves par rapport à la dispersion que nous pouvons

¹⁶ Ces écarts-types sont grisés dans le tableau.

observer selon les résultats du CE1D. Concernant les autres enseignants, un seul d'entre eux note ses élèves avec la même dispersion que celle des résultats au CE1D tandis que les autres attribuent des notes avec une dispersion moindre que l'épreuve standardisée.

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

Dans cette partie, nous présentons les résultats des analyses de régression. Celles-ci ont pour but de tester si certaines variables ont un lien avec les notes que l'enseignant donne à ses élèves. Dans un premier temps, nous observerons le lien entre les résultats au CE1D et les notes données par l'enseignant. Ensuite, nous tiendrons cette variable, représentant la performance, sous contrôle afin d'observer le lien possible entre les autres variables, de niveau élève puis de niveau classe, et les notes données par l'enseignant, le tout, « **à performances égales** » au CE1D.

4.1. Lien entre les notes des enseignants et les résultats au CE1D

Présenté dans le tableau 3, ce premier modèle a pour but de répondre à notre troisième hypothèse : « Il y a un lien entre **les performances des élèves** et les notes qui leur sont données par les enseignants. »

Tableau 3 Analyse de régression entre les résultats au CE1D et les notes données par l'enseignant

Variable dépendante : notes données par l'enseignant		
	Modèle 1	
	β	s. e.
Résultats au CE1D	10.29*	0.59

*Note. * $p < .05$.*

Le coefficient obtenu pour l'analyse de régression entre les résultats obtenus au CE1D et les notes données par l'enseignant est significatif. Cela confirme les résultats obtenus grâce à la corrélation de rang réalisée précédemment. Celle-ci permettait d'affirmer que les notes des enseignants avaient un lien avec les performances des élèves aux épreuves standardisées, car les enseignants classaient leurs élèves relativement de la même manière que le CE1D. Ce premier modèle permet de confirmer qu'il y a effectivement un lien significatif entre les notes données par l'enseignant et les résultats obtenus au CE1D. Les enseignants ne jugent pas leurs élèves arbitrairement.

4.2. Lien entre les notes des enseignants et les variables de niveau « élève »

Tableau 4 Analyses de régression entre les différentes variables de niveau élève et les notes données par l'enseignant, sous contrôle de la performance au CE1D

	Variable dépendante : notes données par l'enseignant							
	Modèle 2		Modèle 3		Modèle 4		Modèle 5	
	β	s. e.	β	s. e.	β	s. e.	β	s. e.
Niveau élève								
Résultats au CE1D	10.47*	0.62	10.50*	0.60	10.29*	0.59	10.23*	0.61
SES de l'élève	0.40	0.60						
Langue parlée à la maison			-1.46	1.80				
Garçon					-4,07*	1.13		
Retard							-4.11*	1.93

Note. * $p < .05$.

- Le niveau socio-économique et culturel de l'élève et la langue parlée à la maison

Dans le tableau 4, le modèle 2 permet de répondre à notre septième hypothèse en observant s'il existe un lien entre les notes données par l'enseignant et le **niveau socio-économique et culturel de l'élève (SES)**. Le modèle 3, quant à lui, permet d'observer si ces notes ont un lien avec le fait de parler français à la maison. À notre connaissance, aucune étude n'ayant été réalisée concernant cette variable, cette analyse est exploratoire. Nous n'avons donc pas émis d'hypothèse quant au résultat.

Dans ces deux modèles, les seuils de signification dépassent 0.05. Ces variables sont donc **non significatives**.

- Cela signifie donc que pour notre échantillon, **à performances égales** au CE1D, il n'y a pas de lien significatif entre les notes données par l'enseignant et :
 - le niveau socio-économique et culturel (SES) de l'élève ;
 - la langue parlée à la maison.

- Le sexe de l'élève et le retard scolaire

Les modèles 4 et 5 permettent de répondre à nos hypothèses 5 et 6 en observant s'il existe un lien d'une part entre les notes de l'enseignant et le **sexe de l'élève** et d'autre part, entre ces notes et le **retard scolaire** de l'élève. Les coefficients de régression obtenus sont significatifs pour ces deux variables. Les notes données par les enseignants

ont donc un lien avec par le sexe de l'élève et avec le retard scolaire de celui-ci, indépendamment de sa performance.

- Pour notre échantillon, à **performances égales** au CE1D, il y a un lien significatif entre :
 - **le sexe de l'élève** et les notes données par l'enseignant.
 - ⇒ À **performances égales**, les garçons obtiennent de moins bonnes notes (4,07 points de moins sur 100) par les enseignants que les filles et inversement.
 - **le retard scolaire** et les notes données par l'enseignant.
 - ⇒ À **performances égales**, les élèves « en retard » obtiennent de moins bonnes notes (4,11 points de moins sur 100) par les enseignants que les élèves « à l'heure » et inversement.

Les variables concernant le sexe de l'élève et le retard scolaire étant significatives lorsqu'elles étaient étudiées séparément, nous avons créé un modèle supplémentaire dans lequel nous avons introduit ces deux variables afin d'observer si, sous contrôle l'une de l'autre, elles avaient toujours un effet significatif.

Tableau 5 Analyse de régression entre les variables sexe de l'élève, retard scolaire et les notes données par l'enseignant, sous contrôle de la performance au CE1D

Variable dépendante : notes données par l'enseignant		
	Modèle 6	
	β	s. e.
Niveau élève		
Variables significatives		
Score au CE1D	10.04*	0.60
Garçon	-4.01*	1.12
Retard	-4.06*	1.91

Note. * $p < .05$.

Nous observons dans le tableau 5 que ces deux variables ont toujours un effet significatif lorsqu'en plus d'être analysées sous contrôle de la performance, elles le sont également l'une sous contrôle de l'autre. Nous pouvons donc conclure, pour notre échantillon :

- **À performances égales et sous contrôle du retard scolaire**, il existe un lien significatif entre les notes données par l'enseignant et **le sexe de l'élève**. Cela signifie qu'à performances égales et à retard scolaire égal, une fille et un garçon, ne seront pas notés de la même façon par l'enseignant.
 - ⇒ À performances égales, un garçon en retard scolaire aura de moins bonnes notes qu'une fille en retard également et un garçon « à l'heure » aura de moins bonnes notes qu'une fille « à l'heure » elle aussi. Ils obtiendront 4,01 points de moins sur 100.
- **À performances égales et sous contrôle du sexe de l'élève**, il existe un lien significatif entre les notes données par l'enseignant et **le retard scolaire**. Cela signifie qu'à performances égales, deux élèves de même sexe ne seront pas notés de la même façon par l'enseignant selon qu'ils soient « à l'heure » ou qu'ils accusent un retard scolaire.
 - ⇒ À performances égales, un garçon en retard aura de moins bonnes notes qu'un garçon « à l'heure » et une fille en retard aura de moins bonnes notes qu'une fille « à l'heure ». Ils obtiendront 4,06 points de moins sur 100.

4.3. Lien entre les notes des enseignants et les variables de niveau « classe »

Certaines variables de niveau classe ont été testées pour répondre à des hypothèses de recherche. Ainsi, les modèles 7, 8, 9 et 10 ont pour objectif de répondre à nos hypothèses 8, 9, 10 et 11 en observant s'il existe un lien entre les notes données par les enseignants et :

- le sexe de l'enseignant
- l'expérience de l'enseignant
- la moyenne de la classe
- l'indice socio-économique de l'établissement.

L'analyse concernant la variable concernant **la taille de la classe** présentée par le modèle 11 est de type **exploratoire** puisqu'à notre connaissance, aucune étude permettant d'émettre une hypothèse n'a été réalisée concernant celle-ci. Le tableau 6 reprend les résultats des analyses de régression avec ces variables de niveau classe.

Tableau 6 Analyses de régression entre les différentes variables de niveau classe et les notes données par l'enseignant, sous contrôle de la performance au CE1D

Variable dépendante : notes données par l'enseignant										
	Modèle 7		Modèle 8		Modèle 9		Modèle 10		Modèle 11	
	β	s. e.	β	s. e.	β	s. e.	β	s. e.	β	s. e.
Niveau classe										
Résultats au CE1D	10.16*	0,60	10.05*	0.60	10.25*	0.59	10.25*	0.62	10.29*	0.59
Sexe de l'enseignant	-3.70	2,58								
Expérience de l'enseignant			0.73	1.91						
Moyenne de la classe					0,23	1.05				
ISE de l'école							0.54	0.46		
Taille de la classe									3.13	2.04

Note. * $p < .05$.

Toutes ces variables de niveau classe sont **non significatives**.

- Pour notre échantillon, à **performances égales**, il n'y a pas de lien significatif entre les notes données par l'enseignant et :
 - le sexe de l'enseignant
 - l'expérience de l'enseignant
 - la moyenne de la classe
 - l'indice socio-économique de l'école (ISE)
 - la taille de la classe

VI. Interprétations et discussion des résultats

1. Concordance des notes données par les enseignants avec les résultats au CE1D

Pour étudier cette concordance, nous avons tout d'abord effectué une corrélation de rang (Spearman). Nos résultats vont dans le même sens que la recherche de Bressoux et Pansu (2003). Nous observons en effet qu'au sein de notre échantillon, les élèves sont « classés » relativement bien de la même façon par leurs enseignants que sur base des résultats au CE1D. Notre coefficient de corrélation de rang (0.6) est néanmoins moins élevé que celui qui était calculé par Hoge et Coladarci dans leur méta-analyse en 1989 (0.8). Nous émettons l'hypothèse que cela pourrait être en partie lié à l'utilisation de méthodologies différentes. Les enseignants de notre échantillon n'ont pas jugé leurs élèves en se basant sur l'épreuve standardisée, mais grâce à leurs propres évaluations alors que dans un certain nombre d'études, le jugement était donné par les enseignants uniquement pour participer à la recherche. Il nous semble donc cohérent que notre coefficient de corrélation soit plus faible que celui obtenu dans cette méta-analyse.

Toutefois, notre corrélation apparaît tout de même comme assez élevée. Pour expliquer cela, nous nous sommes intéressée aux constats de la méta-analyse de Südkamp et al. (2012). Ces auteurs avaient analysé la concordance des notes données par les enseignants et les performances des élèves mesurées grâce à un test standardisé en distinguant les jugements « directs » lorsque les enseignants avaient connaissance du test standardisé et « indirects » lorsqu'ils n'avaient aucune information le concernant. Ils avaient observé une corrélation de Pearson plus élevée lorsque le jugement était « direct » (0.8) que lorsqu'il était « indirect » (0.6). Dans notre recherche, les notes récoltées étant celles que les enseignants avaient attribuées aux élèves lors de la dernière période de l'année, leur jugement n'était pas directement lié à l'épreuve externe. Toutefois, si les enseignants n'ont jamais connaissance des questions du CE1D lorsqu'ils notent leurs élèves durant l'année, ils ont normalement l'habitude du « type » de questions qui composent cette épreuve externe. Leur jugement n'est donc pas totalement « indirect ». Si nous ne pouvons pas comparer directement les coefficients de notre recherche avec ceux de Südkamp et al. car ils n'ont pas été obtenus grâce à la même analyse statistique (corrélation de Pearson pour Südkamp et al. et corrélation de Spearman pour notre recherche), nous pouvons tout de même supposer

qu'une partie de la concordance des notes données par les enseignants avec celles obtenues au CE1D serait due au fait que les enseignants connaissent relativement bien cette épreuve et que leur jugement est donc « direct » d'une certaine manière.

Ensuite, Hoge et Coladarci (1989) avaient observé que les corrélations (Pearson) entre les notes données par les enseignants et les résultats à une épreuve standardisée étaient plus élevées lorsqu'elles étaient calculées pour chaque classe séparément. Nous nous sommes donc demandé si nous pouvions observer la même chose concernant les corrélations de rang. Il apparaît dans nos résultats que certaines corrélations de rang sont effectivement plus élevées lorsqu'on les calcule classe par classe plutôt que sans en faire la distinction. Celles-ci s'étendent de 0.4 à 0.8 alors que, pour rappel, la corrélation de rang concernant l'échantillon complet était de 0.6. Nous observons donc que le classement sur base des notes données par certains enseignants concorde davantage que pour d'autres avec celui obtenu sur base des résultats au CE1D.

Si les résultats que nous avons obtenus montrent que, de manière générale, les élèves les plus performants sont également ceux qui obtiennent les meilleures notes par leurs enseignants et les moins performants les notes les plus faibles, cela ne veut pas forcément dire que les résultats obtenus à l'évaluation standardisée et les notes données par l'enseignant sont les mêmes. Il peut y avoir des écarts entre l'évaluation de l'enseignant et la performance mesurée au CE1D même si les élèves sont classés dans le même ordre. Grisay (1992) attirait l'attention sur le mécanisme de gonflement des variations interindividuelles des élèves par les enseignants. Elle expliquait cela par la tendance de l'enseignant à étaler la dispersion des résultats des élèves de sa classe. Cela correspond à l'effet Posthumus qui, selon Crahay (2007), pousserait les enseignants à ajuster leurs notes pour approcher une distribution gaussienne et donc à surestimer les écarts entre les élèves au sein de leurs classes. Si les enseignants peuvent donc classer leurs élèves relativement de la même façon que l'épreuve standardisée, il est probable que l'écart entre le plus faible et le plus performant soit plus important à l'intérieur de chaque classe.. Nos résultats vont dans ce sens pour une partie des enseignants qui composent notre échantillon. Pour un peu plus de la moitié, nous observons que la dispersion de leurs notes en dernière période est en effet plus importante que celle des résultats au CE1D.

Nous nous questionnons dès lors sur la façon dont cet écart peut être accentué. Hoge et Coladarci (1989) observent que les élèves les plus performants sont surestimés par

leurs enseignants et les plus faibles, sous-estimés. Par contre, Coladarci (1986) et Sanrey et al. (2018) font des observations différentes. Pour Sanrey et al., seuls les plus performants sont surestimés par leurs enseignants alors que pour Coladarci, les élèves les plus performants sont évalués par leurs enseignants avec plus de précision par rapport à la mesure du test standardisé que les élèves les moins performants. Les différences d'écart-types que nous observons sur base de nos résultats montrent que certains enseignants gonflent bien les écarts entre les élèves les plus faibles et les plus performants au sein de leur classe, mais nous n'avons pas observé si certains élèves étaient sur ou sous-estimés dans notre échantillon.

Enfin, pour terminer cette partie concernant la concordance, nous nous sommes attardée sur le lien qui pouvait exister entre les notes données par les enseignants et les résultats au CE1D en calculant un coefficient de régression. Celui-ci est significatif et va donc dans le sens de la littérature. Hoge et Coladarci (1989), Südkamp et al. (2012) et Sanrey et al. (2018) observaient des corrélations assez importantes entre ces deux variables. Le résultat que nous obtenons, même s'il ne se base pas sur la même analyse statistique, montre également qu'il y a un lien entre celles-ci. Cela veut donc dire que les enseignants n'attribuent pas les notes aux élèves de façon arbitraire ; les notes que reçoivent les élèves de la part de leur enseignant ont bien un lien avec leurs performances mesurées par le CE1D, ce qui est plutôt rassurant.

Si le lien entre les notes données par les enseignants et la performance des élèves n'est plus à démontrer, Hoge et Coladarci (1989) soulignaient tout de même que le pourcentage de variance expliquée par les performances des élèves n'était que de 50 %. D'autres variables sont donc à prendre en compte pour expliquer les différences de jugement des enseignants pour des élèves qui ont des performances égales. C'est ce que nous avons tenté de faire en introduisant différentes variables de niveau élève et de niveau classe dans d'autres analyses de régression.

2. Les biais du jugement de l'enseignant

2.1. Le sexe de l'élève

Les résultats que nous avons obtenus concernant le lien entre le sexe de l'élève et les notes données par les enseignants montrent qu'à performances égales, les filles obtiennent de meilleures notes que les garçons. Plus précisément, dans notre

échantillon, elles obtiennent 4,07 points sur 100 de plus que les garçons. Cela va dans le sens de la recherche menée par Felouzis (cité par Bressoux & Pansu, 2003) qui avait réalisé une étude concernant des élèves de 4^e secondaire en français. C'est sur celle-ci que notre hypothèse était basée, considérant que l'âge des élèves n'était pas très éloigné de celui des élèves de notre échantillon et qu'elle concernait la même matière. Par contre, nos résultats vont à l'encontre de ceux obtenus par Sanrey et al. (2018) qui montraient que les garçons obtenaient de meilleures notes que les filles dans une recherche qui concernait les domaines du français et des mathématiques ensemble. Bressoux et Pansu (2003) ayant trouvé un effet significatif du sexe de l'élève en faveur des garçons pour les mathématiques, nous supposons que cette matière a pris le dessus dans l'étude de Sanrey et al. et que celle-ci n'est donc pas représentative du lien qui peut exister entre le sexe de l'élève et le jugement de l'enseignant en français uniquement.

Nous avons réalisé une deuxième analyse de régression concernant le sexe de l'élève toujours réalisée à performances égales, mais également **sous contrôle du retard scolaire**. Les résultats montrent que, pour notre échantillon, à performances égales et à retard scolaire égal, les garçons obtiennent 4,06 points de moins que les filles. Il apparaît donc que les notes données par l'enseignant sont toujours en lien avec le sexe de l'élève en tenant sous contrôle le retard scolaire, mais que ce coefficient est un peu moins élevé que lorsqu'on ne tient pas compte de cette variable. Bressoux et Pansu (2003) n'avaient, eux, pas observé d'effet significatif du sexe de l'élève sur le jugement de l'enseignant en français, à performances et à retard égaux. Cette différence entre leur étude et la nôtre pourrait être expliquée par le fait que le niveau d'étude concerné n'était pas le même.

Cet effet significatif du sexe de l'élève nous pose question. Pourquoi les notes que donne un enseignant à ses élèves sont-elles différentes en fonction du sexe de ces derniers ? Nous nous sommes intéressée à deux facteurs qui ont été étudiés dans la littérature, à savoir les stéréotypes en fonction du sexe de l'élève par rapport à la matière étudiée ainsi que la perception du comportement des élèves par les enseignants.

Premièrement, concernant **les stéréotypes**, Jarlegan et al. (2011) rapportent que les enseignants perçoivent les garçons comme étant plus compétents dans des disciplines à connotation masculine comme les mathématiques et les filles comme étant plus compétentes dans celles qui ont une connotation féminine comme les langues. La

différence de notation que nous observons en faveur des filles pourrait donc être expliquée par ce stéréotype. Les enseignants percevant les filles comme plus compétentes que les garçons en français, ils leur attribueraient de meilleures notes qu'aux garçons qui ont les mêmes performances.

Deuxièmement, concernant **la perception du comportement**, Bennett et al. (1993) ont réalisé une étude concernant des élèves de maternelle et de primaire. Ils n'observaient pas de lien significatif entre le sexe de l'élève et le jugement de l'enseignant. Par contre, ils observaient une relation significative entre le sexe de l'élève et la perception du comportement par les enseignants et également entre cette perception du comportement et les notes données par les enseignants. Le sexe de l'élève pourrait donc avoir un lien avec les notes données par l'enseignant via le comportement de l'élève. Selon leur recherche, les garçons étant plus souvent considérés comme ayant un moins bon comportement que les filles, ils obtiendraient donc, à performances égales, de moins bonnes notes que celles-ci. La méta-analyse de Dusek et Joseph (1983) montre également que les élèves présentés comme ayant un moins bon comportement en classe sont jugés comme ayant moins de chance de réaliser de bonnes performances scolaires, mais également comme plus enclins à redoubler. N'ayant pas étudié la perception du comportement dans notre recherche, nous ne pouvons que nous questionner et émettre l'hypothèse selon laquelle, à performances égales, les garçons pourraient obtenir de moins bonnes notes que les filles parce qu'ils sont perçus comme ayant un moins bon comportement.

2.2. Le retard scolaire

Les résultats que nous obtenons concernant le lien entre le retard scolaire et les notes données par les enseignants vont dans le sens de l'étude de Bressoux et Pansu (2003) ainsi que celles qu'ils ont rapportées. Dans celles-ci, à performances égales, les élèves en retard scolaire obtenaient de moins bonnes notes que leurs camarades « à l'heure » pour le niveau primaire ainsi que pour le début du secondaire. Les seuls résultats non significatifs apparaissaient pour les élèves de fin du secondaire. Notre échantillon ne concernant pas ce niveau, notre hypothèse était basée sur les premières études et celle-ci est donc confirmée pour notre échantillon. Il existe un lien entre le retard scolaire et les notes données par l'enseignant ; à performances égales, les élèves en retard scolaire obtiennent 4,11 points de moins sur 100 que les élèves « à l'heure ».

Nous avons également analysé le lien qui pourrait exister entre les notes données par les enseignants et le retard scolaire, à performances égales et **sous contrôle du sexe de l'élève**. Nos résultats vont toujours dans le même sens, c'est-à-dire que pour des élèves qui ont les mêmes performances au CE1D et qui sont en retard scolaire, les garçons sont moins bien notés que les filles, ils obtiennent 4,01 points en moins sur 100.

Nous nous questionnons dès lors sur les possibles explications de ces différences de notation en fonction du retard scolaire de l'élève. Pourquoi les élèves « en retard » obtiennent-ils de moins bonnes notes que leurs camarades « à l'heure » alors qu'ils ont pourtant des performances égales ?

« L'institution scolaire n'est pas immunisée contre l'émergence des stéréotypes et leurs impacts sur les comportements. » (Desombre et al., 2008, p.4) En 1968, Rosenthal et Jacobson (cités par Desombre et al., 2008) faisaient remarquer que les membres de groupes défavorisés avaient davantage de risques d'induire des attentes défavorables par rapport à ceux qui étaient plus favorisés. Potvin et Rousseau (1994) observaient également que les enseignants avaient des attitudes moins positives envers les élèves qui étaient en retard au niveau scolaire par rapport à ceux qui étaient « à l'heure ». Les stéréotypes concernant les élèves qui sont en retard scolaire amèneraient donc les enseignants à avoir de moins bonnes attentes envers ces élèves et à leur donner de moins bonnes notes qu'à ceux qui sont « à l'heure » même s'ils ont des performances égales.

2.3. Le niveau socio-économique et culturel de l'élève

Nos résultats montrent que, pour notre échantillon, les notes données par les enseignants ne sont pas liées au niveau socio-économique de l'élève. Cela va à l'encontre de plusieurs études qui ont été réalisées à divers niveaux scolaires. Bressoux et Pansu (2003) observaient un effet significatif de cette variable pour des enseignants en CE2 (3^e primaire) et Duru-Bellat et Mingat (1993) observaient la même chose en 5^e (2^e secondaire). Ces deux études concluaient que les enseignants attribuaient de moins bonnes notes aux élèves issus de milieux défavorisés qu'aux autres, alors qu'ils avaient pourtant des performances égales. La seule étude que nous avons consultée et qui n'observait pas d'effet significatif concernait des élèves au niveau du lycée. Les élèves de notre échantillon n'étant pas de ce niveau scolaire, nous ne nous basons pas sur celle-ci pour justifier l'effet non significatif que nous observons.

Nous émettons plusieurs hypothèses sur le fait que nos résultats vont à l'encontre de ces deux premières études. Tout d'abord, même si cela nous semble peu probable au vu de la littérature, ce résultat non significatif pourrait refléter le fait que le niveau socio-économique et culturel n'a effectivement pas de lien avec les notes que donnent les enseignants de notre échantillon à leurs élèves. Ceux-ci pourraient ne pas juger leurs élèves différemment en fonction de leur niveau socio-économique.

D'autres hypothèses nous semblent toutefois plus probables. Cela pourrait premièrement être dû au fait que notre échantillon n'est pas assez important pour observer l'effet du niveau socio-économique des élèves comme l'ont mentionné Dusek et Joseph (1983) en précisant que si les attentes des enseignants étaient en lien avec le SES de l'élève pour un échantillon comprenant plusieurs études, un certain nombre d'entre elles ne montraient pas d'effet significatif individuellement. Notre deuxième hypothèse concerne la différence de méthodologie pour calculer cet indice socio-économique. La plupart des études consultées prenaient uniquement en compte le métier des parents alors que nous nous sommes également basée sur certains biens matériels que possédaient les élèves.

2.4. La langue parlée à la maison

À notre connaissance, aucune étude n'a été réalisée concernant un possible lien entre le jugement de l'enseignant et la langue parlée par l'élève à la maison. Nous avons donc étudié cette variable de manière exploratoire dans notre recherche. Pour notre échantillon, le coefficient de régression n'est pas significatif. Il n'y a donc pas de lien entre cette variable et les notes que l'enseignant donne à ses élèves. Les enseignants ne donnent donc pas de notes différentes aux élèves qui ont des performances égales selon qu'ils parlent français chez eux ou pas. Si le pourcentage d'élèves qui ne parlent que parfois ou même jamais français chez eux dans notre échantillon (10,35 %) est assez proche de celui qui est présenté en 2019 dans les premiers résultats de PISA 2018 en Fédération Wallonie-Bruxelles (17,4 %), il est tout de même un peu plus bas. Ce pourcentage, au niveau de notre échantillon qui contient 434 élèves ne représente pas énormément d'élèves. Cela ne permet donc peut-être pas d'observer adéquatement, par analyse statistique, la variabilité des notes que les enseignants donnent à leurs élèves, sur base de cette variable.

2.5. Le sexe de l'enseignant

Concernant le lien entre les notes données par l'enseignant, les études consultées ne s'accordent pas. Chatel (cité par Merle, 1998) observait que les enseignants masculins donnaient de moins bonnes notes à leurs élèves que leurs collègues de sexe féminin pour le cours de sciences sociales alors que ces élèves avaient pourtant un même niveau de performance. Par contre, d'autres recherches comme celle de Good et al. en 1974 (cités par Brophy, 1985) ainsi que celle de Parker et Claxton en 1996 n'observaient pas de différence de notation des enseignants en fonction de leur sexe. Lafontaine et Monseur (2009) observaient quant à eux, de manière générale, que les enseignants, quel que soit leur sexe, donnaient de moins bonnes notes aux élèves du même sexe qu'eux. Néanmoins, ils invitaient à être prudent par rapport à ces résultats car en réalisant d'autres analyses, ils n'observaient pas cela pour toutes les copies. Chouinard (2002) précise également qu'il faut faire preuve de vigilance par rapport aux interprétations des analyses qui concernent le sexe de l'enseignant car, souvent, les enseignants masculins sont peu nombreux dans les échantillons.

Dans notre recherche, nous n'observons pas d'effet significatif du sexe de l'enseignant. Autrement dit, nous ne pouvons pas affirmer que les enseignants, selon qu'ils soient de sexe féminin ou masculin, attribuent des notes différentes à des élèves qui ont des performances égales. Notre échantillon, même s'il reflète assez bien la composition des enseignants selon leur sexe en Fédération Wallonie-Bruxelles, ne comporte effectivement pas énormément d'enseignants de sexe masculin comme le faisait remarquer Chouinard. Nous restons donc prudente quant à l'interprétation de nos résultats.

2.6. L'expérience de l'enseignant

À notre connaissance, peu d'études ont été réalisées concernant le lien entre le jugement des enseignants et leur expérience. Selon celle que nous avons consultée, réalisée en 2009 par Krolak-Schwerdt et al. (cités par Artelt & Rausch, 2014), les élèves recevaient des notes plus concordantes avec leur performance par les enseignants plus expérimentés que par ceux qui débutaient. Nos résultats ne rejoignent pas cette observation, nous n'observons pas de différence significative entre les notes données par les enseignants en fonction de leur expérience. Cela pourrait s'expliquer par l'utilisation de méthodologies différentes. L'étude de Krolak-Schwerdt et al. (2009) a

été réalisée sur base d'un échantillon d'enseignants en fonction ainsi que d'étudiants, futurs enseignants alors que la nôtre concernait uniquement des enseignants en fonction. De plus, les élèves de l'étude consultée étaient fictifs, ce qui n'est pas le cas non plus dans notre recherche. Les enseignants et les élèves de notre échantillon étaient en effet des sujets réels qui sauf exception, se connaissaient depuis au moins un an étant donné que les données ont été récoltées en fin d'année scolaire. Il y aurait donc probablement d'autres variables qui interviennent et qui modèrent l'effet de l'expérience de l'enseignant dans le cas où ils connaissent leurs élèves.

2.7. La moyenne de la classe

Les résultats obtenus pour notre échantillon vont à l'encontre de la plupart des résultats observés dans la littérature et donc de l'hypothèse que nous avons posée. En effet, nous n'observons pas de lien significatif entre les notes qui sont données par les enseignants et la moyenne de la classe qui a été calculée sur base des résultats obtenus par les élèves au CE1D. Cette différence entre les études antérieures et la nôtre pourrait tout d'abord s'expliquer par une trop faible variabilité des moyennes des classes ; néanmoins, les moyennes observées pour notre échantillon allant de 47,7 % à 74,9 %, nous ne pensons pas que cela en soit la raison. Nous expliquerions plutôt cela par la différence des matières étudiées. Bressoux et Pansu (2003) observaient que le jugement des enseignants était lié au niveau moyen de la classe en mathématiques et en français pour des élèves de niveau primaire, en CE2. Meissel et al. (2017), pour un échantillon qui comportait des élèves âgés de 8 à 13 ans, observaient également ce lien pour des tâches d'écriture, mais pas forcément en lecture. Si nous nous attendions donc à observer également un lien significatif de la moyenne de la classe sur le jugement des enseignants en 2^e secondaire, notre étude porte sur toutes les compétences demandées en français à ce niveau et pas uniquement sur des tâches d'écriture. Cela pourrait donc expliquer que nous n'observons pas de lien pour notre échantillon.

Nous trouvons tout de même étonnant que nos résultats montrent que les notes données par les enseignants n'aient pas de lien avec la moyenne de la classe pour notre échantillon. Selon Crahay (2007) l'effet Posthumus pouvait expliquer, que pour une performance donnée, un élève pouvait être moins bien noté dans une classe dont la moyenne était élevée que dans une classe dont la moyenne était plus faible. « Un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des

performances des élèves de façon à conserver d'année en année, approximativement une distribution (gaussienne) des notes ». (Crahay, 2007, p.79) Selon Bressoux et Pansu (2003), les enseignants jugeraient leurs élèves de cette manière pour obtenir une moyenne « socialement acceptable ». Si nous avons observé précédemment que les enseignants de notre échantillon exagéraient l'écart autour de la moyenne entre l'élève le plus faible et le plus performant au sein de leurs classes, ils ne notent pas leurs élèves différemment en fonction de celle-ci.

2.8. La taille de la classe

Le lien entre les notes données par l'enseignant et la taille de la classe n'a, à notre connaissance, pas été étudié dans la littérature. Pour notre échantillon, nous n'observons pas de lien significatif entre ces deux variables alors que les classes qui le composaient contenaient, pour 30 % d'entre elles, moins de 20 élèves, les autres étant composées de 20 à 23 élèves. Les notes données par les enseignants à des élèves qui ont un niveau de performance égal ne diffèrent donc pas en fonction de la taille de la classe.

2.9. L'indice socio-économique de l'établissement

Cette variable n'a, à notre connaissance, pas été largement étudiée dans les études qui concernent le jugement des enseignants. Les résultats que nous obtenons vont dans le sens de l'étude de Meissel et al. (2017) qui montrait que des élèves qui avaient des performances égales n'étaient pas notés différemment par leurs enseignants selon l'indice socio-économique de l'établissement. Cela voudrait dire que les enseignants de notre échantillon n'adaptent pas leurs notes en fonction du public auquel ils sont confrontés de manière générale dans leur établissement.

3. Biais et limites de la recherche

La première limite de notre recherche est liée à l'échantillon. En effet, si la demande de participation a été faite à un grand nombre d'écoles de la Fédération Wallonie-Bruxelles par mail aux directions, seuls les enseignants qui ont accepté de participer font partie de celui-ci. Notre échantillon n'est, de ce fait, pas aléatoire et simple et comprend uniquement 21 enseignants en français ainsi que 434 élèves. Ce nombre peut ne pas être assez important pour observer certains effets qui pourraient être significatifs avec un plus grand nombre d'observations. De plus, il ne permet pas d'élargir les résultats et les conclusions à l'ensemble de la Fédération Wallonie-Bruxelles. Les observations réalisées ainsi que les conclusions que nous tirons sur base de notre recherche sont donc uniquement applicables à notre échantillon.

Deuxièmement, si certaines études se rapprochent de la nôtre au niveau méthodologique, d'autres diffèrent sur ce point. Dans notre recherche, le jugement correspond à la note donnée par l'enseignant à la dernière période de l'année alors que dans un certain nombre des études réalisées préalablement, le jugement de l'enseignant récolté n'était pas donné en situation réelle. Celui-ci était émis uniquement à la demande des chercheurs. De plus, pour certaines, le jugement de l'enseignant n'est pas représenté par une note, mais par exemple par une prédiction de la réussite ou non à l'ensemble du test ou à chacun des items. Nous pouvons donc tenter d'interpréter nos résultats en fonction de celles-ci, mais cette différence de méthodologie peut biaiser la comparaison.

Enfin, concernant les données chiffrées récoltées, nous sommes confrontée à deux limites. Tout d'abord, si nous avons choisi d'utiliser les notes données par l'enseignant à la dernière période de l'année, c'est surtout parce qu'elles sont les plus proches au niveau temporel de celles obtenues par les élèves au CE1D. Néanmoins, cette épreuve couvre la matière des deux premières années du secondaire, la note donnée par l'enseignant n'est sûrement pas basée sur l'entièreté de celle-ci. N'ayant pas eu accès aux évaluations sur lesquelles sont basées ces notes, nous ne pouvons pas affirmer que les enseignants ont évalué la même chose que ce qui l'a été au CE1D. Ensuite, si aucune mesure n'est la mesure « réelle » de la performance, utiliser les résultats du CE1D pour la calculer permet d'avoir le même instrument de mesure pour tous les élèves et ainsi espérer diminuer les biais. Les élèves passent en effet tous la même épreuve avec les mêmes consignes. Néanmoins, passant les épreuves au sein de leurs propres écoles, il est possible que certains aient bénéficié d'explications supplémentaires par rapport à

d'autres. De plus, pour le CE1D, si les critères de correction sont les mêmes pour tous, celle-ci est néanmoins confiée aux enseignants. Cela peut alors biaiser la comparabilité des résultats obtenus à cette épreuve que nous considérons comme la performance des élèves. Nos observations ne sont sans doute pas toutes complètement « à performances égales », mais s'en rapprochent autant que possible.

VII. Conclusion et perspectives

Si les enseignants doivent remplir des missions telles que transmettre des connaissances et faire acquérir des compétences aux élèves, ils doivent également émettre un jugement scolaire sur ceux-ci (Bressoux & Pansu, 2003).

Les résultats de notre recherche concernant la concordance des notes données par les enseignants en français avec celles que les élèves obtiennent au CE1D montrent que les enseignants classent assez bien leurs élèves de la même manière que le classement obtenu sur base des résultats au CE1D. Ces classements concordent encore davantage au sein de chaque classe séparément. Il apparaît également qu'il y a bien un lien entre les notes données par les enseignants et la performance des élèves telle qu'elle mesurée par l'épreuve externe. Toutefois, nous avons observé qu'au sein de notre échantillon, un peu plus de la moitié des enseignants dispersaient davantage les notes de leurs élèves que les résultats du CE1D. Autrement dit, certains enseignants exagèrent la différence entre l'élève le plus performant et le plus faible de leur classe.

Nous nous sommes ensuite intéressée à certaines variables de niveau élève et de niveau classe qui pourraient avoir un lien avec ces notes. Deux variables de niveau élève se sont révélées significatives dans notre recherche, à savoir le sexe de l'élève et le retard scolaire. Pour notre échantillon, nous observons qu'à performances égales, les filles sont mieux notées en français que les garçons et les élèves en retard scolaire sont moins bien notés que leurs camarades « à l'heure ». Si nous pensons que ces deux variables ont un effet significatif à cause de certains stéréotypes (Jarlegan et al., 2011, Desombre et al., 2008), concernant plus précisément le sexe de l'élève, nous pensons qu'il pourrait être intéressant d'observer si le comportement de l'élève modère l'effet de cette variable comme le précisaient Bennett et al., 1993 et Dusek & Joseph, 1983. Les autres variables de niveau élève n'ont pas révélé d'effet significatif. Pour notre échantillon, il n'y a donc pas de lien entre les notes données par l'enseignant et la langue parlée par les élèves à la maison ni entre ces notes et le niveau socio-économique de l'élève.

Toutes les variables de niveau classe que nous avons étudiées se sont également révélées non significatives. Nous n'observons pas de lien entre les notes données par les enseignants et l'indice socio-économique de l'établissement, le sexe de l'enseignant, son expérience ainsi que la moyenne de la classe. Si cela paraît cohérent pour la plupart

des variables en comparaison avec la littérature, il est étonnant d'obtenir ces résultats pour la moyenne de la classe. Nous pensons qu'il serait utile de poursuivre les recherches concernant cette variable avec par exemple un échantillon plus important, car il nous est difficile de justifier qu'elle ne soit pas significative pour notre échantillon.

En terminant ce travail, nous sommes convaincue que le jugement des enseignants est extrêmement important pour le bien des élèves, car il peut notamment permettre de détecter leurs difficultés afin de pouvoir les aider, mais également mener au redoublement de manière abusive et influencer positivement comme négativement la perception que les élèves ont d'eux-mêmes. S'il nous semble rassurant que les enseignants classent leurs élèves relativement bien dans le même ordre que le CE1D et qu'ils notent assez bien leurs élèves par rapport à leurs performances, il apparaît tout de même que d'autres variables ont un lien avec ce jugement. Si un jugement ne peut être totalement objectif, nous pensons qu'il est important que les enseignants soient conscients que leurs notes ne sont pas uniquement le reflet de la performance des élèves même s'ils pensent s'en approcher. Cela pourrait peut-être permettre de diminuer certains biais dans leur notation notamment par rapport au sexe de l'élève ainsi qu'à son retard scolaire. Comme le précisait Lafontaine (2001), nous espérons qu'une épreuve obligatoire et commune comme le CE1D puisse permettre aux enseignants d'avoir un point de repère par rapport à leurs évaluations et à leurs notes afin de limiter certains biais.

IX. Bibliographie

- Allal, L., & Lafortune, L. (2008). A la recherche du jugement professionnel. In L. Allal, & L. Lafortune (Eds), *Jugement professionnel en évaluation* (pp.1-14). Presses de l'Université du Québec.
- 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>
- Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347-356. <https://doi.org/10.1037/0022-0663.85.2.347>
- Beswick, J., Willms, J., & Sloat, E. (2005). A comparative study of teacher ratings of emergent literacy skills and student performance on a standardized measure. *Education, 126* (1), 116-137.
- Bressoux P. (1998). Felouzis (Georges). - L'efficacité des enseignants. In *Revue française de pédagogie, 124*, Sociologie de l'éducation. 169-171.
- Bressoux, P., & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Presses universitaires de France.
- Bressoux, P. (2018). 1. Comment se fabrique le jugement des enseignants ? *Regards croisés sur l'économie, 22* (1), 15-23. <https://doi.org/10.3917/rce.022.0015>

- Brophy, J. (1985). Interactions of male and female students with male and female teachers. In L.C. Wilkinson, & C.B. Marrett (Eds.), *Gender influences in classroom interaction*. Academic Press.
- Cardinet, J. (1987). *L'objectivité de l'évaluation*. Institut romand de recherches et de documentation pédagogiques.
- 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° 6977 de la communauté française du 6 février 2019 concernant les dispositions relatives à l'octroi du Certificat d'études de base (CEB) à l'issue de l'épreuve externe commune pour l'année scolaire 2018-2019. (2019).
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78 (2), 141-146.
<http://dx.doi.org/10.1037/0022-0663.78.2.141>
- Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire?* (3rd ed.). De Boeck Supérieur. doi:10.3917/dbu.craha.2007.01.
- Décret de la Communauté française du 24 juillet 1997 définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre. (1997). *Moniteur belge*, 23 septembre, p.24653
- 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.

- Décret de la Communauté française du 3 mai 2012 relatif à l'enseignement fondamental et à l'enseignement secondaire ordinaires, notamment en matière de taille de classe. (2012). *Moniteur belge*, 15 juin, p.33035.
- Desombre, C., Delelis, G., Lachal, M., Urban, E., Roye, L., Gaillet, F., & Antoine, L. (2008). Stéréotypes de la difficulté scolaire : un outil de recueil. *L'orientation scolaire et professionnelle*, 37 (2), 215-239. <https://doi.org/10.4000/osp.1673>
- Duru-Bellat, M. & Mingat, A. (1993). *Pour une approche analytique du fonctionnement du système éducatif*. Presses Universitaires de France.
- Dusek, J. & Joseph, G. (1983). The Bases of Teacher Expectancies: A Meta-Analysis. *Journal of Educational Psychology*, 75(3), 327-346. <https://doi.org/10.1037/0022-0663.75.3.327>
- Fédération Wallonie-Bruxelles. (2019a) *Épreuve externe certificative – CE1D*. <http://www.enseignement.be/index.php?page=26247&navi=30570>.
- Fédération Wallonie-Bruxelles. (2019b) *Les outils d'évaluation*. <http://www.enseignement.be/index.php?page=24345>
- Fédération Wallonie-Bruxelles (2019). *Les indicateurs de l'enseignement* (14^e éd.).
- Fédération Wallonie-Bruxelles. (2020). Référentiels de base – Présentation des programmes. <http://www.enseignement.be/index.php?page=25279&navi=297>
- Gerard, F.-M. (2002). L'indispensable subjectivité de l'évaluation. *Antipodes*, 156, 26-34.
- Gerard, F.-M. (2017). Objectiver la subjectivité. In D. Leduc & S. Béland (Eds.), *Regards sur l'évaluation des apprentissages en arts à l'enseignement supérieur, Tome 1* (29-47). Presses de l'Université du Québec.

- Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Soc Psychol Educ, 16*, 111-127. <https://doi.org/10.1007/s11218-012-9197-z>
- Grisay, A. (1992). Examens et échecs dans l'enseignement fondamental. Une recherche-action sur les pratiques d'évaluation-bilan de cinquante écoles primaires en Communauté française de Belgique. *Le Nouvel Observateur*, (1437).
- Hoge, R. D., & Coladarci, T. (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research, 59*(3), 297-313. <https://doi.org/10.3102/00346543059003297>
- Hutmacher, W. (1993). Quand la réalité résiste à la lutte contre l'échec scolaire. Analyse du redoublement dans l'enseignement primaire genevois. Genève : *Service de la recherche sociologique, Cahier 36*. Réédition numérique par le service de la recherche en éducation, 2010.
- Jarlegan, A., Tazouti, Y., & Flieller, A. (2011). L'hétérogénéité sexuée en classe : effet de genre sur les attentes des enseignant(e)s et les interactions verbales enseignant(e) -élève. *Les dossiers des sciences de l'éducation, 26*, 33-50. <https://doi.org/10.4000/dse.1073>
- Jussim, L., & Eccles, J. S. (1992). Teacher expectations: II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology, 63*(6), 947-961. <https://doi.org/10.1037/0022-3514.63.6.947>
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. [Abstract]. *Zeitschrift für Pädagogische Psychologie, 23*, 175-186. <https://doi.org/10.1024/1010-0652.23.34.175>.
- 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>
- Lafontaine, D. (Ed.), Quittre, V., & Dupont, V. (2019). *TALIS 2018 : Paysage de l'enseignement du 1^{er} degré du secondaire*. Service d'analyse des Systèmes et des Pratiques d'enseignement.
- Lafontaine, D. (Ed.), Schillings, P., Dupont, V., Dejaegher, C., Geron, S., & Matoule, A. (2018). *PIRLS 2016 : Étude internationale sur les compétences en lecture. Rapport final*. Service d'analyse des Systèmes et des Pratiques d'enseignement.
- Leclercq, D., Nicaise, J., & Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. In M. Demeuse (Ed.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (pp.273-292). Les éditions de l'Université de Liège.
- Levine, E. (1981). *Teachers' academic and psycho-social expectations for children from single-parent families*. Doctoral Dissertations 1896 – February 2014. https://scholarworks.umass.edu/dissertations_1/3690
- Machts, N., Kaiser, J., Schmidt, F., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research review*, 19, 85-103. <https://doi.org/10.1016/j.edurev.2016.06.003>
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Presses universitaires de France.

- 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>
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research*, 79(3), 1129-1167. <https://doi.org/10.3102/0034654309337522>
- 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.
- Parker-Price, S. & Claxton, A. F. (1996). *Teachers' perceptions of gender differences in students*. Communication présentée à la Convention annuelle de la «National Association of School Psychologist ». <https://files.eric.ed.gov/fulltext/ED397373.pdf>
- Potvin, P., & Rousseau, R. (1994). Attitudes des enseignants envers les élèves en difficulté scolaire. *Canadian Journal of Education / Revue canadienne de l'éducation*, 18 (2), 132-149. <https://doi.org/10.2307/1495186>
- Ritts, V., Patterson, M., & Tubbs, M. (1992). Expectations, Impressions, and Judgments of Physically Attractive Students: A Review. *Review of Educational Research*, 62(4), 413–426. <https://doi.org/10.3102/00346543062004413>
- Roegiers, X. (2010). Chapitre 2. Un cadre de référence en matière d'évaluation des acquis des élèves. In X. Roegiers (Ed.), *L'école et l'évaluation: Des situations complexes pour évaluer les acquis des élèves* (pp. 51-108). De Boeck Supérieur.
- Rosenzweig, P. (2009). L'effet halo, ou les mirages de la performance. *Le journal de l'école de Paris du management*, 79(5), 9-16. <https://doi.org/10.3917/jepam.079.0009>

- Sanrey, C., Pansu, P., & Bressoux, P. (2018). *Le jugement des enseignants reflète-t-il les compétences réelles des élèves ?* Conférence : Congrès de la Société Française de Psychologie.
- Santrock, J. W., & Tracy, R. L. (1978). Effects of children's family structure status on the development of stereotypes by teachers. *Journal of Educational Psychology, 70*(5), 754–757. <https://doi.org/10.1037/0022-0663.70.5.754>
- 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>
- Tournaki, N. (2003). Effect of Student Characteristics on Teachers' Predictions of Student Success. *The Journal of Educational Research, 96*(5), 310-319. <https://doi.org/10.1080/00220670309597643>
- Vial, M. (2012). Chapitre 1. Le modèle de la mesure : L'évaluation des produits (des états) : l'évaluation comme mesure (des effets, des impacts, des comparaisons de bilans). In M. Vial, *Se repérer dans les modèles de l'évaluation : Méthodes – Dispositifs – Outils* (pp. 19-48). De Boeck Supérieur.
Doi : 10.3917/dbu.vial.2012.01

X. Table des tableaux et figures

Figure 1 Schéma des étapes de l'évaluation sur base de Mottier Lopez (2008) et Rœgiers (2010)	11
Figure 2 A model of teacher-based judgments of students' academic achievement. (Südkamp et al., 2012, p.756)	20
Figure 3 Répartition des enseignants selon le réseau	43
Figure 4 Répartition des élèves en fonction du sexe	49
Figure 5 Répartition des élèves en fonction du retard scolaire	49
Figure 6 Répartition des élèves en fonction de l'utilisation du français à la maison ...	50
Figure 7 Répartition des enseignants selon le sexe	50
Figure 8 Répartition des enseignants selon l'ancienneté.....	50
Figure 9 Répartition des classes selon leur taille	51
Figure 10 Répartition des enseignants selon leur perception du nombre d'élèves faibles dans leur classe.	51
Figure 11 Répartition des enseignants selon leur perception du nombre d'élèves attentifs dans leur classe.	51
Figure 12 Répartition des enseignants selon leur estimation du nombre de réussites au CE1D dans leur classe.	51
Tableau 1 Corrélation de Spearman entre les notes de la 3e période et les notes du CE1D par classe	53
Tableau 2 Écarts-types des notes données par les enseignants et des résultats obtenus au CE1D	54
Tableau 3 Analyse de régression entre les résultats au CE1D et les notes données par l'enseignant	55
Tableau 4 Analyses de régression entre les différentes variables de niveau élève et les notes données par l'enseignant, sous contrôle de la performance au CE1D	56
Tableau 5 Analyse de régression entre les variables sexe de l'élève, retard scolaire et les notes données par l'enseignant, sous contrôle de la performance au CE1D.....	57
Tableau 6 Analyses de régression entre les différentes variables de niveau classe et les notes données par l'enseignant, sous contrôle de la performance au CE1D	59

XI. Annexes

1. Annexe 1 : Document destiné à la récolte des notes

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 destiné aux enseignants



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 : À 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.				

3. Annexe 3 : Questionnaire destiné aux élèves¹⁷



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..... 2005..... 2008.....

2003..... 2006..... 2009.....

2004..... 2007..... 2010.....

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

1

¹⁷ Questionnaire créé sur base des questionnaires élèves de PIRLS 2016 et de PISA 2009. OECD PISA. (2008). *Questionnaire élève PISA 2009*. IEA & aSpE Université de Liège. (2015). *PIRLS 2016 : Questionnaire élève*.

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



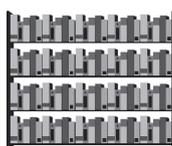
Assez pour remplir une étagère
(entre 11 et 25 livres) 2

Ceci représente 25 livres



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

Ceci représente 100 livres



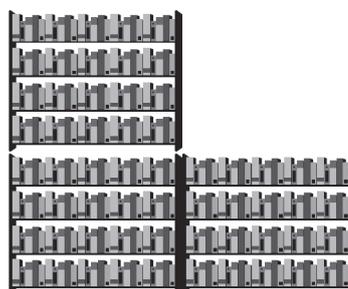
Assez pour remplir
deux bibliothèques
(entre 101 et 200 livres) 4

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 |

4. Annexe 4 : Données brutes pour la décomposition de la variance des notes scolaires

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	codeecole	40.9751
Residual		204.08