
Opportunités d'apprentissage, styles de réponse et performance en culture mathématique : analyse des données de PISA 2012

Auteur : Pressia, Fabian

Promoteur(s) : Monseur, Christian

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

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

Année académique : 2018-2019

URI/URL : <http://hdl.handle.net/2268.2/7932>

Avertissement à l'attention des usagers :

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

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



Faculté de Psychologie, Logopédie et Sciences de l'Éducation

**Opportunités d'apprentissage,
styles de réponse et
performance en culture
mathématique : analyse des
données de PISA 2012**

Mémoire présenté en vue de l'obtention du grade de Master en Sciences
de l'Éducation, à finalité spécialisée en Enseignement

Fabian PRESSIA

Promoteur : Christian Monseur

Lectrices : Lafontaine Dominique

Baye Ariane

Année académique 2018-2019

Je tiens tout d'abord à remercier mon promoteur, Monsieur Christian Monseur, pour sa disponibilité, ses conseils et son expertise. Merci aussi à lui pour le temps qu'il m'a consacré durant les deux années écoulées.

Je remercie également Mesdames Dominique Lafontaine et Ariane Baye pour l'intérêt qu'elles ont manifesté à ce travail en acceptant d'en être les lectrices.

Mes remerciements vont aussi à toutes les personnes, amis, collègues, membres de ma famille, qui m'ont encouragé depuis trois ans et m'ont permis d'aller jusqu'au bout.

Je tiens aussi à remercier tous les professeurs qui m'ont enseigné, les assistants qui m'ont accompagné et, bien sûr, les étudiants avec lesquels j'ai vécu cette aventure et qui sont devenus bien plus que de simples condisciples.

Enfin, merci à mes parents de n'avoir jamais douté de moi, depuis le début. Et surtout, merci à celui qui m'a supporté au quotidien durant ces trois années, et qui a tout rendu possible.



Faculté de Psychologie, Logopédie et Sciences de l'Éducation

**Opportunités d'apprentissage,
styles de réponse et
performance en culture
mathématique : analyse des
données de PISA 2012**

Mémoire présenté en vue de l'obtention du grade de Master en Sciences
de l'Éducation, à finalité spécialisée en Enseignement

Fabian PRESSIA

Promoteur : Christian Monseur

Lectrices : Lafontaine Dominique

Baye Ariane

Année académique 2018-2019

Table des matières

Introduction	1
Partie théorique	3
1 La culture mathématique en FWB	4
1.1 L'enquête PISA.....	4
1.1.1 Définition de la culture mathématique dans PISA 2012	5
1.1.2 Procédure d'échantillonnage de PISA 2012	9
1.1.3 Contenu des questionnaires PISA 2012.....	10
1.2 Résultats de PISA 2012 en FWB	12
2 Les opportunités d'apprentissage	15
2.1 Origine et définition des OTL.....	15
2.2 Mesure des OTL dans les enquêtes internationales	17
2.2.1 Les OTL dans PISA 2012.....	19
2.3 Élargissement du concept d'OTL	22
2.4 OTL et performance académique.....	28
3 Questionnaires contextuels et biais méthodologiques.....	33
3.1 Origine des styles de réponse.....	33
3.2 Conséquences des styles de réponse	35
3.3 Comment contourner les biais de styles de réponse ?	37
4 En conclusion	43
Partie pratique	44
1 Question de recherche et hypothèses.....	45
2 Méthodologie	47
2.1 Échantillon	47
2.2 Méthodes et instruments	48
2.2.1 Variables utilisées dans nos analyses	49
3 Analyse et interprétation des données de PISA 2012	50

3.1	Distribution des OTL en FWB.....	50
3.2	Familiarité avec des concepts mathématiques et niveau académique.....	52
3.3	Performance académique et styles de réponse.....	56
3.4	Modélisation factorielle des styles de réponse.....	62
3.4.1	Modèle 1	62
3.4.2	Modèle 2	65
3.4.3	Modèle 3	67
3.4.4	Modèle 4	76
4	Conclusions	83
5	Limites et perspectives	88
	Bibliographie.....	90
	Table des illustrations.....	93
	Annexes.....	I
	Annexe A : Illustration de la procédure de sélection des écoles avec pas d'échantillonnage. I	
	Annexe B : Syntaxe SAS de la partie 3.1	II
	Annexe C : Syntaxe SAS de la partie 3.2	VI
	Annexe D : Syntaxe SAS de la partie 3.3.....	VIII
	Annexe E : Syntaxe SAS de la partie 3.4, modèle 1	XII
	Annexe F : Syntaxe Mplus de la partie 3.4, modèle 1	XIII
	Annexe G : Syntaxe SAS de la partie 3.4, modèle 2	XIV
	Annexe H : Syntaxe Mplus de la partie 3.4, modèle 2	XV
	Annexe I : Syntaxe SAS de la partie 3.4, modèle 3, groupe 1	XVI
	Annexe J : Syntaxe Mplus de la partie 3.4, modèle 3, groupe 1	XXI
	Annexe K : Syntaxe Mplus de la partie 3.4, modèle 3, groupe 2.....	XXII
	Annexe L : Syntaxe Mplus de la partie 3.4, modèle 3, groupe 3	XXIII
	Annexe M : Syntaxe SAS du modèle 4 (Van de Gaer)	XXIV
	Annexe N : Syntaxe Mplus du modèle 4 (Van de Gaer)	XXXI

Annexe O : Corrélations entre MATHEFF et la performance en culture mathématique dans 37 systèmes éducatifs, sans et avec prise en compte des styles de réponse.	XXXII
Annexe P : Corrélations entre SCMAT et la performance en culture mathématique dans 37 systèmes éducatifs, sans et avec prise en compte des styles de réponse.	XXXIII

Introduction

Depuis maintenant plus de cinquante ans, les enquêtes à large échelle tentent de relever ce défi : mesurer la performance académique des élèves d'un grand nombre de systèmes éducatifs et comprendre les facteurs qui permettent d'influencer cette performance. Dans notre recherche, nous nous sommes intéressé plus spécifiquement au cycle 2012 du Programme International pour le Suivi des Acquis des élèves (PISA), qui avait pour domaine majeur la culture mathématique.

La raison de cet intérêt est double : un cadre conceptuel pour l'évaluation des mathématiques complètement redéfini, qui place la résolution de problèmes au centre des processus à évaluer (Demonty, Blondin, Matoul, Baye, & Lafontaine, 2013), d'une part et des questionnaires contextuels adressés aux élèves remaniés, qui évaluent de nouvelles variables relatives à la qualité de l'enseignement, d'autre part.

Parmi ces nouvelles variables introduites figurent les opportunités d'apprentissage (opportunity to learn – OTL). En effet, PISA 2012 est le premier cycle à en intégrer de façon explicite la mesure dans les questionnaires contextuels qui l'accompagnent (Schmidt, Zoido, & Cogan, 2014). Les OTL constituent un champ de recherche qui remonte aux années 1960, et qui a connu son heure de gloire dans les années 1990, avec les enquêtes internationales de l'IEA (*International Association for the Evaluation of Educational Achievement*). Elles permettent de mesurer si tous les élèves ont eu les mêmes chances d'apprendre, ont été exposés aux mêmes contenus, aux mêmes pratiques pédagogiques, et ont reçu le même temps d'apprentissage.

Les OTL posent donc des questions d'équité, car il ne paraît pas concevable d'évaluer de la même manière des élèves qui n'ont pas reçu les mêmes chances de développer leurs connaissances (Scherff & Piazza, 2008). Nous avons ainsi choisi de nous intéresser aux opportunités d'apprentissage en mathématiques en Fédération Wallonie-Bruxelles (FWB), à leur distribution parmi les élèves (que nous supposons inégale) et aux liens qui peuvent être établis entre ces OTL et le niveau de performance des élèves.

Mesurer de façon précise l'impact des OTL sur la performance des élèves nécessitera de prendre en considération les styles de réponse. Il s'agit d'un biais méthodologique qui s'observe lorsqu'on mesure des variables de contexte à l'aide, notamment, d'échelles de Likert, comme c'est le cas dans les enquêtes à large échelle. Ils apparaissent lorsque les répondants formulent

leurs réponses à cause d'un manque d'implication, de facteurs personnels ou culturels, plutôt que pour ce que la question prétend mesurer (Lafontaine, 2017). Ils constituent dès lors une source de pollution de la mesure et peuvent conduire, entre autres choses, au paradoxe attitude-performance. De nombreuses recherches se sont penchées sur les styles de réponse et ont apporté des pistes sur la façon de les quantifier et de les contrôler, afin d'améliorer la mesure des concepts sous-jacents. Nous nous sommes donc inspiré de ce qui a déjà été fait en ce sens afin d'essayer, à notre tour, de tenir compte des biais liés aux styles de réponse.

Notre objectif est donc d'améliorer la mesure existante des OTL en mathématiques au départ de l'enquête PISA 2012 en tenant les styles de réponse sous contrôle, afin de pouvoir tirer des conclusions valides sur les liens qui peuvent exister en Fédération Wallonie-Bruxelles entre les OTL et la performance en culture mathématique des élèves.

Le présent mémoire s'articule autour de deux parties : l'une théorique, l'autre pratique. La partie théorique consiste en une revue de la littérature dans laquelle nous présentons la méthodologie et les principaux résultats de l'enquête PISA 2012, en termes de performance et d'équité, pour la FWB, ainsi que l'état des lieux des connaissances en matière d'opportunités d'apprentissage et de styles de réponse. La partie pratique décrira, quant à elle, notre méthodologie, les différentes étapes par lesquelles nous sommes passé, les traitements statistiques que nous avons employés et les enseignements que nous pouvons en tirer à la lumière de notre question de recherche.

Partie théorique

1 La culture mathématique en FWB

1.1 L'enquête PISA

Le Programme International pour le Suivi des Acquis des élèves (*Programme for International Student Assessment – PISA*) est une enquête à large échelle conduite tous les trois ans depuis 2000 par l'OCDE (Organisation de Coopération et de Développement Économiques). Cette étude a pour objectif principal de mesurer, dans un nombre grandissant de systèmes éducatifs¹, les acquis des élèves de quinze ans en compréhension de l'écrit, culture mathématique et culture scientifique. Lors de chaque cycle PISA, un des trois domaines est spécifiquement ciblé. À PISA 2012, que nous avons exploitée dans le cadre de notre recherche, c'est la culture mathématique qui était le domaine majeur.

À la différence des enquêtes de l'IEA² (comme l'enquête PIRLS³ par exemple), qui basent le contenu de leurs questions sur l'étude des curriculums nationaux, et dont les experts extraient le plus grand dénominateur commun afin de rédiger les items, les enquêtes PISA posent un cadre de référence différent. Dans PISA, ce n'est pas le contenu des curriculums nationaux qui est évalué, mais les compétences que les jeunes de quinze ans devraient maîtriser afin de pouvoir s'insérer dans la société. En effet, dans de nombreux pays, l'âge de quinze ans correspond à la fin de la scolarité obligatoire. C'est donc à cet âge que les jeunes vont être amenés à poser des choix quant à leur avenir. Les experts PISA évaluent donc, de façon prospective, dans quelle mesure les jeunes de quinze ans sont aptes à entrer dans la vie active et à poser des choix conscients, indépendamment de leur niveau d'études ou de leur parcours antérieur (Demonty, Blondin, Matoul, Baye, & Lafontaine, 2013).

Chaque cycle de l'enquête PISA est le fruit d'un processus long, qui fait intervenir de nombreux experts nationaux et internationaux, ainsi que les gouvernements des pays participants. Le but de ce travail de longue haleine est de garantir la fidélité des items ainsi que la validité et la comparabilité des résultats obtenus. Le cycle PISA 2012 ne fait pas exception,

¹ 32 pays ont pris part à PISA 2000, 41 à PISA 2003, 57 à PISA 2006 (Baye, Quittre, Hindryckx, Fagnant, & Lafontaine, 2007). Pour PISA 2012, ce sont 65 pays et systèmes éducatifs qui ont pris part à l'enquête.

² International Association for the Evaluation of Educational Achievement, Association internationale pour l'Évaluation du Rendement scolaire.

³ Progress in International Reading Literacy Study, Programme international d'évaluation des compétences en lecture

puisque'il a été précédé d'un essai de terrain à large échelle, lequel a notamment permis de mesurer la fidélité des items soumis aux élèves et, le cas échéant de supprimer les items problématiques. Il a aussi permis d'établir la difficulté relative des questions portant sur les domaines cognitifs, étape essentielle pour permettre d'établir les scores de performance des élèves après l'implémentation de l'enquête.

1.1.1 Définition de la culture mathématique dans PISA 2012

Les cadres conceptuels relatifs à la compréhension de l'écrit et à la culture scientifique dans PISA 2012 sont respectivement identiques à ceux utilisés lors des cycles 2009 et 2006. À l'inverse, pour ce qui est de la culture mathématique, qui n'avait plus été domaine majeur depuis 2003, le cadre conceptuel a totalement été remanié, afin de mieux appréhender les processus par lesquels passent les élèves lorsqu'ils sont amenés à résoudre des problèmes, tout en conservant la comparabilité avec les résultats des cycles précédents. C'est ainsi que les experts responsables de l'élaboration de PISA 2012 ont redéfini la culture mathématique comme suit :

La culture mathématique est l'aptitude d'un individu à formuler, employer et interpréter des mathématiques dans un éventail de contextes, soit de se livrer à un raisonnement mathématique et d'utiliser des concepts, procédures, faits et outils mathématiques pour décrire, expliquer et prévoir des phénomènes. Elle aide les individus à comprendre le rôle que les mathématiques jouent dans le monde et à se comporter en citoyens constructifs, engagés et réfléchis, c'est-à-dire à poser des jugements et à prendre des décisions en toute connaissance de cause (OCDE, 2013, p. 27).

Cette définition laisse bien transparaître la vision prospective de l'évaluation PISA, puisque'elle insiste sur le fait que la culture mathématique doit préparer les individus à s'insérer dans la société et à poser des choix et des actes de façon consciente. Les trois verbes d'action (formuler, employer et interpréter) renvoient aux trois processus qui sont évalués au travers des questions posées dans PISA 2012.

- **Formuler** renvoie à l'idée « [d'] identifier des possibilités d'appliquer et d'utiliser les mathématiques » (OCDE, 2013, p 27). Il s'agit donc pour l'élève de comprendre l'utilité des mathématiques pour résoudre des situations concrètes et de traduire ces situations sous des formes qui en permettront une résolution mathématique.

- **Employer** les mathématiques correspond aux étapes de traitement, de calcul. L'élève doit donc mobiliser les outils et procédures qu'il connaît afin de dégager une solution à un problème mathématique. Ce processus recouvre également l'analyse de données mathématiques, telles que des diagrammes ou des graphiques.
- **Interpréter** consiste à se questionner sur la plausibilité d'une solution mathématique (l'évaluer) et à replacer une solution mathématique dans le cadre d'une situation problème donnée afin de dire si elle convient à cette situation.

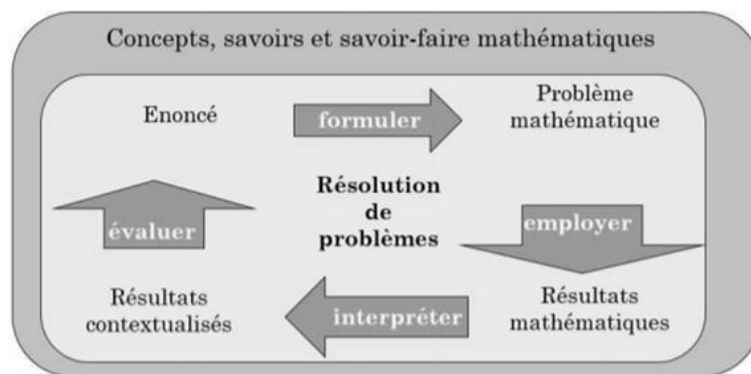


Figure 1 : cycle de modélisation mathématique dans PISA2012.

Ces trois processus, qui rappellent fortement le modèle de modélisation mathématique de Verschaffel (2000, 2008, 2010, cités par Fagnant & Demonty, 2018 ; Marcoux 2014) mettent la résolution de problème au centre des préoccupations de l'enquête PISA 2012 et insistent sur l'importance des étapes par lesquelles les élèves passent pour aboutir à une solution contextualisée au départ de données brutes. Il est en effet possible de faire un parallèle entre les processus PISA et le modèle de Verschaffel. Ainsi, *formuler* correspond aux étapes de compréhension et de modélisation, *employer* renvoie à l'analyse mathématique et *interpréter* recouvre tout à la fois les étapes d'interprétation, d'évaluation et de communication.

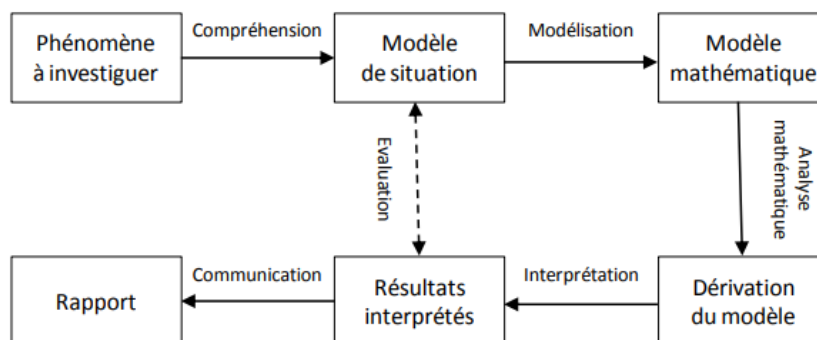


Figure 2 : diagramme de modélisation mathématique de Verschaffel.

Il est en outre intéressant de signaler que les processus mathématiques de modélisation seront toujours au centre du cycle PISA 2021, dans lequel la culture mathématique sera à nouveau domaine majeur. Les trois processus de PISA 2012 sont toujours présents, et encore plus proches de ceux de Verschaffel (interpréter a été renommé en interpréter et évaluer). S'ajoute un quatrième processus : raisonner, qui décrit la capacité des élèves à planifier les étapes de leur raisonnement en vue de résoudre un problème. On insistera donc davantage dans PISA 2021 sur l'importance de raisonner en vue de modéliser un problème sous une forme mathématique, de le résoudre pour apporter une solution et d'évaluer la pertinence de cette réponse en fonction de la situation de départ (Fagnant & Demonty, 2018).

En plus des trois processus, le cadre de référence PISA 2012 définit deux dimensions dans lesquelles s'inscrivent les problèmes mathématiques : le contexte et le contenu. Le contexte mathématique décrit l'éventail des situations de sa vie où un individu peut être confronté à un problème mathématique. Il peut être personnel (achats, jeux, transports, santé...), professionnel (devis, comptabilité, commande de matériaux...), sociétal (élections, démographie, publicité...) ou scientifique (climat, médecine, génétique...).

Le contenu mathématique explique le type de phénomène qui sous-tend le problème visé. Puisque l'enquête PISA définit la culture mathématique en termes de processus plutôt que de savoirs, les contenus s'éloignent des branches des mathématiques enseignées à l'école (géométrie, algèbre, arithmétique...). Cela s'explique en outre par le fait qu'un même phénomène mathématique peut être abordé sous différents angles, relevant de différentes branches des mathématiques. Les items sont donc regroupés en :

- « variations et relations » (relations provisoires ou permanentes entre des systèmes d'objets, des quantités ou des variables dans lesquelles les éléments s'influencent les uns les autres),
- « espace et formes » (propriétés, description, représentation et interactions avec des objets du réel, qui dépassent le cadre de la géométrie seule),
- « quantité » (raisonnement quantitatif, écritures multiples des nombres, argumentation basée sur la quantité),
- « incertitude et données » (probabilités et statistiques, variations dans des processus, prise de décisions malgré l'incertitude).

Enfin, le cadre de référence de PISA 2012 définit également sept facultés mathématiques fondamentales, qui sont autant de compétences que les individus activent lorsqu'ils réalisent une tâche mathématique. Ces facultés interviennent dans chacun des processus décrits. Leur activation est un bon indicateur de la compétence en mathématiques des individus, donc de la difficulté des items proposés. Ces sept facultés sont :

- la communication, soit la capacité des individus à extraire des données d'un problème ou d'une situation, à résumer et présenter leurs résultats durant le processus de résolution et à présenter leur solution sous quelque forme que ce soit, voire à la justifier,
- la mathématisation, qui recouvre les activités fondamentales de transposition d'éléments du monde réel en langage mathématique formel et l'interprétation de données mathématiques en fonction du réel,
- la représentation, aptitude à sélectionner, comprendre différentes représentations d'une même réalité et à passer de l'une à l'autre,
- le raisonnement et l'argumentation, qui permettent de réaliser des inférences, de justifier une solution ou de vérifier la véracité d'une justification,
- la conception de stratégies de résolution de problèmes, que l'on peut relier au concept de raisonnement décrit plus haut et qui « permet aux individus de sélectionner ou de concevoir une approche ou une stratégie permettant d'utiliser les mathématiques pour résoudre les problèmes qui se posent dans une tâche ou dans un contexte, mais aussi de guider sa mise en œuvre » (OCDE, 2013, p. 33),
- l'utilisation d'opérations et d'un langage symbolique, formel et technique, qui implique de connaître et de respecter la grammaire des mathématiques et d'appliquer correctement les règles qui régissent les opérations mathématiques que l'individu emploie,
- l'utilisation d'outils mathématiques, à savoir la connaissance des outils à disposition qui peuvent servir dans un contexte donné (instruments de mesure, calculatrice, ordinateur...).

1.1.2 Procédure d'échantillonnage de PISA 2012

Les responsables de PISA 2012 ont demandé aux experts de chaque pays participant de tirer un échantillon de 150 écoles au moins, au sein desquelles 35 élèves sont sélectionnés pour être soumis au questionnaire. En Belgique, ce sont 246 établissements qui ont pris part à PISA 2012, dont 110 en Fédération Wallonie-Bruxelles. La procédure d'échantillonnage dans le cadre de PISA est extrêmement complexe et représente plusieurs mois de travail, afin de s'assurer que les échantillons de chacun des pays soient valides et recouvrent bien l'entièreté de la population dont ils sont issus.

Tout d'abord, la totalité des établissements scolaires qui sont susceptibles d'accueillir des élèves de quinze ans (en FWB, les écoles secondaires, sur base de leur numéro FASE) sont classées selon un certain nombre de variables de stratification explicites. Chez nous, ces variables étaient «(1) le type d'enseignement (ordinaire/en alternance/spécialisé), (2) les niveaux d'études organisés dans l'établissement (1er degré uniquement/autres écoles), ainsi que (3) les filières proposées dans l'établissement (uniquement général et transition/mixtes/uniquement technique de qualification et professionnel)» (Demonty et al., 2013). Ces variables de stratification permettent de créer des sous-ensembles dans la population, au sein desquels vont être tirées les écoles cibles. Cela garantit de retrouver tous les types d'établissements de la population au sein de l'échantillon constitué. C'est également à cette étape qu'une pondération est attribuée aux élèves de chacun des groupes, afin que l'importance relative de ceux-ci soit la même que dans la population.

Une fois ces groupes constitués, les établissements au sein de chacun d'entre eux sont triés selon des variables de stratification implicites. En FWB, ces variables étaient le réseau d'enseignement, le taux de retard scolaire et le pourcentage de filles (uniquement dans le qualifiant). Cela permet par la suite de tirer aussi bien des établissements officiels que libres, des écoles avec de forts et de faibles taux de retard, et des écoles qualifiantes avec peu ou beaucoup de filles. Après cela, il convient de définir un pas d'échantillonnage, qui dépend du nombre total d'établissements, et du nombre qu'on souhaite en choisir. Il ne reste alors *plus qu'à* déterminer la première école sélectionnée sur base d'un nombre aléatoire et à suivre le pas

d'échantillonnage jusqu'à la fin de la liste des établissements pour avoir l'échantillon de premier ordre.⁴

Une fois les écoles sélectionnées, 35 élèves sont choisis de façon aléatoire parmi les élèves de quinze ans que compte chacun des établissements (ou moins si l'école ne comprend pas un minimum de 35 élèves de quinze ans). C'est ainsi que 3457 élèves ont été sélectionnés en Fédération Wallonie-Bruxelles pour participer à PISA 2012. Il y a donc une dépendance dans l'échantillon, puisque les élèves sélectionnés dépendent des écoles préalablement tirées. Il ne s'agit donc pas d'un échantillon aléatoire et simple. Les traitements statistiques sur l'échantillon PISA doivent donc être adaptés. En effet, les traitements statistiques classiques du logiciel SAS font le postulat du caractère aléatoire et simple de l'échantillon, ce qui conduit à sous-estimer l'erreur-type des paramètres calculés, donc biaise les résultats (avec un risque de rejeter une hypothèse nulle qui aurait dû être acceptée) (Monseur, 2017). Les traitements statistiques que nous avons utilisés dans la présente recherche tiennent dès lors compte de la dépendance qui existe dans l'échantillon PISA.

1.1.3 Contenu des questionnaires PISA 2012

Grâce à l'essai de terrain qui a préalablement eu lieu, il a été possible de mesurer, non seulement la difficulté des items, mais aussi le temps nécessaire aux élèves pour y répondre. Cela a permis de créer différentes formes du questionnaire cognitif (plan d'évaluation incomplet). Les items qui évaluent la culture mathématique, la culture scientifique et la compréhension de l'écrit ont été rassemblés en unités de questions (PM1 à 7 pour les mathématiques, PS1 à 3 pour les sciences et PR1 à 3 pour la lecture). Ces unités ont été groupées en treize carnets (*booklets*), d'une durée estimée pour les élèves de deux heures, qui sont autant de formes du questionnaire. Ces carnets ont été répartis de façon aléatoire entre les élèves qui ont pris part à l'enquête PISA. Le plan d'évaluation incomplet de PISA 2012 est détaillé dans le tableau suivant.

⁴ Pour plus d'informations, voir l'annexe A, tirée du rapport technique PISA 2012 (OCDE, 2014), qui donne un exemple fictif de procédure de sélection d'écoles sur base d'un pas d'échantillonnage.

Carnet	Unités de questions			
B1	PM5	PS3	PM6	PS2
B2	PS3	PR3	PM7	PR2
B3	PR3	PM6	PS1	PM3
B4	PM6	PM7	PR1	PM4
B5	PM7	PS1	PM1	PM5
B6	PM1	PM2	PR2	PM6
B7	PM2	PS2	PM3	PM7
B8	PS2	PR2	PM4	PS1
B9	PR2	PM3	PM5	PR1
B10	PM3	PM4	PS3	PM1
B11	PM4	PM5	PR3	PM2
B12	PS1	PR1	PM2	PS3
B13	PR1	PM1	PS2	PR3

Tableau 1 : design du plan d'évaluation incomplet de PISA 2012

Chaque unité de questions apparaît dans quatre carnets, qui comptent forcément quatre unités différentes chacun. Une même unité se retrouve tantôt en début, tantôt au milieu et tantôt à la fin d'un carnet, ce qui permet de supprimer un biais lié à l'effet de fatigue des élèves durant l'avancement du test, lequel diminue la performance moyenne des élèves et augmente la variance de cette performance (Monseur, 2017). Un certain nombre d'unités (PM1, PM2, PM3 pour les mathématiques) sont identiques à celles utilisées lors de PISA 2009. Ces items d'ancrage permettent de comparer les résultats sur certaines questions précises d'un cycle PISA à l'autre.

Concernant les questions portant sur la culture mathématique, elles ont été rédigées afin de couvrir les différents aspects de la culture mathématique, telle que définie dans PISA 2012. Ainsi, chacun des quatre contextes (voir supra) est couvert par 25 % des items, de même que chacun des quatre contenus. Pour ce qui est des processus, 25 % portent sur *formuler*, 50 % sur *employer* et 25 % sur *interpréter*. Environ un tiers des items sont des questions ouvertes, pour deux tiers de questions à choix multiples.

En plus des questionnaires cognitifs, les élèves sont soumis à un questionnaire contextuel, d'une durée de trente minutes environ, qui permet de recueillir de précieuses informations sur des variables telles que l'engagement vis-à-vis des mathématiques, la motivation, l'anxiété ou les opportunités d'apprentissage. Les directions des établissements sélectionnés répondent également à un questionnaire contextuel, qui permet de mesurer des variables relatives à l'école (informations administratives, climat scolaire, ressources financières...).

1.2 Résultats de PISA 2012 en FWB

Avec un score moyen de 493 points (SE=2,92), la Fédération Wallonie-Bruxelles ne diffère statistiquement pas de la moyenne de l'OCDE (494). En revanche, la moyenne de la FWB est significativement inférieure à celles obtenues par les Communautés flamande (531) et germanophone (511). Ce score de 493 est inférieur de 5 points à celui obtenu par la FWB en 2003, lorsque la culture mathématique était aussi le domaine majeur, mais cette différence n'est pas statistiquement significative (Demonty, et al., 2013). Il est également intéressant de noter que la FWB a réduit son écart-type sur la performance en mathématiques des élèves entre 2003 et 2012. Il est passé de 108 à 96. Cette réduction de l'écart entre les élèves les plus forts et les moins forts permet à la FWB de ne plus différer statistiquement de l'OCDE pour ce qui est de l'écart-type sur la performance en mathématiques. Cependant, comme le montre la figure 3 (Demonty et al., 2013, p. 20), il existe malgré tout en FWB de fortes disparités entre les élèves quant à leur performance en mathématiques, au regard de certaines variables contextuelles.

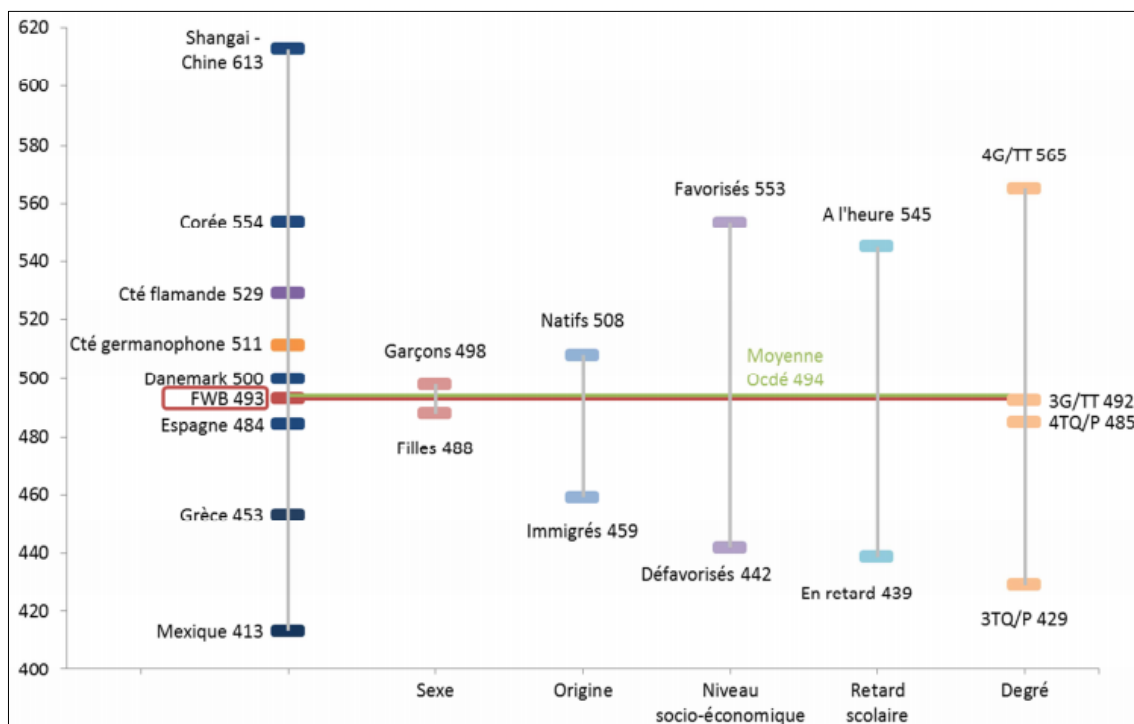


Figure 3 : différences de scores moyens en culture mathématique entre différentes catégories d'élèves à PISA 2012

Ainsi, bien qu'il existe une différence de performance en mathématiques de dix points en faveur des garçons, le genre de l'élève n'a finalement qu'un faible impact en comparaison de son statut socioéconomique ou de son statut d'immigration. Le retard scolaire, qui caractérise notre système éducatif, joue également un rôle important dans la performance des élèves : on observe 106 points d'écart entre les élèves à l'heure et les élèves en retard, en faveur des premiers. On remarque aussi que la moyenne des élèves à l'heure (545) est bien supérieure à celle de l'OCDÉ. Malheureusement, notre système éducatif se caractérise par un fort taux de redoublement. Au moment du test PISA 2012, comme à l'heure actuelle, un élève sur deux en troisième secondaire a connu le redoublement (FWB, 2013 ; FWB 2019). On comprend dès lors que la lutte contre le redoublement soit une des priorités du Pacte pour un Enseignement d'Excellence (FWB, 2017).

Les écarts se marquent également très fort en FWB entre les établissements scolaires. Il y a 181 points de différence sur l'échelle de performance en culture mathématique entre les 25 % d'écoles les plus performantes et les 25 % d'écoles les moins performantes. C'est près du double de certains autres pays (notamment la Norvège). Cela témoigne d'une forte variance inter-écoles, ce qui témoigne du caractère inégal de notre système éducatif. En effet, la variance inter-écoles à PISA 2012 est de 55,8 %, contre moins de 20 % en Finlande, en Islande ou au Danemark (Danhier, Jacobs, Devleeshouwer, Martin, & Alarcon, 2014).

De la même façon, le statut socioéconomique de l'élève est également un indicateur de l'iniquité de notre système éducatif. En FWB, 20,6 % de la variance observée sur la performance en mathématiques des élèves peut s'expliquer par leur ESCS⁵. Ce pourcentage de variance expliqué est non seulement statistiquement supérieur à celui de l'OCDÉ, mais il est plus de deux fois plus élevé que celui des pays nordiques, signe de l'important déterminisme social qui existe en Fédération Wallonie-Bruxelles.

⁵ Economic, social and cultural status.

En résumé...

PISA 2012 est le cinquième cycle de l'enquête à large échelle PISA menée par l'OCDE. Cette enquête a pour objectif de mesurer les compétences en culture mathématique, culture scientifique et compréhension de l'écrit des jeunes de quinze ans des pays de l'OCDE et d'une série de pays partenaires, avec une vision prospective. Les mathématiques étaient le domaine majeur de PISA 2012.

L'enquête PISA 2012 se caractérise par un cadre de référence totalement neuf pour la culture mathématique, qui la définit en termes de processus, contenus et contextes. Ce cadre place la résolution de problèmes au centre des préoccupations de l'enquête et s'intéresse aux étapes par lesquelles les élèves passent lorsqu'ils sont confrontés à ce type de tâches.

Les aspects techniques de l'enquête PISA 2012 sont le fruit d'un long travail, dirigés par des experts internationaux de renom, que ce soit pour la rédaction des questions, l'implémentation de l'essai de terrain, l'échantillonnage des écoles et élèves participants, le design d'évaluation incomplet... Cette rigueur technique qui caractérise les enquêtes de l'OCDE garantit la validité des questions, la représentativité des échantillons et la comparabilité des résultats obtenus.

Les résultats de la FWB à PISA 2012 sont, *a priori*, satisfaisants, puisque la performance moyenne des élèves de quinze ans, de même que l'écart qui existe chez nous entre les élèves les plus forts et les élèves les plus faibles se situent dans la moyenne des pays de l'OCDE. Cependant, une analyse plus fine des résultats révèle l'iniquité qui subsiste en FWB entre les élèves. Ainsi, le fait d'être à l'heure, le statut socioéconomique, l'école et la filière fréquentées sont, hélas, de puissants prédicteurs du niveau de performance d'un élève en FWB.

Dans cette recherche, nous nous sommes attaché à voir si les OTL en mathématiques peuvent également être reliées à la performance en culture mathématique des élèves, sur base des données issues de PISA 2012. Nous nous sommes aussi intéressé à la distribution des OTL en FWB, en particulier aux liens qui peuvent exister entre OTL et ESCS, d'une part et entre OTL et parcours scolaire, d'autre part.

2 Les opportunités d'apprentissage

2.1 Origine et définition des OTL

Comme l'expliquent Schmidt et Maier (2006), le concept d'opportunité d'apprentissage est loin d'être neuf. En effet, il fait écho aux travaux menés dans les années 1960 par le psychologue américain John B. Carroll, repris par la suite, entre autres, par Benjamin Bloom. Ce concept renvoie à une idée simple : « ce que les élèves apprennent à l'école dépend de ce qui est enseigné à l'école »⁶ [traduction libre] (Schmidt & Maier, 2006, p. 541). La première définition du concept d'opportunité d'apprentissage correspondait au temps attribué à une tâche, à une matière d'étude. En effet, dans ses travaux, Carroll (1984, cité par Schmidt & Maier, 2006) a essayé d'expliquer les variations d'aptitudes entre les individus ; son postulat était que le temps consacré à l'étude d'un sujet est le principal facteur qui permet d'expliquer pourquoi certaines personnes sont plus performantes que d'autres, aboutissant à la formule bien connue :

$$\text{Degré de Maîtrise} = f\left(\frac{\text{temps réellement consacré}}{\text{temps requis}}\right)$$

Dans la définition originelle, le temps réellement investi dans une tâche est donc la première opportunité d'apprentissage reconnue. Tout individu est donc capable de réussir n'importe quel apprentissage, pour autant qu'il y consacre le temps nécessaire, en fonction du temps que cet apprentissage requiert. Plus tard, dans son modèle élargi, Carroll a pris en considération des caractéristiques de l'individu ou du système dans le concept de temps requis (aptitude propre, compréhension, qualité de l'instruction...). L'idée de Carroll a par la suite été incorporée aux travaux de Bloom (1968, cité par Schmidt & Maier, 2006), qui, dans son modèle de la pédagogie de la maîtrise, soutient qu'avec suffisamment de temps, des objectifs clairement définis, un guidage approprié et un enseignement de qualité, n'importe quel élève peut arriver à réaliser des apprentissages de haut niveau cognitif (Crahay, 1999).

De là découle l'idée défendue par Bloom que « à partir du moment où les enseignants et des décideurs curriculaires sont à même de définir un critère de réussite approprié, il est de la

⁶ « What students learn in school is related to what is taught in school. »

responsabilité des enseignants et des écoles de donner le temps nécessaire pour que les élèves atteignent ce critère »⁷ [traduction libre] (Bloom, 1974, p. 683, cité par Schmidt & Maier, 2006). À cela s'ajoute la conception apportée par Berliner (1979, cité par Schmidt & Maier, 2006) selon laquelle le temps alloué n'est pas suffisant par lui-même, mais qu'il doit s'agir d'un temps durant lequel l'élève est effectivement engagé dans un curriculum en lien avec l'objectif fixé. C'est ainsi que les OTL ont commencé à glisser d'un concept uniquement théorique vers une piste d'amélioration de la qualité de l'enseignement.

Une définition plus récente des OTL a été donnée par Stevens, Wiltz, et Bailey (1998, cités par Abedi, Courtney, Leon, Kao, & Azzam, 2006) et reprise par Wang (1998). Selon eux, les OTL peuvent être décrites selon quatre variables qui reviennent le plus régulièrement dans la littérature scientifique :

- **la couverture du contenu** (*content coverage*), à savoir la proportion du curriculum prescrit (pour un grade ou une matière spécifique) que les élèves ont couverte ;
- **l'exposition au contenu** (*content exposure*) fait référence à ce que les élèves ont bien été exposés à un contenu spécifique, que ce contenu leur a bien été enseigné. Il s'agit donc principalement du temps que les enseignants ont consacré à l'enseignement de ce contenu lors de leurs leçons, rejoignant ainsi la définition originelle des OTL ;
- **l'emphase sur le contenu** (*content emphasis*) nous indique l'importance qui est accordée par l'enseignant à un contenu au sein des leçons dispensées aux élèves. Avec l'exposition au contenu, elle permet de distinguer le curriculum prescrit du curriculum réellement implémenté ;
- **la qualité de l'enseignement** (*quality of instruction delivery*) apparaît aussi comme une variable essentielle en termes d'opportunités d'apprentissage. L'idée étant que la façon dont les enseignants vont présenter un contenu aux élèves va influencer la manière dont ils vont comprendre ce qui leur est enseigné. Cette idée

⁷ « If teachers and curriculum makers can define an appropriate criterion of achievement, it then becomes the responsibility of the teachers and the schools to provide the time necessary for the students to attain the criterion. »

se retrouvera notamment dans la manière dont les OTL seront mesurées dans les enquêtes à large échelle (voir infra).

2.2 Mesure des OTL dans les enquêtes internationales

Avec l'apparition des premières enquêtes à large échelle en enseignement (First International Mathematics Study, FIMS, IEA, 1967 ; Second International Mathematics Study - SIMS, IEA, 1980), les OTL sont devenues pour l'IEA un indicateur de l'exposition de l'élève à un curriculum ou à un contenu. L'idée étant de savoir dans quelle mesure les performances de l'élève sont influencées par le fait qu'il ait, ou n'ait pas eu l'opportunité d'étudier le sujet visé par le test ou de s'exercer au type particulier de problème présenté. En particulier, lors de l'enquête SIMS, les opportunités d'apprentissage ont servi à mesurer les différences qui peuvent exister entre le curriculum prescrit, d'une part, et l'enseignement effectivement dispensé dans les classes, d'autre part (McDonnell, 1995).

Bien que la mesure des OTL dans SIMS ait posé quelques problèmes, notamment en termes de désirabilité sociale sur les pratiques déclarées, elle a permis de tirer des conclusions plus fines et plus approfondies sur les performances en mathématiques des élèves. Ainsi, il a notamment été possible d'observer dans quelle mesure le curriculum d'un système éducatif est en effet implémenté dans les classes, ou si, à l'inverse, certaines parties en sont moins enseignées. De la même façon, cela a permis d'étudier la façon dont les différents pays peuvent mettre l'accent sur telle ou telle composante des mathématiques dans leurs curriculums, et si ces composantes sont travaillées en classe de façon intensive ou si elles sont uniquement survolées. D'autre part, les OTL ont également renseigné les chercheurs sur la façon dont les systèmes scolaires peuvent être équitables ou non : au sein d'un même système, les élèves reçoivent-ils les mêmes chances d'apprendre ou y a-t-il des disparités entre élèves à qui l'on donne de meilleures occasions et ceux qui en obtiennent moins ?

Ainsi, dans les années 1980, les opportunités d'apprentissage ont constitué à la fois une mise en lumière d'un dysfonctionnement de certains systèmes éducatifs et une piste de solution possible. En effet, il est apparu clairement que certaines parties de curriculums sont moins mises en œuvre dans les classes et, surtout, qu'il existe entre les écoles des différences dans les opportunités d'apprentissage données aux élèves, desquelles découlent des différences de performances entre les élèves. Mais la solution est dans le problème : pour pallier ces différences de performances, il convient alors de donner à tous de meilleures occasions

d'apprendre. En particulier, « l'exposition au curriculum pourrait être un levier efficace dans les efforts visant à améliorer le rendement des élèves et à répartir plus équitablement les possibilités d'apprentissage »⁸ [traduction libre] (McDonnell, 1995, p. 308).

Dans les années 1990, en particulier à travers les enquêtes SMSO (Survey of Mathematical Opportunities) et TIMSS (Third International Mathematics and Science Study, IEA, 1995), les chercheurs ont essayé d'améliorer la manière dont le concept d'opportunités d'apprentissage peut décrire comment les élèves sont exposés à certains contenus, toujours dans un but d'amélioration de la validité de leurs résultats. Ainsi, pour ce qui est du curriculum prescrit, une mesure plus directe a été mise en place. Plutôt que de se baser sur des interprétations indirectes, il a été décidé d'aller chercher les informations à la source, auprès des pouvoirs politiques des systèmes éducatifs concernés. Il a donc été demandé à l'autorité en charge (généralement, le ministre de l'Éducation) de décrire précisément, sur base des directives officielles, quels contenus devaient être maîtrisés pour chaque grade d'enseignement. En outre, grâce à cela, les chercheurs ont pu améliorer la façon dont ils mesuraient le curriculum implémenté. Plutôt que de questionner les enseignants sur tous les aspects d'une discipline, ils ont pu cibler quelles matières de cette discipline sont abordées lors de chaque année d'étude, donc réduire le nombre d'items auxquels les enseignants devaient répondre et les formuler d'une façon plus adaptée en fonction du niveau auquel on se place.

À ce stade, les chercheurs se sont également intéressés à une nouvelle dimension de l'exposition curriculaire : le « curriculum potentiellement implémenté »⁹ [traduction libre] (Schmidt & Maier, 2006, p.545). Il s'agissait en fait de considérer les manuels scolaires et leur utilisation dans les classes comme des sources possibles d'OTL, lesquels peuvent réduire considérablement, ou au contraire, améliorer l'exposition des élèves à certains contenus et savoirs. Enfin, pour ce qui est du curriculum atteint, les chercheurs des enquêtes TIMSS 1995 et SMSO ont développé, en plus du score global de l'élève, les sous-scores qui reflètent leurs performances pour les contenus auxquels ils ont effectivement été exposés, sur base du curriculum prescrit.

⁸ « (...) curriculum exposure could be an effective lever in efforts to improve student achievement and to distribute learning opportunities more equitably. »

⁹ « potentially implemented curriculum. »

Avec les années 2000 sont arrivées les enquêtes PISA de l'OCDE. Cependant, il a fallu attendre le cycle PISA 2012 pour qu'une mesure des OTL soit pour la première fois introduite dans les questionnaires contextuels adressés aux élèves (Schmidt, Zoido, & Cogan, 2014). C'est justement sur base de ces questionnaires que nous avons établi notre recherche.

2.2.1 Les OTL dans PISA 2012

Dans la mesure où la culture mathématique dans PISA 2012 n'est pas tant définie en termes de contenus qu'en termes de compétences fondamentales, la définition des OTL dans cette enquête à large échelle s'écarte quelque peu de celles proposées jusque-là, notamment dans TIMSS 1995 (OCDE, 2013). C'est pourquoi, dans PISA 2012, sont mesurées des OTL de contenu (*OTL content*), des OTL liées aux pratiques d'enseignement (*OTL teaching practices*) et des OTL liées à la qualité de l'enseignement (*OTL teaching quality*) (OCDE, 2014).

Les opportunités d'apprentissage de contenu sont celles qui nous ont intéressé dans notre recherche. Elles sont mesurées dans le questionnaire contextuel adressé aux élèves grâce à trois séries de questions.

- La question ST61 (expérience avec des tâches mathématiques) interroge les élèves sur la fréquence à laquelle ils rencontrent différents types de tâches mathématiques durant le temps scolaire. Deux sous-échelles ont été créées à partir de cette question, aboutissant à deux variables : l'une liée aux tâches mathématiques pures (items e, g et i), l'autre aux tâches mathématiques appliquées (items a, b, c, d, f, g et h).

Q38 (ST61) À quelle fréquence avez-vous eu les types de tâches mathématiques suivantes à l'école ? (Cochez une case par ligne.)		Fréquemment	Parfois	Rarement	Jamais
a)	Utiliser un <horaire de trains> pour calculer combien de temps prendrait le trajet d'un endroit à un autre.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
b)	Calculer l'augmentation du prix d'un ordinateur après ajout de la taxe.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
c)	Calculer combien de mètres carrés de dalles il faut pour carreler un sol.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
d)	Comprendre des tableaux scientifiques présentés dans un article.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
e)	Résoudre une équation du type : $6x^2 + 5 = 29$.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
f)	Calculer la distance réelle entre deux endroits sur une carte à l'échelle 1/10 000.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
g)	Résoudre une équation du type : $2(x+3) = (x + 3) (x - 3)$.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
h)	Calculer la consommation hebdomadaire d'un appareil électrique.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
i)	Résoudre une équation du type : $3x + 5 = 17$.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

Figure 4 : items de la question ST61 — expérience avec des tâches mathématiques.

- La question ST62 (familiarité avec des concepts mathématiques) interroge les élèves sur treize concepts mathématiques. Pour chacun d’eux, il leur est demandé de déterminer à quel point ils sont familiers de ces concepts, sur une échelle de Likert allant de « je n’en ai jamais entendu parler » à « je connais et comprends le concept ». L’intérêt particulier de cette question est que trois des treize items (d, j et l) sont de faux concepts mathématiques. Il s’agit dès lors d’items pièges, qui permettent de capter la tendance de certains élèves à recourir à l’acquiescement ou de répondre par désirabilité sociale, soit d’être influencés par un biais de styles de réponse (voir infra). C’est pour cette raison que la partie pratique de notre recherche sera basée sur la question ST62. Deux indices ont été dérivés de cette question dans la base de données PISA 2012. L’indice FAMCON (*familiarity with mathematics concepts*) reprend les treize items ST62 tandis que FAMCONC (*familiarity with mathematics concepts – corrected for overclaiming*) soustrait au précédent les trois items pièges.

Q39		Pensez aux concepts mathématiques. Dans quelle mesure les termes suivants vous sont-ils familiers ?				
(ST62)		<i>(Cochez une case par ligne.)</i>				
		Je n'en ai jamais entendu parler	J'en ai entendu parler une fois ou deux	J'en ai entendu parler quelques fois	J'en ai souvent entendu parler	Je connais et comprends le concept
a)	Fonction exponentielle.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
b)	Diviseur.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
c)	Fonction du second degré.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
d)	<Nombre propre>.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
e)	Équation linéaire.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
f)	Vecteurs.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
g)	Nombre complexe.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
h)	Nombre rationnel.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
i)	Radicaux.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
j)	<Échelle subjonctive>.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
k)	Polygone.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
l)	<Fraction déclarative>.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
m)	Figure isométrique.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
n)	Cosinus.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
o)	Moyenne arithmétique.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
p)	Probabilité.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Figure 5 : items de la question ST62 — familiarité avec des concepts mathématiques.

- Les questions ST73 à ST76 (expérience de différents types de problèmes mathématiques vus à l'école) proposent aux élèves des types de tâches qu'ils auraient pu rencontrer à l'école. ST73 illustre deux problèmes mathématiques (*Algebraic Word Problem*), ST74 relève de tâches procédurales, touchant plutôt la pensée algébrique (*Procedural Task*). ST75 montre des tâches de raisonnement mathématique pures (*Pure Math Reasoning*) tandis que ST76 illustre également des tâches de raisonnement mathématique, mais contextualisées. Pour chaque type de tâche, il est demandé aux élèves de se positionner deux fois : une fois pour exprimer à quelle fréquence ils ont rencontré ce type de tâches lors des cours de mathématiques et une fois pour faire de même, mais dans le cadre des évaluations. Cela a permis la création de huit indices (par type de tâche et par situation).

Q45 (ST74)		Voici des exemples d'un autre ensemble de compétences mathématiques. 1) Résoudre $2x + 3 = 7$. 2) Trouver le volume d'une boîte dont les côtés mesurent 3 m, 4 m et 5 m.			
		Nous voulons savoir si vous avez eu affaire à ces types de problèmes à l'école. Il est inutile de résoudre ces problèmes ! (Cochez une case par ligne.)			
		Fréquemment	Parfois	Rarement	Jamais
a)	Avez-vous vu ces types de problèmes en cours de mathématiques ?	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b)	Avez-vous eu ces types de problèmes lors des évaluations à l'école ?	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Figure 6 : items de la question ST74 — expérience de différents types de problèmes mathématiques, tâches procédurales.

Les opportunités d'apprentissage liées aux pratiques d'enseignement ont été mesurées grâce à des items adaptés de l'enquête TALIS (*Teaching and Learning International Survey*, enquête internationale sur l'enseignement et l'apprentissage) de l'OCDE. Dans cette étude, il était demandé aux enseignants de déclarer à quelle fréquence ils réalisaient un certain nombre de pratiques de classe (pratiques structurantes, pratiques axées sur les élèves, pratiques de renforcement) (OCDE, 2013). Ces items ont été adaptés aux élèves et d'autres, spécifiques à l'enseignement des mathématiques, ont été ajoutés. En résulte une série de treize items (ST79) qui mesurent la façon dont les élèves perçoivent les pratiques de classe de leur enseignant.

c)	Le professeur donne des travaux différents aux élèves qui ont des difficultés d'apprentissage ou à ceux qui progressent plus vite.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
d)	Le professeur donne des travaux de recherche qui prennent au moins une semaine de travail.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e)	Le professeur me dit si je travaille bien en mathématiques.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f)	Le professeur nous pose des questions pour s'assurer que nous avons compris le contenu enseigné.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Figure 7 : exemple d'items de la question ST79 — comportement de l'enseignant

Enfin, les opportunités d'apprentissage liées à la qualité de l'enseignement ont été mesurées en se basant sur les trois piliers du modèle tridimensionnel de Klieme, également décrits par Pianta et Hamre (2009, cité par OCDE, 2014), à savoir une gestion claire et structurée de la classe, un climat de classe soutenant et tourné vers l'élève et une forte activation cognitive (Klieme, Pauli, & Reusser, 2009). Les deux premières dimensions du modèle ont été mesurées grâce aux indices DISCLIMA (climat disciplinaire, items ST81) et TEACHSUP (support de l'enseignant, items ST77). Ces indices existaient déjà tous deux dans les versions antérieures de PISA. En revanche, la mesure de l'activation cognitive était jusque-là inédite dans les questionnaires PISA. Elle s'est faite par le biais de la question ST80, composée de neuf items, qui a permis de construire l'indice COGACT (*cognitive activation*).

Q50 (ST80)		Pensez au professeur de mathématiques qui vous a donné votre dernier cours de mathématiques. À quelle fréquence chacune des situations suivantes se présente-t-elle ? (Cochez une case par ligne.)			
		Toujours ou presque toujours	Souvent	Parfois	Jamais ou presque jamais
a)	Le professeur pose des questions qui nous font réfléchir au problème.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b)	Le professeur nous donne des problèmes qui nécessitent une longue réflexion de notre part.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
c)	Le professeur nous demande d'élaborer nos propres procédures pour résoudre des problèmes complexes.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
d)	Le professeur nous donne des problèmes pour lesquels la méthode de résolution n'apparaît pas immédiatement.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e)	Le professeur nous donne des problèmes dans différents contextes afin que nous vérifiions que nous avons compris les concepts.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f)	Le professeur nous aide à tirer les leçons de nos erreurs.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
g)	Le professeur nous demande d'expliquer comment nous avons résolu un problème.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
h)	Le professeur nous présente des problèmes qui nous amènent à appliquer dans de nouveaux contextes ce que nous avons appris.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
i)	Le professeur nous donne des problèmes qui peuvent être résolus de différentes manières.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Figure 8 : items de la question ST80 — activation cognitive.

Une recherche menée sur 42 pays ayant participé à PISA 2012 (Schmidt et al., 2014) a permis de conclure à un lien significatif entre les OTL mesurées dans cette enquête et la performance en mathématiques, pour une majorité de pays. Ce lien a pu être établi au niveau du pays, de l'école et de l'élève.

2.3 Élargissement du concept d'OTL

Parallèlement à leur développement comme outil de recherche, les opportunités d'apprentissage ont, dans la deuxième moitié des années 1980, également commencé à être utilisées comme des indicateurs d'enseignement (« OTL as an Education Indicator », McDonnell, 1995, p. 309), en lien avec plusieurs études américaines de mesure des processus

scolaires et de classe. Avant cela, les grandes collectes de données en éducation s'intéressaient surtout aux extrants de l'enseignement (le niveau des élèves) et à des éléments tels que les dépenses par élève. Le recueil d'informations quant à l'exposition curriculaire ou à la façon dont les contenus sont enseignés ne se faisait alors que sur des échantillons réduits, de façon non systématique. L'idée de cette diversification dans la collecte de données en enseignement était d'une part, de pouvoir comparer les résultats des élèves à travers le temps ou d'un environnement à un autre et, d'autre part, de servir de base pour la prise de décisions au niveau politique, afin de comprendre et de reproduire les situations favorables aux apprentissages des élèves.

De ce tournant dans l'utilisation des OTL a logiquement résulté un élargissement du concept en lui-même. À la « simple » idée d'exposition au curriculum, telle qu'elle est utilisée dans l'enquête SIMS, est venue s'ajouter celle que la façon dont ce contenu est présenté et par qui il est présenté doivent aussi jouer un rôle déterminant dans la manière dont les élèves peuvent améliorer leurs apprentissages. Pourtant, bien que les recherches menées à cette époque commencent à s'intéresser aux qualifications des enseignants et à la manière dont les écoles organisent leurs cours, aucune donnée n'est encore recueillie en ce qui concerne les stratégies pédagogiques mises en place dans les classes. D'autre part, ces recherches en lien avec le concept se limitaient presque toujours aux domaines de l'enseignement des mathématiques et à celui des sciences. Notons également que les données recueillies à ce stade posent des problèmes par deux aspects. Tout d'abord, pour ce qui est de l'exposition au curriculum et des priorités des écoles, les chercheurs se basaient sur les intitulés des cours et sur le nombre d'élèves inscrits à ces cours. Or, vu l'importante variation qui peut exister dans la couverture du sujet au sein d'un même cours, d'une école à l'autre, cet indicateur est peu révélateur des réelles OTL qui sont données aux élèves. Ensuite, les données recueillies dans les écoles, même à un niveau national, étaient collectées par le biais de questionnaires adressés aux enseignants (méthode peu coûteuse en temps et en argent). Mais cela ne permet pas de réellement mesurer la complexité des pratiques curriculaires, comme l'aurait permis l'observation directe des pratiques d'enseignement dans les classes.

D'indicateurs d'enseignement, les OTL ont aussi, dans les années 1990, commencé à servir d'instrument politique aux États-Unis («OTL as a Policy Instrument», McDonnell,

1995, p. 311). En effet, en 1992, le Conseil national sur les Normes et les Tests pédagogiques¹⁰ a émis un rapport qui a contribué à placer les opportunités d'apprentissage au centre des préoccupations politiques en enseignement. Bien que le concept ait été renommé dans le rapport en « *school delivery standards* », la définition que le NCEST en donne fait clairement référence aux OTL¹¹. L'idée générale du rapport était qu'il convenait d'établir, au niveau national, un ensemble de normes de rendement, qui décriraient ce que les élèves doivent savoir, le niveau qu'ils doivent atteindre, et que les établissements scolaires sont responsables de leur fournir les opportunités d'arriver à ces normes, en leur donnant notamment accès à des enseignants bien formés, du matériel pédagogique approprié et des installations adéquates.

Cette préoccupation de fournir à chaque élève des OTL qui lui permettent d'atteindre les objectifs fixés par le système fait écho à une recherche menée par Gau (1997, cité par Abedi et al., 2006). En effet, Gau a étudié les OTL, leur distribution et leurs effets sur la performance en mathématiques des élèves. Dans le concept d'OTL, Gau a intégré les connaissances des enseignants, le contenu et le niveau de l'enseignement et les ressources matérielles en mathématiques de l'école. Les résultats de la recherche montrèrent que les OTL sont positivement liées à la performance en mathématiques, mais aussi que la distribution des OTL entre les catégories d'écoles est inégale.

Dans les années qui suivirent, les décisions politiques mises en place aux États-Unis allèrent dans le sens des recommandations du rapport du NCEST pour aboutir à la réforme Goals 2000, qui redéfinit clairement les opportunités d'apprentissage comme « les critères et la base de l'évaluation de la suffisance ou de la qualité des ressources, des pratiques et les conditions nécessaires à chaque niveau du système éducatif (écoles, agences éducatives locales et États) pour donner aux élèves l'occasion d'apprendre la matière dans les normes volontaires de contenu national ou normes de contenu de l'État »¹² [traduction libre] (McDonnell, 1995, p. 312). Cette définition, qui dépasse l'exposition curriculaire définie dans l'enquête SIMS, fait

¹⁰ National Council on Education Standards and Testing (NCEST)

¹¹ « School delivery standards should provide a metric for determining whether a school “delivers” to students the “opportunity to learn” well the material in the content standards » (NCEST, 1992, p. E-5, cité par McDonnell, 1995).

¹² « (...) the criteria for, and the basis of, assessing the sufficiency or quality of the resources, practices, and conditions necessary at each level of the education system (schools, local educational agencies, and States) to provide all students with an opportunity to learn the material in voluntary national content standards or State content standards ».

des OTL un enjeu majeur pour permettre à tous les élèves d'atteindre les objectifs que le système fixe pour eux.

Une recherche menée en 2004 par l'Université du Michigan (Study of Instructional Improvement - SII) a permis de mettre en lumière les différences d'opportunités d'apprentissage qui peuvent exister, non pas entre les écoles, mais entre élèves d'une même classe. En effet, pour mener leur étude, les chercheurs, plutôt que de s'intéresser à ce qui est enseigné à la classe dans son ensemble, ont demandé à des enseignants de rapporter par écrit, à plusieurs moments de l'année, les pratiques d'enseignement dispensées et l'exposition curriculaire pour huit élèves cibles de leur classe. Cette étude s'est intéressée à la fois aux mathématiques et à la langue maternelle, dans des classes de grades 4 et 5. En plus de leurs notes écrites, il a été demandé aux enseignants participants de classer le curriculum de cours (divisé en plusieurs composantes) selon deux dimensions : l'exposition (la composante a-t-elle fait l'objet ou non d'un enseignement) et, pour la langue maternelle, la difficulté des compétences visées (plus une compétence a été enseignée souvent, plus elle est jugée comme facile). En mathématiques, la deuxième composante n'était pas la difficulté des compétences visées, mais le type de méthodes utilisées par l'enseignant (enseignement direct, exercices...).

La validité des résultats de la mesure des OTL sur base des carnets de bord des enseignants a fait l'objet de nombreuses études, qui se sont attachées à comparer les carnets de bord avec des observations par des tiers, les carnets de bord avec des enquêtes transversales et des enquêtes transversales avec des observations par des tiers.

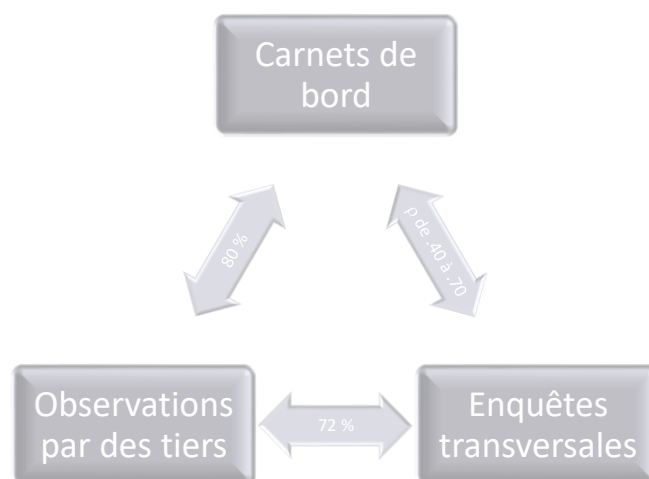


Figure 9 : taux de correspondance et corrélation entre les différentes sources de données de l'étude SII 2004.

Globalement, le lien qui existe entre les traces écrites des enseignants dans leurs carnets de bord et les observations réalisées par des tiers est très fort, avec dans la plupart des cas, plus de 80% de concordance entre les deux sources de données. (Camburn & Barnes, 2004, cités par Schmidt & Maier, 2006) De la même façon, il existe une forte corrélation (supérieure à 0,70) entre les données récoltées par une étude à large échelle transversale et celles issues de déclarations quotidiennes des enseignants. (Smithson & Porter, 1994, cités par Schmidt & Maier, 2006) Cependant, cette corrélation, tout comme la concordance entre carnets de bord et observations, diminue lorsque les questions posées vont plus dans le détail (on arrive alors à des corrélations de l'ordre de 0,40). Pour ce qui est de la comparaison entre études à large échelle et observations, la recherche est plus discrète. Une recherche (Burstein et al., 1995, cités par Schmidt & Maier, 2006) compare cependant les données d'une étude à large échelle avec celles recueillies par l'analyse de matériels de classes (cahiers d'exercices par exemple). Le taux de correspondance est assez élevé (72%), ce qui va dans le sens de la validité des différents types de mesures.

Nous retenons de la recherche SII l'importance de définir clairement le cadre de contenu pour mesurer les OTL, et de ne pas s'arrêter aux intitulés de cours, qui peuvent parfois fausser la mesure. En effet, « les titres des cours ne sont que des indicateurs génériques, et capturent probablement plus la nature des filières dans ce pays que les particularités des opportunités d'apprentissage »¹³ [traduction libre] (Schmidt & Maier, 2006, p.550).

Une recherche menée auprès de 1801 élèves de Floride (Scherff & Piazza, 2008) pose la question des OTL (ici, en langue maternelle) en termes d'équité. Cette recherche, qui s'appuyait sur les déclarations des élèves (sur base des questionnaires à items fermés et de commentaires écrits libres), a amené les auteures à conclure qu'il est « irresponsable et répréhensible de demander à chaque étudiant de répondre aux mêmes normes lors de tests à enjeux élevés lorsqu'il ne dispose pas des ressources nécessaires pour réussir »¹⁴ [traduction libre] (Scherff & Piazza, 2008, p. 350).

¹³ « (...) course titles are only generic indicators, and probably capture more the nature of tracking in this country the particulars of opportunity to learn. »

¹⁴ « (...) it is irresponsible and reprehensible to hold every student accountable to the same standards on high-stakes tests when they are not provided with the necessary resources to enable them to be successful. »

Les données récoltées dans cette recherche ont notamment permis de mettre en lumière le rôle des filières d'enseignement, mais aussi des politiques éducatives de reddition de compte dans la distribution inégale des OTL. Les filières d'enseignement, comme on peut s'y attendre, jouent un rôle de facilitateur ou de limitateur des opportunités d'apprentissage données aux élèves en compréhension de l'écrit. Il ressort de l'étude que les élèves des filières supérieures en Floride (*Advanced Placement et International Baccalaureate*) lisent des textes plus riches et plus diversifiés que ceux des filières inférieures (*General and Special Education*), les amenant ainsi à développer davantage de compétences en littérature.

Concernant les politiques éducatives, les chercheuses ont mis en lumière un lien entre l'année d'étude (*grade*) et les OTL. La raison en est que l'état de Floride organise aux grades 9 et 10 des tests à hauts enjeux (*Florida Writes!* et *Florida Comprehensive Achievement Test*, organisés par le *Florida Department of Education*), lesquels aboutissent à un système de notation des établissements scolaires, avec des répercussions en termes de subsides, de curriculums et une politique de reddition de compte pour les établissements les moins performants. Or, les données récoltées par les auteures montrent justement qu'aux grades 9 et 10, les élèves avaient justement moins accès à des activités de rédaction riches, à du matériel didactique diversifié et à des supports de lecture variés (autant d'OTL). Le temps d'apprentissage était à ces grades en grande partie consacré à la préparation des tests de l'état, par le biais de matériel de préparation (fourni par l'état) ou de livres d'exercices (*workbooks*).

En conclusion, Scherff et Piazza (2008) insistent sur l'importance de replacer les OTL au centre du débat concernant l'apprentissage de la langue maternelle (ici, la langue anglaise). En effet, il paraît impensable d'évaluer tous les étudiants sur les mêmes bases en dépit de leurs disparités propres, sans chercher à leur donner les moyens de se développer au maximum de leurs capacités. Selon les auteures, « jusqu'à ce que nous considérions certaines des questions sous-jacentes en matière d'OTL, l'inégalité continuera à être omniprésente dans les systèmes scolaires américains et il y aura des écarts entre les étudiants riches et pauvres, les étudiants blancs et les étudiants de couleur, les étudiants surperformants et les étudiants avec de faibles capacités »¹⁵ [traduction libre] (Scherff & Piazza, 2008, p. 350).

¹⁵ « However, until we consider some of the underlying issues in regard to OTL, inequality will continue to be pervasive in U.S. school systems and there will be discrepancies between rich and poor students, white students and students of color, and students of high abilities and low abilities. »

2.4 OTL et performance académique

Pour ce qui est des liens qu'il est possible d'établir entre les OTL et la performance des élèves, de nombreuses études ont permis d'établir ce lien de façon répétée, du moins, dans les domaines mathématiques et scientifiques. (Brophy & Good, 1986 ; Floden, 2002, cités par Schmidt & Maier, 2006) À titre d'exemple, citons les enquêtes menées par l'IEA. Ainsi, à travers l'enquête FIMS, il a été possible de conclure à l'existence significative d'un lien entre les opportunités d'apprentissage et le niveau académique des élèves, bien que variable et parfois ténu. Les corrélations obtenues entre les OTL et la performance des élèves variaient de 0,16 à 0,32 lorsqu'on prenait en considération l'ensemble des élèves, mais elles ont pu monter jusqu'à 0,80 lorsque les chercheurs ont comparé les pays entre eux, ce qui leur a permis de conclure que de plus grandes opportunités d'apprentissage pour les élèves étaient liées à une meilleure réussite.

Avec un système de mesure des OTL plus détaillé et un nombre de pays participants plus élevé, l'enquête SIMS a, elle aussi, permis de déduire des corrélations intrapays faibles ou inexistantes. Les chercheurs ont émis l'hypothèse que ces faibles corrélations sont en partie dues à une forme d'homogénéité imposée par les curriculums : les pays où il y a le moins de variations dans l'exposition curriculaire étant ceux où le lien établi était le plus faible. De la même façon, les données SIMS ont permis d'établir des corrélations bien plus consistantes entre les pays qu'à l'intérieur de ceux-ci (jusqu'à 0,57). De plus, il a été noté à partir des données SIMS que les pays pour lesquels le curriculum se concentrait sur un domaine particulier plutôt que survoler le curriculum voyaient leur performance plus élevée dans ce domaine particulier. Les chercheurs en ont conclu que « donner plus de contenu à plus d'étudiants produit plus de gains » (Kifer & Burstein, 1992, p. 337, cités par Schmidt & Maier, 2006).

Exactement de la même façon, les résultats de l'enquête TIMSS appuient, voire amplifient, ceux issus de FIMS et SIMS. À nouveau, les chercheurs ont pu établir un lien significatif entre OTL et niveau des élèves entre les pays, avec de grosses variations dans les opportunités données aux élèves d'un pays à l'autre. De plus, dans l'enquête TIMSS, comme dit précédemment, une importante collecte de données quant aux curriculums a été menée, avec des différences entre le curriculum prescrit et le curriculum implémenté. Or, ces différences se sont révélées être significativement reliées à des gains de performance. (Schmidt et al., 2001, cités par Schmidt & Maier, 2006) Pour ce qui est des effets des OTL dans les pays, ils ont pu cette fois être mieux établis que lors des deux enquêtes précédentes (effet significatif dans

24 pays sur 29, avec un R^2 variant de 0 à 0,67), sans doute grâce à la plus grande importance accordée aux données curriculaires.

Des résultats similaires sont observés concernant les sciences (notamment à travers la First IEA Science Study), mais avec des résultats plus consistants à l'intérieur des pays. Cependant, concernant l'apprentissage de la langue maternelle, il n'y a que peu de résultats de recherche disponibles, et encore moins qui réussissent à établir un lien significatif entre OTL et performance des élèves. Plus surprenant, certaines recherches ont même conclu à des corrélations négatives entre les opportunités d'apprentissage et la performance en lecture des élèves entre les pays, mais à des corrélations positives à l'intérieur des pays, car les items qui mesurent ces opportunités induisent un biais en fonction de la culture du répondant. (Lundberg, 1994, cité par Lafontaine, Baye, Vieluf, & Monseur, 2015)

Ces résultats décevants peuvent être expliqués de plusieurs façons. Une première piste tient au fait que la mesure des OTL se base sur des faits autorapportés, ce qui engendre des biais méthodologiques. Une deuxième piste serait que le fait de tenir sous contrôle le ESCS des élèves conduit à sous-estimer l'importance des opportunités d'apprentissage, à cause de l'existence d'un effet joint de ces deux variables sur la performance des élèves. Enfin, une troisième piste réside dans l'idée que l'apprentissage de la lecture est un processus long, résultant de l'accumulation de plusieurs années d'enseignement-apprentissage. En conséquence, mesurer l'impact des OTL par une enquête transversale, en demandant à l'élève de se positionner sur son seul enseignant actuel, n'est pas un fonctionnement optimal pour saisir la complexité du phénomène, d'autant plus que des preuves tangibles de l'effet positif à long terme de certaines stratégies sur la compréhension de l'écrit des élèves existent dans la littérature scientifique.

Sur base des données PISA 2012, Schmidt et al. (2014) ont étudié les relations qui existent entre les OTL et la performance en mathématiques. Ils ont basé leur recherche sur trois indices OTL dérivés des questionnaires contextuels PISA. Le premier, « mathématiques formelles » (*formal mathematics*) décrit la mesure dans laquelle les élèves ont été exposés à des concepts mathématiques liés à l'algèbre et à la géométrie durant les cours de mathématiques. « Problèmes » (*word problems*) indique si les élèves ont eu l'occasion de s'exercer à ce type d'exercices en classe et « mathématiques appliquées » (*applied mathematics*) mesure si les élèves ont eu l'occasion de réaliser des exercices en contexte lors des cours dispensés.

Sur les trois indices, la Belgique obtient des scores respectifs de 1,8 (SE=0,0), 1,9 (SE=0,0) 1,9 (SE=0,0). La Belgique obtient donc un score statistiquement supérieur à celui de l'OCDE pour l'indice « mathématiques formelles », mais n'en diffère pas pour les deux autres. Une analyse de régression sur 41 pays a permis aux auteurs de conclure à une relation linéaire entre les indices OTL de problèmes et de mathématiques appliquées. Cependant, la relation entre l'indice OTL de mathématiques formelles et la performance de ces pays est de nature quadratique.

À l'intérieur des pays, tant au niveau de l'école que de l'élève, Schmidt et al. (2014) ont conclu à l'existence d'une relation statistiquement significative entre les indices OTL et la performance en mathématiques. En Belgique, cette relation n'est significative que pour l'indice « mathématiques formelles » et est linéaire. Les auteurs ont considéré la Belgique dans son ensemble. Nous ne disposons donc pas dans cette étude de données relatives à la FWB. La Belgique est également un des pays où la différence d'exposition aux OTL en fonction du ESCS de l'école est la plus forte, ce qui pourrait renforcer les inégalités dues à cet indice.

En 2015, Lafontaine et al. ont mené une recherche dont le but était de montrer dans quelle mesure les OTL auxquelles les élèves sont exposés varient en fonction de l'école qu'ils fréquentent, d'une part, et comment ces OTL peuvent influencer sur la performance en compréhension de l'écrit de ces élèves, d'autre part. Pour la mener, les auteurs se sont penchés sur les données de l'enquête PISA 2009. En effet, cette année-là, le thème majeur était la compréhension de l'écrit.

Comme le soulignent les auteurs, un problème majeur dans la mesure des OTL en compréhension de l'écrit est que cette dernière ne fait pas l'objet d'un cours spécifique, mais se travaille de façon transversale, à travers les différents cours suivis par l'élève (ce qui est d'autant plus vrai dans l'enseignement secondaire). Dès lors, il sera moins évident de mesurer les opportunités d'apprentissage dans ce domaine que ce n'est le cas en mathématiques et en sciences. Une première façon de procéder consiste à comparer les questions de l'épreuve et les matériels utilisés à ceux qui figurent dans le cours de l'élève. Cela revient à se demander si le contenu visé a bien été couvert par l'éducation qu'a reçue l'élève. Une autre manière de faire, plus proche de la notion précédemment évoquée d'exposition curriculaire, est de vérifier dans quelle mesure les élèves ont pu apprendre les processus visés par les questions de l'épreuve.

La recherche menée par Lafontaine et al. (2015) pose la question des opportunités d'apprentissage en termes d'équité, ce qui se rapproche dans une certaine mesure de la

perspective envisagée par Scherff et Piazza (2008). Dans cette notion d'équité (qui renvoie également aux notions d'égalité des chances, d'égalité de traitement et d'égalité des acquis, mais aussi à celle de discrimination négative, neutre ou positive) s'affrontent deux points de vue : l'un méritocratique, l'autre correctif. (Perrenoud, 1995, cité par Lafontaine, 2016) Si l'on se place sous un regard méritocratique, la réflexion en matière d'OTL sera de considérer qu'il est de la responsabilité de l'élève de capter un maximum d'opportunités d'apprendre dans ce que l'école lui propose et que cette dernière a pour rôle d'offrir exactement la même chose à chacun (stricte égalité de traitement). À l'inverse, selon un principe de justice corrective, il convient plutôt de combler les difficultés des élèves ou les inégalités qui existent entre eux en leur apportant davantage ; il est donc du devoir de l'institution scolaire d'offrir plus à ceux qui ont moins, pour garantir à tous les mêmes chances de réussite (égalité des acquis).

Dans tous les cas, peu importe le type d'égalité prôné, il semble inéquitable qu'un système éducatif reproduise, voire amplifie les inégalités sociales. C'est pourquoi dans leur recherche, Lafontaine et al. (2015) ont considéré qu'un système était d'autant plus équitable que la variance interécole dans la distribution des OTL serait faible et que le lien entre le contexte social des écoles et la répartition des OTL serait proche de zéro, voire négatif (ce qui témoignerait de l'existence d'une politique compensatoire). Cette idée d'équité dans les opportunités d'apprentissage renvoie directement à leur rôle d'instrument politique (McDonnell, 1995), tel qu'expliqué précédemment.

La recherche menée par Lafontaine et al. (2015), bien qu'elle présente certaines limites (en particulier le fait de ne pas pouvoir tirer de conclusions au niveau de la classe, donc de l'acte enseignant), a permis d'établir de façon claire qu'il existe un lien positif et significatif entre les opportunités d'apprentissage offertes aux élèves et leurs performances en compréhension de l'écrit et que les disparités qui existent dans certains systèmes quant à l'offre d'OTL en lecture sont une source potentielle de renforcement des inégalités entre les élèves.

En résumé...

Les opportunités d'apprentissage constituent un sujet d'étude depuis plus de cinquante ans. Leur définition originelle de temps consacré à l'apprentissage d'un sujet s'est vue élargie et complexifiée au fil du temps pour inclure, notamment des variables relatives à la qualité de l'enseignement.

L'histoire des OTL est étroitement liée à celle des enquêtes à large échelle, notamment avec les études de l'IEA menées dans les années 1980 et 1990. Concernant les enquêtes PISA, c'est en 2012 pour la première fois que les OTL ont été mesurées par le biais des questionnaires contextuels adressés aux élèves. Cela a été fait en termes de contenu, de pratiques d'enseignement et de qualité de l'instruction.

De nombreuses études ont investigué les OTL au fil des années, avec des résultats parfois contradictoires. Il en ressort généralement un lien positif entre l'exposition des élèves à de plus grandes opportunités d'apprentissage et leur niveau académique.

Dès lors, l'exposition à de mêmes chances d'apprendre peut constituer un indicateur de l'équité qui existe dans un système éducatif. Il semble en effet inéquitable que des élèves ne reçoivent pas les mêmes OTL sur base de leur origine, de leur niveau socioéconomique ou de l'école qu'ils fréquentent. Au contraire, de plus grandes OTL dans des milieux défavorisés pourraient même être le signe d'une justice compensatoire dans un système.

Dans cette recherche, nous nous sommes intéressé à un aspect particulier des OTL en mathématiques, le concept de familiarité avec des concepts mathématiques (question ST62). Nous avons étudié les liens qui existent entre cette mesure des OTL et la performance en mathématiques des élèves en FWB, ainsi que la distribution de ces OTL en fonction des caractéristiques des élèves.

3 Questionnaires contextuels et biais méthodologiques

La récolte de données contextuelles dans les enquêtes à large échelle n'est pas neuve puisqu'elle caractérisait déjà les premières enquêtes de l'IEA conduites dans les années 1970. Cependant, l'importance et le crédit accordés aux questionnaires contextuels n'ont fait qu'augmenter au fil des années. À titre d'exemple, il est possible de citer les enquêtes PISA. Entre 2000 et 2015, dans PISA, les questionnaires contextuels « passent d'un statut quasi confidentiel à la pleine consécration » (Lafontaine, 2017, p. 27). Depuis 2009, ils font l'objet d'un chapitre indépendant dans le cadre de référence PISA, alors qu'avant ce cycle, une seule page leur était consacrée et que les travaux relatifs à leur élaboration n'existaient que sous forme de documents de travail. Les cycles 2012 et 2015 de PISA vont encore plus loin, puisque les experts en charge des questionnaires contextuels s'attachent, lors de ces cycles, à mesurer de nouvelles variables relatives à la qualité de l'enseignement (notamment les OTL) et à garantir la validité et la comparabilité internationale des résultats.

Cette volonté d'amélioration de la mesure des variables de contexte dans les enquêtes internationales ne peut qu'être saluée, mais elle s'accompagne d'un certain nombre de défis méthodologiques qu'il convient de relever pour garantir la validité de cette mesure. En particulier, la manière dont les questions sont adressées aux élèves, enseignants, parents, directeurs dans ces questionnaires (sous forme d'échelle de Likert) engendre des biais liés aux spécificités des répondants : les styles de réponse.

3.1 Origine des styles de réponse

Toute mesure en éducatrice s'accompagne d'une certaine forme d'erreur de mesure, de biais. Il peut s'agir d'une erreur d'échantillonnage, liée à la façon de sélectionner les individus, d'une erreur de mesure au niveau individuel, liée à la nature de l'instrument employé, ou d'une erreur d'ancrage, liée à un mauvais choix d'items pour comparer les résultats au fil du temps (Monseur, 2016). Les styles de réponse (*response styles*) sont du second type. En effet, ils sont dus à la manière dont sont collectées les données contextuelles dans les enquêtes internationales, c'est-à-dire sous forme d'échelle de Likert. Ce type d'échelle induit chez le répondant un biais qui l'amène à répondre, non pas sur la base de ce que la question prétend mesurer, mais à cause de caractéristiques personnelles ou culturelles qui lui sont propres. La mesure du *construct* visé par l'item s'en trouve dès lors biaisée.

Les styles de réponse peuvent être définis comme « la probabilité qu'un répondant sélectionne un sous-ensemble spécifique de catégories de réponse »¹⁶ (Weijters, 2006) ou « des tendances systématiques et stables dans la manière de répondre qui ne s'expliquent pas par le contenu d'une question ni par ce que celle-ci vise à mesurer » (Yang, Harkness, Chin, & Villar, 2010, cités par Lafontaine, 2017, p. 28). Ils sont donc bien dus à des caractéristiques individuelles (sexe, traits de caractère) ou culturelles du répondant, plutôt qu'au *construct* visé par la question, qu'ils viennent dès lors polluer. Ils peuvent aussi s'expliquer par un manque d'engagement des répondants dans le questionnaire, qui répondent uniformément aux items d'un même bloc (comportement satisfaisant) ou par un manque de maîtrise de la langue des items.

Les principaux styles de réponse connus sont :

- l'acquiescement (*Acquiescence Response Style*) : la tendance à faire un usage disproportionné de catégories de réponses allant dans le sens de l'accord ou de l'approbation, peu importe le contenu de la question ;
- le désacquiescement (*Disacquiescence Response Style*) : la tendance à faire un usage disproportionné de catégories de réponses allant dans le sens du désaccord ou de la désapprobation, peu importe le contenu de la question ;
- le choix des extrêmes (*Extrême Response Style*) : la tendance à faire un usage disproportionné des réponses situées aux extrémités de l'échelle de Likert ;
- le choix des médians (*Midpoint Response Style*) : la tendance à faire un usage disproportionné des réponses médianes sur l'échelle de Likert ;
- la désirabilité sociale, qui consiste à répondre, non pas sur ce qu'on pense réellement, mais en fonction des normes de la société, ou des attentes sociales supposées par le répondant ;
- le choix des réponses au hasard.

¹⁶ « the probability that a respondent selects a specific subset of response categories »

3.2 Conséquences des styles de réponse

La conséquence évidente des biais de styles de réponse est que les résultats obtenus en utilisant des échelles de Likert ou des mesures autorapportées vont être difficilement comparables d'un pays à l'autre. Ainsi, pour une même échelle, les répondants des pays asiatiques auront tendance à répondre plutôt dans les catégories médianes alors que les personnes issues de pays occidentaux, marqués par une culture plus individualiste, répondront davantage dans les extrêmes (Grisay, 2009). De la même manière, les ressortissants du sud de l'Europe sont plus marqués par l'acquiescement que ceux du Nord, tout comme les répondants venant des États-Unis. Pour garantir la comparabilité des résultats d'un pays à l'autre (équivalence cross-culturelle), les enquêtes à large échelle recourent à des procédures qui testent l'invariance des *constructs*, indépendamment des caractéristiques culturelles des répondants.

Il existe trois types d'invariance qui garantissent l'équivalence cross-culturelle : l'invariance configurale (qui garantit que le *construct* mesuré a la même configuration factorielle d'un pays à l'autre), l'invariance métrique (qui garantit que le *construct* a la même unité de mesure dans chaque pays) et l'invariance scalaire (en plus d'une même unité de mesure, les pays partagent sur le *construct* une origine commune). La première invariance est nécessaire afin d'affirmer qu'il s'agit bien d'un seul *construct* dans les différents pays. Elle est généralement atteinte grâce à la rigueur des cadres conceptuels dans les enquêtes à large échelle et aux essais de terrain qui la vérifient. La seconde invariance doit être atteinte afin qu'il soit valide de comparer des corrélations sur le *construct* visé à l'échelle internationale. Enfin, la dernière invariance, si elle est attestée, permet de comparer les moyennes du *construct* d'un pays à l'autre. Cette dernière forme d'invariance n'est que rarement atteinte dans les enquêtes à large échelle, car les procédures statistiques utilisées pour la calculer sont particulièrement sensibles aux échantillons inégaux entre les pays (Van de Vijver, & He, 2014, cités par Lafontaine, 2017).

Une autre conséquence bien documentée dans la littérature scientifique, et qui a d'ailleurs mis les chercheurs sur la piste des styles de réponse, est le paradoxe attitudes-performances (*attitude-achievement paradox*). Il s'agit d'une situation dans laquelle on observe des corrélations de signes opposés, à l'intérieur des pays et entre ceux-ci, entre un *construct* donné et la performance académique des sujets. Il semblerait en effet logique qu'un *construct* corrélé positivement avec la performance à l'intérieur des pays le soit également lorsqu'on compare

les pays entre eux, et inversement. Une recherche récente (Lafontaine, Dupont, Jaegers, & Schillings, 2019) a illustré ce paradoxe sur base des données issues de PIRLS 2011. Comme le montre la figure 10, les auteures ont mis en évidence une corrélation négative entre le concept de soi en lecture et la performance en compréhension de l'écrit des élèves (corrélation moyenne de -0,20 pour les pays de l'OCDE). Pourtant, à l'intérieur des pays, les corrélations sont toutes positives et avoisinent généralement les 0.30, voire 0,40.

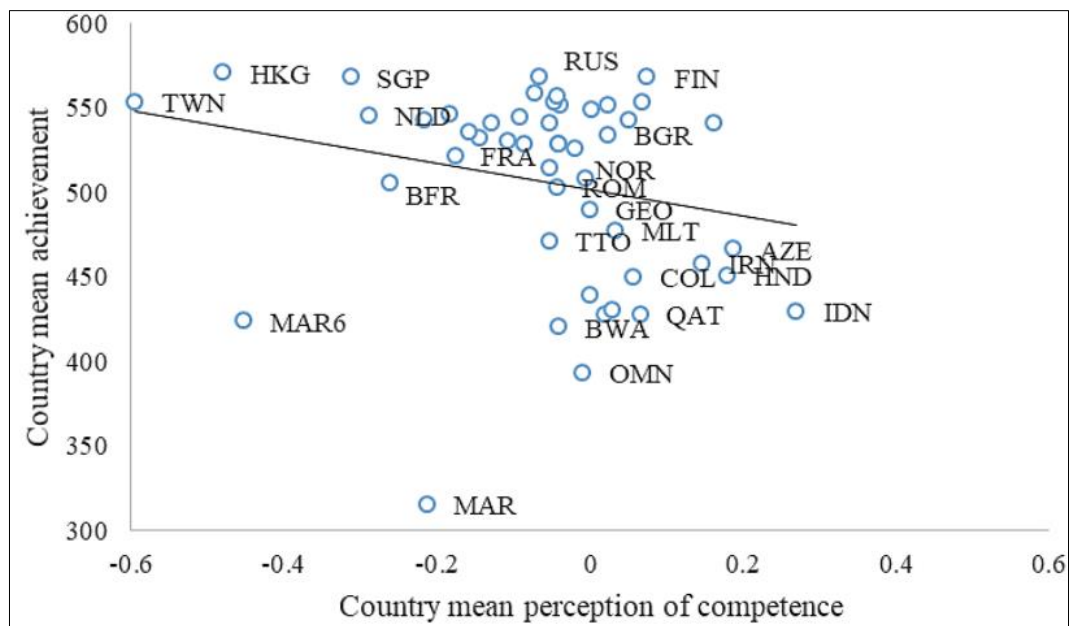


Figure 10 : corrélation entre le concept de soi et les performances en lecture au niveau pays à PIRLS 2011.

Ce paradoxe s'explique par les styles de réponse. En effet, parmi les plus performants en compréhension de l'écrit, on retrouve Singapour, Taïwan, Hong Kong. Or, ces pays, comme dit précédemment, sont justement ceux où les élèves ont tendance à la modestie, qui auront donc un concept de soi faible. À l'inverse, dans les pays les moins performants (Indonésie, Honduras, Azerbaïdjan, Iran, Colombie, Qatar...) les élèves ont tendance à se surestimer, ce qui biaise les résultats et explique les corrélations négatives observées au niveau international. Il est aussi intéressant de noter que, dans leur recherche, Lafontaine et al. (2019) ont mis en évidence qu'il n'existait pas de paradoxe attitude-performance pour le pendant négatif du concept de soi, à savoir la perception de la difficulté (corrélations négatives tant entre les pays qu'à l'intérieur), ce qui soulève des questions sur la manière dont la conception des échelles peut influencer les styles de réponse.

3.3 Comment contourner les biais de styles de réponse ?

De nombreuses recherches se sont attachées, par différentes procédures, à quantifier les biais de styles de réponse afin de les contourner et d'améliorer la mesure des *constructs* sous-jacents. En dresser la liste exhaustive n'est pas possible dans le cadre du présent mémoire, mais nous avons sélectionné plusieurs d'entre elles dont nous allons faire la synthèse.

À partir de données issues de PISA 2003, Lie et Turmo (2005, cités par Lafontaine, 2017), ont eu recours aux analyses factorielles de composant principal (*principal component analysis*) pour modéliser les styles de réponse. Ils ont forcé un ensemble de dix variables contextuelles à saturer sur un seul et même facteur, qu'ils ont nommé « *superconstruct* ». Ce facteur commun reflète, selon les auteurs, la tendance des répondants à donner des réponses formatées, soit le biais de styles de réponse. Un score élevé pour un pays à ce *superconstruct* témoigne d'une tendance plus fortement marquée pour les élèves de ce pays à répondre de façon stéréotypée. Les auteurs ont mis en évidence une corrélation négative entre le niveau de performance des pays et leur score sur le *superconstruct*, ce qui fournit une piste d'explication au paradoxe attitude-performance (Grisay, 2009).

Le *superconstruct* a ensuite été utilisé comme variable de contrôle pour ajuster les scores obtenus sur une série de variables contextuelles, ce qui a permis de calculer de nouvelles corrélations, bien moins négatives, voire positives, entre ces variables et la performance académique, au niveau international. Il est aussi remarquable que Lie et Turmo (2005, cités par Grisay, 2009) aient obtenu des résultats presque identiques au départ de données PISA 2000. Les auteurs ont obtenu une corrélation de 0,89 sur les indicateurs calculés entre ces deux cycles, pour les pays qui ont pris part aux deux. De plus les pays avec les scores de *superconstruct* extrêmes se sont révélés être les mêmes lors des deux cycles. Les scores les plus élevés ont notamment été observés en Turquie, Indonésie, Mexique, Bulgarie, Brésil, Thaïlande et Tunisie, tandis que le Japon, la Corée, les Pays-Bas et les pays scandinaves ont eu les scores les plus faibles au *superconstruct*, tant sur base des données issues de PISA 2000 que de PISA 2003. Ces derniers pays seraient donc ceux qui sont les moins marqués par les biais de styles de réponse.

Aletta Grisay (2009) a répliqué la recherche conduite par Lie et Turmo, mais en partant des données du cycle 2006 de PISA. Comme eux, elle a basé sa recherche sur dix variables contextuelles, qu'elle a fait saturer sur un composant principal, comme le montre la figure 11.

Ce facteur, qui expliquait à lui seul 78% de la variance sur les variables choisies, est donc le *superconstruct*, que Grisay assimile spécifiquement à la désirabilité sociale. Comme dans la recherche précédente, des scores faibles ont été observés au Japon, en Corée, aux Pays-Bas et dans les cinq pays scandinaves. À l'autre bout de l'échelle, on retrouve des pays similaires à ceux pointés précédemment, à quelques nuances près.

(N: 57 countries)

		Factor1
SCIEFUT	Future-oriented science motivation PISA 2006 (WLE)	0.94593
PERSSCIE	Personal value of science PISA 2006 (WLE)	0.93773
JOYSCIE	Enjoyment of science PISA 2006 (WLE)	0.92789
SCIEACT	Science activities PISA 2006 (WLE)	0.92270
INTSCIE	General interest in learning science PISA 2006 (WLE)	0.91811
PV1INTR	Plausible value in interest in science	0.90332
INSTSCIE	Instrumental motivation in science PISA 2006 (WLE)	0.89618
GENSCIE	General value of science PISA 2006 (WLE)	0.83633
PV1SUPP	Plausible value in support for scientific inquiry	0.77648
SCSCIE	Science self-concept PISA 2006 (WLE)	0.75424
Variance Explained : 78,2%		

Figure 11 : analyse de composant principal sur dix variables contextuelles de PISA 2006.

De la même manière que Lie et Turmo, Grisay (2009) a ensuite utilisé le *superconstruct* pour ajuster les corrélations entre plusieurs variables et la performance académique, au niveau des pays. Il en ressort que de nombreuses corrélations très négatives se voient réduites sur de nombreuses variables, comme le concept de soi en sciences qui, avec des items comme « je suis bon en sciences », induit une forte comparaison au groupe de référence de l'élève (*big-fish-little-pond-effect*, *BFLPE*) (Lafontaine, 2017). En revanche, l'indice d'auto-efficacité en sciences, qui pose des questions en termes d'aptitudes spécifiques (je sais faire telle ou telle tâche), et qui induit bien moins de comparaison sociale, n'était dans la recherche de Grisay que faiblement corrélé avec le *superconstruct*. Le changement de corrélation observé en ajustant cet indice avec le *superconstruct* est d'ailleurs négligeable.

Van de Gaer, Grisay, Schulz et Gebhardt (2012), toujours sur base des données de PISA 2006, se sont intéressés plus spécifiquement à l'influence du BFLPE sur les mesures réalisées grâce aux questionnaires contextuels. Ils ont investigué les relations qui existent à différents niveaux de mesure entre la performance académique des élèves et leur concept de soi en sciences. Les auteurs ont mis en évidence l'existence d'un paradoxe attitude-performance sur ledit concept. Les corrélations mesurées au sein des pays, au niveau de l'élève sont positives (sauf exception) et très variables, tandis que la corrélation à l'échelle internationale est fortement négative ($\rho=-0,74$).

En recourant à une analyse de régression multiniveaux, Van de Gaer et al. (2012) ont mis en lumière la relation entre la performance en sciences et le concept de soi en sciences. Toute chose étant égale par ailleurs, les élèves plus performants ont une meilleure image d'eux. Cependant, les auteurs ont révélé une relation négative entre l'année d'étude (*grade*) et le concept de soi. Les élèves dans les années d'études inférieures (donc en retard), qui sont avec des élèves plus jeunes et un programme moins poussé, ont, lorsqu'on tient les autres variables contextuelles sous contrôle, un meilleur concept de soi en sciences que ceux dans les classes les plus avancées (qui sont donc à l'heure, avec des pairs de meilleur niveau). Cela témoigne bien de l'existence d'un BFLPE sur le concept de soi en sciences. De la même manière, plus l'école fréquentée a une moyenne générale faible, ou plus elle a un ESCS faible, plus les élèves déclarent de hauts concepts d'eux-mêmes.

Van de Gaer (2010), en utilisant des données PISA 2003, a également modélisé les styles de réponse afin de donner une réponse au paradoxe attitude-performance. Pour ce faire, elle a travaillé sur un sous-ensemble de onze pays, choisis pour leurs scores élevés, moyens ou faibles sur l'échelle de performance en mathématiques et sur celle d'intérêt pour les mathématiques.

- La Tunisie, l'Indonésie et le Brésil sont caractérisés par une performance moyenne faible et un haut niveau moyen d'intérêt pour les mathématiques.
- La Corée, le Japon, Hong Kong et la Finlande, à l'inverse, ont une performance moyenne élevée, mais un faible intérêt pour les mathématiques.
- Enfin, l'Australie, la France, l'Allemagne et l'Irlande montrent des niveaux moyens sur les deux échelles.

Ensuite, grâce au logiciel Mplus, les données ont été analysées via une analyse factorielle confirmatoire. Trois *constructs* issus des questionnaires contextuels ont été inclus dans le modèle : l'intérêt pour les mathématiques (INTMAT), le concept de soi en mathématiques (SCMAT) et le sentiment d'auto-efficacité en mathématiques (MATHEFF). Le modèle prévoyait quatre variables latentes, comme le montre la figure 12 :

- la variable latente INTMAT qui explique les quatre items s'y rapportant (ST30),

- la variable latente SCMATH¹⁷, qui explique les cinq items ST32,
- la variable MATHEFF, qui explique les huit items ST31,
- une variable latente RESPONSE BIAS, qui explique, conformément aux études précédemment citées, les items relatifs à INTMAT et SCMATH, mais pas ceux qui expliquent MATHEFF.

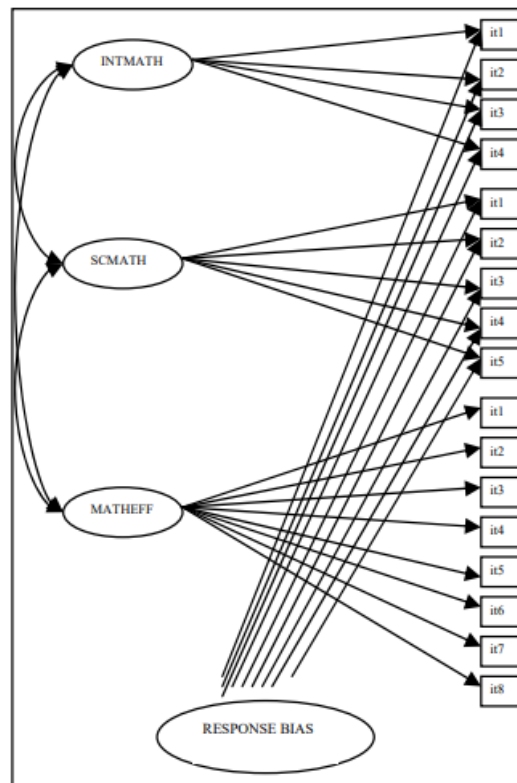


Figure 12 : structure factorielle utilisée par Van de Gaer (2010).

Cette modélisation a permis d'augmenter légèrement les corrélations entre les trois variables contextuelles et la performance en mathématiques à l'intérieur des pays. La variable pour laquelle la différence est la plus marquée (donc qui est la plus entachée par le biais de styles de réponse) est l'intérêt pour les mathématiques. Entre les pays, les corrélations qui étaient largement négatives entre INTMAT et la performance en mathématiques d'une part, et entre SCMAT et cette même performance d'autre part, deviennent toutes deux positives. À nouveau, c'est INTMAT qui est la variable la plus fortement marquée par les *response styles*.

¹⁷ L'orthographe de la variable SCMAT de PISA 2003 a été modifiée en SCMATH dans l'article de Van de Gaer (2010).

Étonnamment, la corrélation à l'échelle internationale entre MATHEFF et la performance passe de 0,37 à 0,02 lorsqu'on prend en compte le biais de styles de réponse. Il est aussi intéressant de noter que la prise en compte de ce biais permet d'améliorer de façon consistante les corrélations qui existent entre les trois indices étudiés. La corrélation entre MATHEFF et INTMAT passe ainsi de -0,11 à 0,77.

En 2015, Lafontaine, Baye, Vieluf, et Monseur ont mené une recherche qui croise les styles de réponse avec les opportunités d'apprentissage, sur base des données issues de PISA 2009. Afin de mener à bien leur recherche et d'apporter des réponses à leurs questions, les auteurs ont dû modéliser le concept d'OTL. Pour ce faire, ils ont sélectionné un ensemble de 17 items du questionnaire contextuel élève, qui décrivent les types de supports utilisés lors des cours de langue maternelle et les types de processus qu'ils ont eu l'occasion d'y exercer. Onze items furent retenus, six ont été rejetés par une analyse confirmatoire.

Le concept d'OTL a alors été modélisé en trois dimensions, pour chacune desquelles cinq valeurs plausibles ont été tirées pour les élèves issus des 34 pays de l'OCDE qui composent l'échantillon. La première dimension constituait le concept d'OTL dans son ensemble et incluait les onze items. Cette dimension servirait de témoin pour vérifier s'il existe un biais lié aux styles de réponse. La seconde dimension comprenait trois items et relevait de la lecture de textes non-continus. Quant à la troisième, constituée de quatre items, elle reflétait la mesure dans laquelle les élèves lisaient des œuvres de fiction et devaient poser des interprétations ou des évaluations sur ces œuvres (processus de haut niveau). C'est par cette décomposition du concept d'OTL que les chercheurs ont pu contourner le biais lié aux styles de réponse des élèves.

Cette modélisation a permis aux auteurs de conclure à des corrélations plus élevées sur les deux dimensions des OTL (fiction et textes non continus), que ce soit au niveau élève, école ou pays, lorsqu'on prend en compte les styles de réponse. À ce dernier niveau en particulier, la corrélation avec la dimension OTL fiction, qui était négative avant la prise en considération des styles de réponse (paradoxe attitude-performance) devient positive. Quant à la dimension « styles de réponse », elle est négativement corrélée avec la performance en compréhension de l'écrit, à tous les niveaux de mesure, ce qui appuie les conclusions des études précédemment citées : ce sont les élèves les plus faibles qui sont le plus fortement marqués par les biais de styles de réponse.

En résumé...

Les questionnaires contextuels qui accompagnent les enquêtes internationales constituent une source d'informations essentielle, mais sont à prendre avec recul et une méthodologie adaptée pour en tirer des conclusions valides.

Au cours des dernières années, une attention grandissante a été accordée aux questionnaires contextuels dans les enquêtes PISA, avec de nouvelles variables mesurées, un cadre conceptuel mieux défini et un souci pour l'équivalence cross-culturelle, prise en compte par la mesure des invariances configurale, métrique et scalaire.

Cependant, le format des items dans les questionnaires contextuels, à savoir les échelles de Likert, engendre des biais méthodologiques liés aux spécificités des répondants : les styles de réponse, qui polluent les mesures des *constructs* que prétendent mesurer ces items.

Ces styles de réponse s'expriment différemment d'un pays à l'autre, en fonction de la culture de ce pays. Ils peuvent aussi trouver leur origine dans des caractéristiques propres à l'individu. Ils conduisent dans de nombreux cas à tirer des conclusions erronées lorsqu'on compare des variables de contexte avec la performance académique au niveau international (paradoxe attitude-performance).

De nombreuses études ont, au départ de données issues d'enquêtes internationales, étudié la façon dont il est possible de contourner ces styles de réponse. Elles diffèrent quant à leur méthodologie, mais recourent généralement à des modélisations factorielles qui permettent de calculer des indices de styles de réponse, lesquels sont par la suite utilisés comme variables de contrôle.

Les études menées dans le domaine vont dans le même sens : la prise en compte des biais de styles de réponse améliore la mesure des variables de contexte et permet, dans de nombreux cas, de dépasser le paradoxe attitude-performance.

Dans notre recherche, nous essaierons également d'améliorer la mesure d'une variable contextuelle en prenant en compte les styles de réponse : les OTL en mathématiques.

4 En conclusion

À la croisée entre les opportunités d'apprentissage, la performance en culture mathématique et les biais de styles de réponse se trouve une opportunité pour la recherche. En effet, il s'agit de domaines largement documentés dans la littérature scientifique lorsqu'ils sont pris indépendamment ou deux à deux, mais il n'existe, à notre connaissance, aucune recherche qui permette de croiser les trois champs à ce jour. Cela est d'autant plus vrai si l'on s'intéresse à la Fédération Wallonie-Bruxelles.

C'est dans cette idée que s'inscrit le présent mémoire. Notre objectif est d'expliquer comment les opportunités d'apprentissage en mathématiques peuvent être reliées à la performance des élèves de la FWB, tout en essayant d'améliorer la mesure de ces OTL, au départ du questionnaire contextuel PISA 2012.

Les opportunités d'apprentissage ont, depuis longtemps été corrélées positivement avec la performance académique des élèves. C'est également le cas avec les données de PISA 2012 (Schmidt et al., 2014). Cependant, nous ne disposons pas de conclusions spécifiques à la Fédération Wallonie-Bruxelles qui vont dans le même sens. C'est pourquoi nous travaillerons au départ des données de la FWB dans notre recherche.

Le questionnaire contextuel adressé aux élèves dans le cadre de PISA 2012 contient en effet plusieurs questions qui prétendent mesurer les opportunités d'apprentissage en mathématiques. Une d'elles, en particulier, nous intéresse pour notre recherche : la question ST62 — familiarité avec des concepts mathématiques. Cette question contient un certain nombre de pièges destinés à mesurer l'implication des élèves dans leurs réponses, qui peuvent être utilisés comme révélateur d'un biais de styles de réponse.

Les recherches menées dans le domaine des styles de réponse constituent autant de pistes méthodologiques que nous pouvons suivre pour répondre aux besoins de notre recherche. Grâce à elles, nous espérons réussir à mettre les styles de réponse sous contrôle afin d'améliorer la mesure des OTL en mathématiques.

Partie pratique

1 Question de recherche et hypothèses

De nombreuses variables contextuelles influencent la performance des élèves de quinze ans de la Fédération Wallonie-Bruxelles en culture mathématique. En particulier, le niveau socioéconomique de l'élève, le fait qu'il soit à l'heure, la filière et l'école qu'il fréquente sont de bons prédicteurs de sa performance. Mais qu'en est-il des opportunités d'apprentissage ? Plusieurs études (Schmidt & Maier, 2006 ; Scherff & Piazza, 2008 ; Schmidt et al, 2014 ; Lafontaine et al., 2015) ont établi qu'il existe un lien positif entre les OTL et le niveau académique. C'est ce que nous comptons investiguer pour la FWB.

Lafontaine et al. (2015) ont mis en évidence le fait que les OTL sont soumises aux biais de styles de réponse, comme cela est souvent le cas avec les variables issues de questionnaires contextuels. C'est pourquoi modéliser les OTL en prenant en considération les styles de réponse fera partie intégrante de notre recherche. De ces considérations découle la question qui guide notre réflexion et à laquelle nous allons essayer d'apporter des éléments de réponses :

Est-il possible d'améliorer la mesure des opportunités d'apprentissage en mathématiques en Fédération Wallonie-Bruxelles, en prenant en considération les styles de réponse des élèves, afin d'établir un lien entre les OTL et la performance en culture mathématique des élèves ?

Au départ de cette question, nous émettons plusieurs hypothèses que nous comptons vérifier.

Premièrement, nous émettons l'hypothèse que la distribution des OTL parmi les élèves n'est pas égale. Nous nous attendons à ce que les élèves à l'heure, ceux qui ont un statut socioéconomique élevé et ceux inscrits dans l'enseignement de transition reçoivent davantage d'opportunités d'apprentissage que les élèves en retard, ceux qui sont défavorisés du point de vue du ESCS ou ceux qui fréquentent l'enseignement qualifiant. Cette hypothèse n'est pas directement reliée à la question de départ, mais elle découle des recherches (Scherff & Piazza, 2008 ; Lafontaine et al., 2015) qui pointent les inégalités qui peuvent exister en termes d'OTL. Nous souhaitons vérifier dans quelle mesure la FWB est un système égalitaire ou non dans les chances d'apprendre qu'elle offre à ses élèves.

En second lieu, nous comptons repartir des corrélations entre les scores en culture mathématique des pays et les indices d'OTL (familiarité avec des concepts mathématiques) définis dans PISA 2012, avec et sans correction pour *overclaiming*, c'est-à-dire avec et sans prise en compte des styles de réponse des élèves. Notre hypothèse est que la prise en considération des styles de réponse devrait permettre d'améliorer la mesure des OTL. Nous prévoyons que, dans une majorité de pays (y compris en FWB), la corrélation entre l'indice FAMCONC et la performance des élèves soit supérieure à celle obtenue entre l'indice FAMCON et cette même performance.

Troisièmement, nous mettrons en relation la performance académique des élèves et leur propension à être affecté par les styles de réponse. Comme dans les recherches déjà menées (Grisay, 2009), nous émettons l'hypothèse d'une relation négative entre la performance académique et les styles de réponse. Les élèves de pays moins performants seraient plus affectés que ceux de pays avec un score moyen élevé. Nous vérifierons également cette hypothèse à l'intérieur d'un même système éducatif (la FWB), où nous nous attendons à des conclusions similaires : les élèves les plus performants, ceux de l'enseignement de transition et ceux à l'heure sont moins sujets aux styles de réponse que les élèves peu performants, de l'enseignement qualifiant, ou en retard.

En quatrième et dernier lieu, nous vérifierons, à partir des données de la FWB, comment il est possible d'améliorer la mesure des OTL en recourant à une nouvelle modélisation de l'indice, en prenant en considération les styles de réponse. Notre hypothèse est que recourir à une analyse factorielle pour modéliser un indice de biais (qui sera utilisé comme variable de contrôle) permettra d'améliorer le lien entre les OTL et la performance en culture mathématique des élèves de quinze ans.

2 Méthodologie

2.1 Échantillon

Dans le cadre de la présente recherche, nous avons utilisé les bases de données de l'enquête PISA 2012. Ces bases de données sont disponibles au téléchargement via le site internet de l'OCDE¹⁸. Pour la Fédération Wallonie-Bruxelles, ce sont quelques 3457 élèves, répartis au sein de 110 établissements scolaires qui ont pris part à l'enquête PISA 2012 et qui constituent dès lors notre échantillon.

Cet échantillon a été tiré en deux étapes (sélection d'écoles, puis d'élèves), en suivant une procédure stricte¹⁹ qui en garantit la représentativité par rapport à la population dont il est issu, à savoir l'ensemble des élèves de la FWB (Demonty et al., 2013 ; OCDE, 2014).

Le tableau 2 présente les proportions respectives d'élèves au sein de l'échantillon PISA 2012 et de la population-source, pour les différentes formes et années d'enseignement. Il montre bien que l'échantillon est extrêmement représentatif de la population dont il est issu, à deux exceptions près : un peu moins d'élèves issus de l'enseignement spécialisé et des CEFA dans l'échantillon que dans la population (Demonty et al., 2013).

Formes et années d'enseignement	Données administratives FWB		Échantillon
	Fréquence absolue	Fréquence relative	
1 ^{er} degré	6936	13,1%	12,35% (1,0)
3 ^e année de transition (G ou TT)	9057	17,1%	17,75% (0,7)
3 ^e année de qualification (TQ ou P)	9629	18,2%	17,3% (1,0)
4 ^e année de transition (G ou TT)	19 524	36,8%	39,3% (1,3)
4 ^e année de qualification (TQ ou P)	3752	7,1%	7,3% (0,9)
3 ^e degré	787	1,5%	1,2% (0,2)
CEFA	808	1,5%	0,9% (0,3)
Enseignement spécialisé (F3 et F4)	2529	4,8%	3,8% (0,5)
TOTAL	53 022	100,0%	100,0%

Tableau 2 : comparaison entre l'échantillon PISA 2012 en FWB et les données administratives de la population-cible, par forme et année d'enseignement.

¹⁸ <http://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm>

¹⁹ Voir 1.1.2. Procédure d'échantillonnage de PISA 2012.

Il est à noter que, bien que la taille initiale de l'échantillon soit de 3457 individus, ce nombre peut varier lors des analyses statistiques que nous avons menées. En effet, des élèves pouvaient être absents lors de l'épreuve ou refuser de répondre. De ce fait, 2947 individus constituent la base de données pour la FWB. De plus, des données manquantes sur certaines variables pourraient nous amener à supprimer les individus ayant produit ces données manquantes. C'est pourquoi la taille de l'échantillon sera systématiquement précisée dans la partie « Analyse et interprétation des données de PISA 2012 » de la présente recherche.

2.2 Méthodes et instruments

Afin de produire nos résultats de recherche, nous avons eu recours à deux logiciels différents : SAS et Mplus. Le premier a été utilisé pour préparer les fichiers de données, d'une part, et pour calculer les paramètres statistiques que nous analyserons dans la partie « Analyse et interprétation des données de PISA 2012 » d'autre part. Le second nous a permis de réaliser des analyses factorielles tant exploratoires que confirmatoires et d'obtenir des indices d'ajustement permettant d'attester ou de réfuter la validité d'un modèle.

Les paramètres statistiques que nous avons calculés sont essentiellement des coefficients de corrélation. Ils nous permettent de quantifier le lien qui peut exister entre deux variables, de savoir si elles évoluent dans un même sens ou en sens contraires, mais en aucun cas d'établir un lien de causalité entre elles. Nous avons également réalisé des analyses de régression linéaire qui permettent d'expliquer la variation d'une variable dépendante (typiquement, la performance académique) par une ou plusieurs variables indépendantes. Comme dit précédemment, l'échantillon PISA 2012 ne constitue pas un échantillon aléatoire et simple, ce dont le logiciel SAS fait le postulat. Pour garantir des résultats valides, nous avons adapté la manière de calculer les paramètres, notamment en ayant recours aux méthodes de répliques. Cela nous a permis de calculer des paramètres plus proches de la vérité, accompagnés de leur erreur standard (SE), qui détermine leur statut statistiquement significatif (avec un risque de première espèce de 5%)²⁰.

²⁰ Dans la partie « analyse et interprétation des données PISA 2012 », les valeurs statistiquement différentes de 0 seront suivies d'un astérisque (*).

Dans le but de garantir la reproductibilité de notre recherche, l'ensemble des syntaxes que nous avons utilisées, tant dans le logiciel SAS que dans le logiciel Mplus, seront placées en annexe.

2.2.1 Variables utilisées dans nos analyses

Le tableau 3 synthétise l'ensemble des variables que nous avons utilisées lors de nos analyses statistiques.

Variable	Signification
FAMCON	<i>Familiarity with mathematics concepts</i> , indice d'OTL de contenu.
FAMCONC	<i>Familiarity with mathematics concepts – corrected for overclaiming</i> , indice d'OTL de contenu qui prend en considération les styles de réponse des élèves.
QUALIFIANT	Fréquentation de l'élève de l'enseignement de transition (=0) ou de qualification (=1). Recodage de la variable ISCEDO.
ISCEDO	Orientation du programme d'études entre général (=1), pré-vocationnel (=2), vocationnel (=3) ou modulaire (=4).
REDOUBLE	Indication du fait que l'élève a déjà connu le redoublement (=1) ou non (=0). Recodage de la question ST01Q01 qui porte sur le grade de l'élève.
ESCS	<i>Economic, social and cultural status</i> , statut socioéconomique de l'élève.
QUARTILE	Indication du quartile de l'échelle ESCS dans lequel se trouve l'élève. Recodage de la variable ESCS sur base des valeurs de p25, p50 et p75.
BIAS	Indication de la tendance de l'élève à être sujet aux styles de réponse. Obtenue par addition de ST62Q04, ST62Q11 et ST62Q13, puis standardisation ($\mu=0$; $\sigma=1$).
RS_BIAS	Indice de styles de réponse calculé grâce à une analyse factorielle confirmatoire dans le logiciel Mplus.
INTMAT	Mathematics Interest, intérêt pour les mathématiques, construit à partir des items ST29.
MATHEFF	<i>Mathematics Self-Efficacy</i> , sentiment d'auto-efficacité en mathématiques, construit à partir des items ST37.
SCMAT	<i>Mathematics Self-Concept</i> , concept de soi en mathématiques, construit à partir des items ST42.
INTMAT_VDG	Indice d'intérêt pour les mathématiques, calculé grâce à une analyse factorielle confirmatoire dans le logiciel Mplus, sur base du modèle de Van de Gaer (2010).
MATHEFF_VDG	Indice d'auto-efficacité en mathématiques, calculé grâce à une analyse factorielle confirmatoire dans le logiciel Mplus, sur base du modèle de Van de Gaer (2010).

SCMAT_VDG	Indice de concept de soi en mathématiques, calculé grâce à une analyse factorielle confirmatoire dans le logiciel Mplus, sur base du modèle de Van de Gaer (2010).
RS_BIAS_VDG	Indice de styles de réponse, calculé grâce à une analyse factorielle confirmatoire dans le logiciel Mplus, sur base du modèle de Van de Gaer (2010).

Tableau 3 : signification des variables utilisées

3 Analyse et interprétation des données de PISA 2012

3.1 Distribution des OTL en FWB

Pour analyser la distribution des opportunités d'apprentissage en Fédération Wallonie-Bruxelles, nous sommes parti de la variable FAMCONC de la base de données PISA 2012. Pour rappel, cet indice mesure la familiarité avec un ensemble de concepts mathématiques, tout en tenant sous contrôle les styles de réponse des élèves, en ayant soustrait à FAMCONC la moyenne de chaque élève sur les trois items pièges. L'indice FAMCONC en FWB a une moyenne de 0,10 (SE=0,03) et un écart-type de 1,14 (SE=0,02).

Nous avons ensuite calculé les corrélations entre la variable FAMCONC et trois autres variables :

- QUALIFIANT (codée 0 si l'élève est dans l'enseignement de transition, 1 s'il est dans le qualifiant),
- REDOUBLE (codée 0 si l'élève n'a jamais connu le redoublement, 1 dans le cas contraire),
- ESCS, l'indice socioéconomique de l'élève.

Comme on pouvait s'y attendre, les corrélations obtenues en croisant FAMCONC et QUALIFIANT d'une part, et FAMCONC et REDOUBLE d'autre part, sont toutes deux négatives et statistiquement significatives. Elles sont de l'ordre de -0,35. La corrélation entre FAMCONC et ESCS est, elle aussi, significative, mais de signe positif et de plus faible ampleur ($r=0,24$).

De plus, il existe des différences de moyennes statistiquement significatives sur l'échelle FAMCONC entre les différents groupes qui composent les variables QUALIFIANT, REDOUBLE et ESCS. Ainsi, il y a une différence de familiarité avec des concepts

mathématiques entre les élèves du qualifiant et ceux du général, en faveur de ces derniers. De la même façon, la différence entre les élèves à l'heure et ceux en retard favorise les élèves qui n'ont jamais connu le redoublement. Pour le ESCS, la différence de moyenne est calculée en comparant les 25% d'élèves qui ont le statut socioéconomique le plus faible (QUARTILE=1) avec les 25% d'élèves les plus favorisés (QUARTILE=4). La différence négative indique une moyenne sur l'échelle FAMCONC plus élevée pour les élèves du quatrième quartile.

Paramètres	QUALIFIANT	REDOUBLE	ESCS
<i>r</i>	-0,35*	-0,36*	0,24*
<i>SE (r)</i>	0,02	0,02	0,02
<i>Différence de moyenne</i>	-0,89*	-0,82*	-0,66*
<i>SE (mean dif)</i>	0,06	0,05	0,08
n = 2778			

Tableau 4 : corrélations et différences de moyennes entre FAMCONC et QUALIFIANT, REDOUBLE et ESCS

Nous pouvons donc dire, conformément à notre première hypothèse, que la distribution des opportunités d'apprentissage définies par l'indice FAMCONC en Fédération Wallonie-Bruxelles est bien inégale. Avec une corrélation de -0,35 et une différence de moyenne de l'ordre de huit dixièmes d'écart-type, nous pouvons affirmer que les élèves de l'enseignement qualifiant déclarent être moins fortement familiarisés avec les concepts mathématiques, ce qui est révélateur de moins d'OTL reçues, en comparaison des élèves de l'enseignement général. Des conclusions similaires sont possibles entre les élèves redoublants et ceux qui sont à l'heure : les premiers ont reçu moins d'OTL que les seconds.

Ces différences en faveur tant des élèves à l'heure que de ceux de l'enseignement de transition s'expliquent assez aisément au regard des items qui composent la question ST62, base de l'échelle FAMCONC. Plusieurs des concepts mathématiques proposés dans les items ne sont abordés en classe, dans l'enseignement de transition, qu'en fin de troisième année secondaire, voire plus tard. C'est notamment le cas du cosinus (abordé en fin de 3^e et développé à partir de la 4^e), de la fonction du second degré (4^e) ou des nombres complexes (6^e, uniquement dans les classes les plus avancées). C'est aussi le cas des probabilités, mais qui constituent un terme faisant partie du langage courant, avec lequel des élèves pourraient être familiers même s'ils ne l'ont pas travaillé en classe.

Or, au moment de PISA 2012, 47% des élèves de quinze ans étaient répartis dans les trois premières années de l'enseignement secondaire, et 24% dans des filières de qualification, où

les programmes développent aussi plus tardivement certains des concepts mathématiques visés à la question ST62. On comprend dès lors que les différences entre les élèves soient aussi marquées. Néanmoins, ces différences restent révélatrices de la façon dont notre système sépare les élèves en ayant recours au redoublement ou aux filières précoces, et des différences en termes d'acquis, à âge égal, que cela peut entraîner entre les élèves.

Ces différences d'opportunités d'apprentissage reçues (ici, en termes de contenu, donc d'exposition curriculaire) constituent également une piste pour expliquer les différences de performances observées en FWB entre les différents groupes d'élèves (voir 1.2, p. 12). À nouveau, comme l'ont dit Scherff & Piazza (2008), il semble inconcevable de demander aux élèves de répondre à des normes similaires lorsqu'ils n'ont pas tous reçu les mêmes chances de réussite.

Concernant le lien entre le statut socioéconomique de l'élève et les OTL de contenu, avec une corrélation de 0,24, il est moins robuste, mais significatif. Il y a tout de même une différence de moyenne sur l'échelle FAMCONC entre les 25% d'élèves les plus favorisés et les 25% d'élèves les plus défavorisés de l'ordre de six dixièmes d'écart-type, qui nous fait dire que les élèves issus de milieux socioéconomiques plus élevés reçoivent de meilleures chances d'apprendre que ceux avec un ESCS plus faible. L'effet du ESCS est donc moins marqué que celui des filières et du redoublement, mais sans doute l'effet entre ces variables connaît-il une forme de redondance, les élèves les plus défavorisés étant souvent ceux qui fréquentent l'enseignement de qualification ou qui ont fait l'expérience du redoublement.

3.2 Familiarité avec des concepts mathématiques et niveau académique

Afin de vérifier notre deuxième hypothèse, c'est-à-dire afin de voir dans quelle mesure prendre en considération les styles de réponse, tel que c'est envisagé dans PISA 2012, peut améliorer la mesure des OTL, nous avons établi les corrélations entre la performance en culture mathématique des élèves et l'indice FAMCON (qui ne tient pas compte des styles de réponse) d'une part, et les corrélations entre la performance en mathématiques et l'indice FAMCONC (qui tient compte de ce biais) d'autre part.

Nous avons utilisé les données de 41 systèmes éducatifs pour réaliser cette analyse : les pays membres de l'OCDE, les membres de l'Union européenne non-membres de l'OCDE, Singapour, Macao et Hong Kong. Les trois communautés linguistiques de la Belgique ont été envisagées séparément pour cette analyse. La Norvège ne fait pas partie des pays retenus, car

aucune mesure pour les indices FAMCON et FAMCONC n'existe pour ce pays. Le nombre total d'individus dans l'échantillon pour cette analyse est de 321 645. Le tableau 5 présente les résultats obtenus.

<i>CNT</i>	<i>corr_famcon</i>	<i>se_corr_famcon</i>	<i>corr_famconc</i>	<i>se_corr_famconc</i>	<i>n</i>
AUS	0,54*	0,01	0,44*	0,01	14 481
AUT	0,50*	0,02	0,48*	0,02	4755
BFL	0,62*	0,01	0,50*	0,02	4877
BFR	0,42*	0,02	0,50*	0,02	2947
BGE	0,27*	0,08	0,44*	0,04	773
BGR	0,38*	0,02	0,40*	0,02	5282
CAN	0,42*	0,01	0,40*	0,01	21544
CHE	0,46*	0,02	0,50*	0,01	11 229
CHL	0,50*	0,02	0,39*	0,02	6856
CZE	0,44*	0,02	0,37*	0,02	5327
DEU	0,49*	0,02	0,52*	0,02	5001
DNK	0,38*	0,02	0,25*	0,02	7481
ESP	0,47*	0,01	0,47*	0,01	25 313
EST	0,32*	0,02	0,28*	0,02	4779
FIN	0,43*	0,02	0,32*	0,02	8829
FRA	0,44*	0,02	0,50*	0,02	4613
GBR	0,48*	0,02	0,34*	0,02	12659
GRC	0,33*	0,02	0,34*	0,02	5125
HKG	0,41*	0,02	0,20*	0,02	4670
HRV	0,40*	0,02	0,44*	0,02	5008
HUN	0,49*	0,03	0,55*	0,02	4810
IRL	0,46*	0,02	0,35*	0,02	5016
ISL	0,21*	0,03	0,33*	0,02	3508
ISR	0,42*	0,02	0,32*	0,02	5055
ITA	0,44*	0,01	0,45*	0,01	31 073
JPN	0,48*	0,02	0,36*	0,02	6351
KOR	0,61*	0,02	0,57*	0,02	5033
LUX	0,36*	0,02	0,38*	0,02	5258
MAC	0,47*	0,02	0,25*	0,02	5335
MEX	0,36*	0,01	0,26*	0,01	33806
NLD	0,54*	0,03	0,44*	0,02	4460
NZL	0,51*	0,02	0,39*	0,02	4291
POL	0,39*	0,03	0,36*	0,02	4607
PRT	0,44*	0,02	0,47*	0,01	5722

ROU	0,37*	0,02	0,28*	0,03	5074
SGP	0,56*	0,01	0,41*	0,02	5546
SVK	0,45*	0,02	0,40*	0,02	4678
SVN	0,39*	0,02	0,44*	0,02	5911
SWE	0,14*	0,02	0,35*	0,02	4736
TUR	0,35*	0,02	0,43*	0,02	4848
USA	0,49*	0,02	0,49*	0,02	4978
	0,41*	0,01	0,47*	0,01	n = 321 645

Tableau 5 : corrélations entre la performance des élèves et FAMCON et FAMCONC dans 41 pays

Dans l'ensemble des 41 pays du groupe, comme au niveau international, il existe des corrélations statistiquement significatives et positives, tant entre la performance des élèves et FAMCON qu'entre la performance et FAMCONC. Les élèves les plus forts sont donc ceux qui déclarent recevoir davantage d'opportunités d'apprentissage. Cependant, l'ampleur de ces liens est extrêmement variable d'un pays à l'autre.

Ainsi, la corrélation entre la performance et FAMCON varie de 0,14 en Suède au quadruple en Corée ou en Communauté flamande. Les coefficients obtenus entre les trois communautés linguistiques belges sont aussi très différents les uns des autres : 0,27 en Communauté germanophone, 0,47 en FWB et 0,62 en Communauté flamande. La corrélation obtenue en comptabilisant les élèves des 41 pays est de 0,41.

Les variations sur les coefficients de corrélation entre FAMCONC et la performance sont également importantes, mais moins marquées. Les coefficients varient de 0,20 à Hong Kong à 0,57 en Corée. Les trois communautés linguistiques belges obtiennent des corrélations de 0,50 (FWB et BFL) et 0,44 (BGE). Au niveau international, la corrélation est de 0,47.

Lorsqu'on compare le coefficient de corrélation entre les pays, on se rend compte que la prise en considération des styles de réponse améliore légèrement le lien entre les opportunités d'apprentissage et la performance en culture mathématique. Ce coefficient passe de 0,41 avec FAMCON à 0,47 avec FAMCONC. Nous sommes ici loin d'un paradoxe attitude-performance ou des résultats obtenus dans d'autres études. Cependant, il semble que la façon dont les styles de réponse sont envisagés dans PISA 2012, c'est-à-dire en insérant des items pièges dans l'échelle de familiarité avec des concepts mathématiques, permette d'améliorer la mesure des OTL, du moins à l'échelle internationale.

En revanche, à l'intérieur des pays, les résultats sont beaucoup plus mitigés. Dans un certain nombre de systèmes éducatifs, la prise en compte des items pièges, donc des styles de réponse, augmente en effet la corrélation entre les opportunités d'apprentissage et la performance en culture mathématique. C'est notamment le cas en Communauté germanophone, en Fédération Wallonie-Bruxelles, en Hongrie, en Islande, en Slovénie, en Suède ou en Turquie. Il semble donc, dans ces systèmes, que les élèves sont bien affectés par la présence des items pièges qui polluent la mesure. En particulier, cela semble témoigner chez les élèves d'une tendance à l'acquiescement ou d'un biais de désirabilité sociale. Ces pays ne sont cependant pas majoritaires.

Dans d'autres systèmes éducatifs, le changement d'indice pour calculer la corrélation avec la performance académique des élèves n'a aucune incidence. La différence dans ces pays entre les deux coefficients de corrélation est nulle ou trop faible pour être statistiquement significative. Les pays concernés ici sont notamment le Royaume-Uni, l'Espagne, la Grèce, l'Italie, le Luxembourg ou les États-Unis. Dans la logique dont l'enquête PISA 2012 envisage les styles de réponse, cela semble donc signifier que les élèves de ces systèmes ne sont pas du tout affectés par la tendance à l'acquiescement ou la désirabilité sociale, et que la mesure avec ou sans prise en compte des styles de réponse est de même validité.

Enfin, dans les systèmes éducatifs restants, le coefficient de régression basé sur FAMCONC est plus faible que celui qui ne prend pas en considération les styles de réponse. Il s'agit, par exemple, de l'Australie, de la Communauté Flamande, de Hong Kong, de la Finlande, de Macao ou de Singapour. Ces pays sont justement ceux, dans les études précédentes (Grisay, 2009), qui sont les moins entachés par les styles de réponse. Ces résultats suggèrent que, pour ces systèmes, la manière dont les styles de réponse sont envisagés dans PISA 2012 rend la mesure des OTL moins valide. En effet, il s'agit de pays asiatiques ou scandinaves pour la plupart, qui sont justement moins affectés par les biais de réponse. Dès lors, la méthode consistant à soustraire le score moyen aux items pièges pour contrôler ce biais n'est peut-être pas la meilleure. Cela pourrait aussi signifier que les élèves sont bien marqués par un biais de styles de réponse, mais qui conduit plutôt à surestimer la corrélation entre les OTL et la performance en culture mathématique.

Les résultats mitigés quant aux variations des coefficients de régression ne nous permettent pas de confirmer notre deuxième hypothèse. Nous pouvons seulement dire que la prise en considération des biais de styles de réponse telle qu'elle est envisagée au travers de

l'échelle FAMCONC a des effets très variables en fonction des pays. Cela nous amène à penser que les autres items de la question ST62 sont sans doute, eux aussi, sujets à un biais de réponse. Il conviendrait dès lors, conformément à notre quatrième hypothèse et aux travaux déjà menés, de modéliser un indice de biais et de l'utiliser comme variable de contrôle afin d'améliorer la mesure. C'est ce que nous essaierons de faire dans la quatrième partie de notre démarche de recherche.

3.3 Performance académique et styles de réponse

Notre troisième hypothèse était qu'il existe un lien entre les scores moyens des pays et la tendance des élèves qui en font partie à être affectés par les biais de styles de réponse. Pour la vérifier, nous avons utilisé les données des mêmes 41 systèmes éducatifs que précédemment. Cependant, nous avons supprimé un certain nombre de données, lorsque les élèves n'avaient pas répondu aux items ST62Q04, ST62Q11 ou ST62Q13.

En effet, c'est en additionnant les scores de ces trois items que nous avons construit un indice BIAS, qui reflète la tendance des élèves à se déclarer familiers avec des concepts mathématiques qui n'existent pas (nombre propre, échelle subjonctive, fraction déclarative). Il s'agit donc d'une mesure de la tendance à acquiescer de l'élève, ou du biais de désirabilité sociale.

Nous avons ensuite normalisé la variable BIAS ($\mu=0$; $\sigma=1$), avant de calculer, pour les 41 pays et au niveau international, la performance moyenne, le score moyen de la variable BIAS et la corrélation entre la performance et la variable BIAS. Le tableau 6 présente les résultats obtenus. Pour en faciliter la lecture, il est trié par performance moyenne décroissante des systèmes éducatifs.

En raison des données manquantes sur les trois items pièges, la taille de l'échantillon est moindre que précédemment. 203 337 individus le composent ici. Il en résulte aussi que les performances moyennes en culture mathématique ne reflètent pas réellement celles des pays, puisqu'un certain nombre d'élèves ont été supprimés. Par exemple, pour la FWB, le score est ici surévalué, puisqu'il passe de 493 à 502. Cependant, cette manipulation était nécessaire pour calculer les corrélations entre la performance et BIAS.

<i>CNT</i>	<i>mean perf</i>	<i>se mean perf</i>	<i>mean bias</i>	<i>se mean bias</i>	<i>corr perf bias</i>	<i>se corr perf bias</i>	<i>n</i>
SGP	574,57*	1,74	0,20*	0,02	0,13*	0,02	3625
HKG	562,94*	3,40	-0,17*	0,03	0,15*	0,02	3005
KOR	553,68*	4,79	-0,80*	0,01	-0,15*	0,03	3292
BFL	539,55*	3,14	-0,08*	0,02	-0,06*	0,02	3070
JPN	539,20*	3,61	-0,56*	0,01	-0,02	0,02	4137
MAC	538,85*	1,49	-0,03	0,02	0,14*	0,02	3458
CHE	533,88*	3,11	-0,15*	0,01	-0,15*	0,02	7190
NLD	530,61*	4,03	-0,02	0,02	0,01	0,03	2725
DEU	525,37*	2,98	-0,19*	0,02	-0,15*	0,02	2665
EST	522,80*	2,34	-0,15*	0,02	-0,07*	0,03	3087
FIN	521,49*	1,74	-0,48*	0,01	0,02	0,02	5610
CAN	521,06*	1,87	0,15*	0,01	-0,01	0,02	13572
POL	518,46*	3,75	0,42*	0,02	-0,02	0,02	2992
BGE	515,89*	3,22	-0,15*	0,04	-0,15*	0,06	493
AUT	510,29*	2,98	-0,22*	0,02	-0,14*	0,02	3000
AUS	509,08*	1,82	0,04*	0,01	0,13*	0,01	9086
CZE	507,31*	3,08	-0,22*	0,02	-0,05*	0,02	3395
NZL	504,14*	2,73	0,02	0,02	0,17*	0,02	2735
IRL	504,00*	2,49	-0,17*	0,02	0,10*	0,02	3215
DNK	503,95*	2,13	0,12*	0,02	0,07*	0,02	4633
SVN	503,64*	1,62	0,24*	0,02	-0,10*	0,02	3715
BFR	502,02*	3,05	-0,05	0,02	-0,23*	0,02	1794
FRA	501,14*	2,52	-0,10*	0,02	-0,20*	0,02	2830
ISL	498,01*	1,93	-0,50*	0,02	-0,09*	0,02	2174
GBR	497,67*	3,21	-0,03	0,02	0,12*	0,02	8035
LUX	494,40*	1,54	0,27*	0,02	-0,01	0,02	3339
PRT	490,84*	4,06	0,22*	0,02	-0,12*	0,02	3590
SVK	488,85*	3,48	-0,12*	0,02	-0,05	0,02	2932
ITA	487,31*	2,03	0,17*	0,01	-0,06*	0,01	19 865
ESP	486,98*	1,90	-0,57*	0,01	-0,13*	0,02	16 114
USA	483,07*	3,74	0,16*	0,02	-0,03	0,02	3139
SWE	482,34*	2,09	-0,60*	0,02	-0,11*	0,02	2966
HUN	480,19*	3,49	-0,14*	0,02	-0,22*	0,02	3071
HRV	472,54*	3,68	0,29*	0,02	-0,08*	0,02	3212
ISR	472,50*	4,83	-0,07*	0,02	0,03	0,02	3057
GRC	453,52*	2,78	-0,04	0,02	-0,05*	0,02	3284
TUR	450,21*	4,95	0,21*	0,03	-0,18*	0,02	3061
ROU	446,28*	3,79	0,64*	0,02	0,04	0,03	3279

BGR	445,26*	3,97	0,38*	0,02	-0,07*	0,02	3262
CHL	424,27*	3,14	0,23*	0,02	0,01	0,02	4302
MEX	415,93*	1,40	0,39*	0,01	0,11*	0,01	21 371
	490,18*	1,19	0,00	0,01	-0,13*	0,01	n = 203 377

Tableau 6 : performance en culture mathématique et biais de styles de réponse dans 41pays.

Un premier élément marquant concernant les items pièges de la question ST62 est qu'ils ne fonctionnent pas du tout de la même façon. Que ce soit en Fédération Wallonie Bruxelles (figure 13) ou au niveau international (figure 14), les pourcentages de réponses pour les différentes modalités des trois items sont très différents.

Les items ST62Q11 et ST62Q13 fonctionnent de façon similaire, avec plus de la moitié des élèves qui répondent 1 (je n'en ai jamais entendu parler), plus de 10% d'élèves qui répondent 2 (j'en ai entendu parler une fois ou deux) et de moins en moins de taux de réponse avec l'avancement des catégories. Seuls 4,79% d'élèves choisissent la catégorie 5 (je connais et comprends le concept) de l'item ST62Q11 en FWB, et ce pourcentage est encore plus faible au niveau international.

En revanche, tant chez nous que dans les autres systèmes éducatifs, l'item ST62Q04 présente des pourcentages très différents. Cet item présente des pourcentages de réponses à ses différentes modalités plutôt symétriques. Les catégories 1 et 5 affichent des taux de réponse similaires et plus élevés que les trois autres. Les modalités 2, 3 et 4 ont quant à elles des taux de réponse très proches. Il y a ainsi en FWB 32,2% d'élèves qui déclarent connaître et comprendre le concept de nombre propre.

Cela semble indiquer que l'item ST62Q04 capte mieux les styles de réponse des élèves, puisqu'il en piège davantage. En particulier, nous avançons l'hypothèse qu'en langue française et en langue anglaise, la ressemblance entre « nombre propre » (*proper number* en anglais) et « nombre premier » (*prime number*), concept avec lequel les jeunes de quinze ans ont de fortes chances d'être familiers, a pu amener un nombre non négligeable d'individus à répondre dans des catégories à droite de l'échelle de Likert. Cela pourrait témoigner d'un comportement satisfaisant des élèves qui parcourent la question sans vraiment s'y investir et croient reconnaître un terme familier.

La procédure FREQ

Overclaiming - Proper Number

ST62Q04	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	475	26.48	475	26.48
2	222	12.37	697	38.85
3	249	13.88	946	52.73
4	270	15.05	1216	67.78
5	578	32.22	1794	100.00

Overclaiming - Subjunctive Scaling

ST62Q11	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	1223	68.17	1223	68.17
2	243	13.55	1466	81.72
3	146	8.14	1612	89.86
4	96	5.35	1708	95.21
5	86	4.79	1794	100.00

Overclaiming - Declarative Fraction

ST62Q13	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	1218	67.89	1218	67.89
2	216	12.04	1434	79.93
3	137	7.64	1571	87.57
4	114	6.35	1685	93.92
5	109	6.08	1794	100.00

Figure 13 : pourcentages de réponses aux différentes modalités des items ST62Q04, ST62Q11 et ST62Q13 en FWB.

La procédure FREQ

Overclaiming - Proper Number

ST62Q04	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	50140	24.65	50140	24.65
2	32783	16.12	82923	40.77
3	37806	18.59	120729	59.36
4	38205	18.79	158934	78.15
5	44443	21.85	203377	100.00

Overclaiming - Subjunctive Scaling

ST62Q11	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	123079	60.52	123079	60.52
2	34136	16.78	157215	77.30
3	24738	12.16	181953	89.47
4	13470	6.62	195423	96.09
5	7954	3.91	203377	100.00

Overclaiming - Declarative Fraction

ST62Q13	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	112444	55.29	112444	55.29
2	35615	17.51	148059	72.80
3	26878	13.22	174937	86.02
4	16737	8.23	191674	94.25
5	11703	5.75	203377	100.00

Figure 14 : pourcentages de réponses aux différentes modalités des items ST62Q04, ST62Q11 et ST62Q13 à l'échelle internationale.

Pour ce qui est du lien qui existe au niveau international entre la performance moyenne des pays et la tendance des élèves à se déclarer familiers avec les concepts pièges, avec une corrélation statistiquement significative de $-0,13$, nous concluons que ce lien existe, mais qu'il est très faible. Les pays les plus performants sont donc ceux au sein desquels les élèves sont le moins affectés par les styles de réponse, et inversement. La figure 15 représente cette corrélation sous forme de nuage de points. On voit bien que la droite de régression qui l'accompagnerait aurait une pente négative, mais on remarque aussi à quel point le nuage est dispersé.

À l'intérieur des pays, donc au niveau élève, les coefficients de corrélations sont également très variables. Dans la majorité des pays, ils sont statistiquement significatifs (30 sur 41). Cependant, leur ampleur, tout comme leur signe, varie selon les systèmes éducatifs. Comme nous nous y attendions, dans la majorité des systèmes où il est significatif, le coefficient de corrélation est de signe négatif (dans 22 pays sur 30). Cela signifie que dans ces pays, les élèves les plus performants sont les moins susceptibles d'être affectés par les styles de réponse. L'ampleur de ce lien est très variable, mais elle atteint au mieux les $-0,20$ dans les pays asiatiques, à savoir en Corée, à Singapour et à Hong Kong.

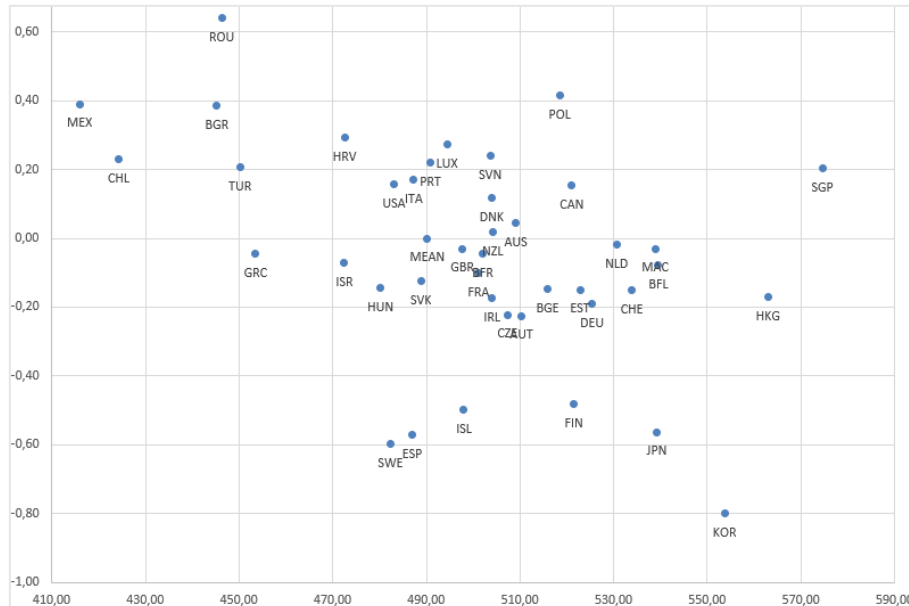


Figure 15 : corrélation entre BIAS et la performance en culture mathématique.

Il y a malgré tout huit systèmes éducatifs dans le groupe (la Croatie, Israël, la Grèce, la Turquie, la Roumanie, le Royaume-Uni, le Chili et le Mexique) où le coefficient de corrélation est significativement positif. Cela semble indiquer que, dans ces pays, les élèves les plus performants sont ceux qui sont le plus affectés par le biais de styles de réponse. Il s'agit en outre des huit pays du groupe de référence qui ont la performance moyenne la plus faible.

Si l'on s'intéresse plus spécifiquement à la Fédération Wallonie-Bruxelles (tableau 7), on voit que les coefficients de corrélation entre la variable BIAS et les variables QUALIFIANT, REDOUBLE et la performance des élèves en culture mathématique sont toutes trois significatives. Les deux premières sont de signe positif alors que la troisième est négative.

Paramètres	QUALIFIANT	REDOUBLE	PERFORMANCE
<i>r</i>	0,16*	0,18*	-0,22*
<i>SE (r)</i>	0,02	0,02	0,02
			n = 1780

Tableau 7 : corrélations entre BIAS et QUALIFIANT, REDOUBLE et la performance en culture mathématique en FWB.

Ces résultats vont dans le sens de ce que nous avons prévu dans nos hypothèses. En FWB, il existe bien un lien, même s'il n'est pas fort, entre le fait d'être sujet aux styles de réponse et la filière fréquentée. Les élèves de l'enseignement qualifiant sont plus susceptibles d'être affectés par ce biais que ceux de l'enseignement de transition. De la même façon, avec un coefficient de corrélation de 0,18, nous pouvons affirmer que les élèves qui sont à l'heure sont moins marqués par les biais de styles de réponse que ceux qui ont fait l'expérience du redoublement. Enfin, la performance est négativement corrélée en FWB avec les styles de réponse, ce qui signifie que les élèves les plus performants sont moins affectés.

Comme c'était le cas pour la distribution des OTL, nous supposons une certaine forme de redondance dans les résultats. Comme les élèves les moins performants sont justement ceux qui ont connu le redoublement et ceux qui fréquentent les filières de qualification, il pourrait être intéressant de déterminer les effets nets et joints de chacune de ces variables sur les styles de réponse, afin de pouvoir parler « toute chose étant égale par ailleurs ».

3.4 Modélisation factorielle des styles de réponse

La quatrième et dernière hypothèse de recherche que nous avons voulu mettre à l'épreuve était qu'il serait possible d'améliorer la mesure des opportunités d'apprentissage en ayant recours à une modélisation du biais de styles de réponse afin de pouvoir le tenir sous contrôle. Pour tester cette hypothèse, nous avons recours à plusieurs modélisations, que nous allons détailler une par une.

Dans toutes nos analyses factorielles, pour pouvoir attester de la validité d'un modèle, nous avons utilisé les balises suivantes :

- **Le CFI (*Comparative Fit Index*) doit avoir une valeur supérieure à 0,95 ;**
- **Le TLI (*Tucker-Lewis Fit Index*) doit avoir une valeur supérieure à 0,95 ;**
- **Le RMSEA (*Root Mean Square Error of Approximation*) doit au moins être inférieur à 0,08 (ajustement raisonnable). Idéalement, il devrait même être inférieur à 0,05 (bon ajustement) (Byrne, 2012).**

3.4.1 Modèle 1

Pour ce premier modèle, nous sommes parti uniquement des données de la Fédération Wallonie-Bruxelles (n=1857). Grâce au logiciel SAS, nous avons exporté les scores des élèves aux seize items de la question ST62 dans un fichier texte. Des scores ont été importés dans le logiciel Mplus afin de réaliser les analyses factorielles.

Notre objectif était d'arriver à une structure en trois facteurs. Le premier facteur, qui aurait regroupé tous les items, aurait été un indice OTL « brut », toujours entaché des styles de réponse. Le second facteur, sur lequel nous voulions voir saturer les items 4, 11 et 13, aurait reflété la tendance des élèves à l'acquiescement ou à la désirabilité sociale (indice de biais). Enfin, le troisième facteur attendu était un indice OTL épuré de ces styles de réponse, reprenant les items « vrais ». Une analyse factorielle exploratoire (*exploratory factor analysis*, EFA) nous a suggéré une réalité différente.

L'EFA réalisée suggère en effet qu'une structure en trois facteurs puisse fonctionner au regard des indices d'ajustement (voir supra), mais les saturations des seize items sur les facteurs

ne correspondent pas du tout à ce que nous attendions (figure 16). Nous sommes donc reparti de cette analyse exploratoire pour essayer de trouver un modèle qui fonctionne également en analyse confirmatoire (*confirmatory factor analysis*, CFA), tout en ayant du sens au regard du contenu des items ST62.

GEOMIN ROTATED LOADINGS			
	1	2	3
A	0.002	0.396	0.398
B	0.693	-0.125	0.062
C	0.241	-0.007	0.655
D	0.426	0.485	-0.190
E	0.447	0.273	0.251
F	0.508	-0.043	0.387
G	0.453	0.576	-0.008
H	0.467	0.233	0.370
I	0.447	0.012	0.505
J	-0.007	0.898	0.046
K	0.800	-0.133	0.004
L	-0.023	0.859	0.002
M	0.716	-0.003	-0.051
N	0.032	-0.112	0.892
O	-0.042	0.315	0.648
P	0.167	0.394	0.299

Figure 16 : saturations sur les trois facteurs du modèle 1, EFA.

Le passage du stade exploratoire au stade confirmatoire n'a pas permis d'obtenir un modèle valide. Le modèle le plus satisfaisant que nous ayons obtenu avait des indices d'ajustement qui s'approchaient des balises fixées (figure 17), mais pas assez bons pour pouvoir attester de sa validité. Seul le TLI avait une valeur acceptable. De plus, la structure factorielle du modèle 1 n'était pas satisfaisante en termes de contenu des items (tableau 8).

Un premier facteur obtenu regroupait les trois items pièges (nombre propre, échelle subjonctive, fonction déclarative), mais aussi un ensemble d'items « vrais » (fonction exponentielle, nombre complexe, probabilité, moyenne arithmétique, équation linéaire). Ce facteur pourrait refléter la tendance des élèves à être touchés par les styles de réponse. En effet, à quinze ans, les élèves ne sont normalement pas familiers avec les notions de nombre complexe, de fonction exponentielle ou de probabilité (même si ce mot fait partie du langage courant), puisqu'elles ne sont abordées que dans les deux dernières années de l'enseignement secondaire.

Pour ce qui est de la moyenne arithmétique et des équations linéaires, nous émettons l'hypothèse que les adjectifs puissent mettre en difficulté les élèves de la FWB. Dans les classes, ce sont plutôt les appellations « équation du premier degré » et « moyenne » qui sont utilisées. Cela pourrait expliquer que ces deux concepts constituent également des indicateurs de styles de réponse dans notre système éducatif.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	463.594*
Degrees of Freedom	34**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	8116.016
Degrees of Freedom	29
P-Value	0.0000
CFI/TLI	
CFI	0.947
TLI	0.955
Number of Free Parameters	94
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.082

Figure 17 : indices d'ajustement du modèle 1, CFA.

Le deuxième facteur obtenu dans cette analyse factorielle aurait pu refléter les OTL de contenu en mathématiques. En effet, il ne contient que des termes « vrais », avec lesquels les élèves de quinze ans sont susceptibles d'être familiers, exception faite de « fonction exponentielle » et « probabilité ». Il en va de même pour le troisième facteur, qui contient majoritairement des items « vrais », mais dans lequel se retrouve le piège « nombre propre » (confusion possible avec nombre premier) et « nombre complexe », qui ne fait normalement pas partie de ce que les élèves connaissent à quinze ans. De plus, les facteurs 2 et 3 ont une majorité de concepts communs, ce qui en rend l'interprétation difficile.

FACTOR 1	FACTOR 2	FACTOR 3
Fonction exponentielle <Nombre propre> Équation linéaire Nombre complexe <Échelle subjonctive> <Fonction déclarative> Moyenne arithmétique Probabilité	Fonction exponentielle Diviseur Fonction du second degré Équation linéaire Vecteurs Nombre rationnel Radicaux Polygone Cosinus Moyenne arithmétique Probabilité	Diviseur Fonction du second degré <Nombre propre> Équation linéaire Vecteurs Nombre complexe Nombre rationnel Radicaux Polygone Figure isométrique Cosinus

Tableau 8 : structure factorielle du modèle 1, CFA.

Cette première tentative de modélisation du biais de styles de réponse n'est donc pas satisfaisante, que ce soit au point de vue des indices des ajustements, ou en termes d'interprétation des dimensions qu'elle suggère.

3.4.2 Modèle 2

Face à la difficulté de trouver une structure factorielle valide pour la Fédération Wallonie-Bruxelles, nous avons essayé d'en faire émerger une sur un plus grand échantillon. Nous avons donc essayé au départ des données de l'ensemble de la Belgique (n=5467). L'idée que nous poursuivions était que cela permettrait de faire apparaître une structure factorielle que nous pourrions ensuite utiliser pour la FWB.

Une analyse factorielle exploratoire a tout d'abord suggéré qu'une structure en trois facteurs puisse fonctionner, comme pour la FWB. Le passage en analyse confirmatoire a permis d'obtenir des indices d'ajustement satisfaisants (figure 18) avec trois facteurs.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	962.048*
Degrees of Freedom	37**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	26310.398
Degrees of Freedom	29
P-Value	0.0000
CFI/TLI	
CFI	0.965
TLI	0.972
Number of Free Parameters	95
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.067

Figure 18 : indices d'ajustement du modèle 2, CFA.

Les dimensions du modèle sont en outre intéressantes. En particulier, nous y retrouvons un facteur sur lequel saturent les mêmes items pièges que dans le modèle 1, ce qui laisse suggérer que ces huit items sont bien ceux qui sont révélateurs d'un biais de styles de réponse. Cette dimension pourrait donc être le biais de styles de réponse.

FACTOR 1	FACTOR 2	FACTOR 3
Fonction exponentielle <Nombre propre> Équation linéaire Nombre complexe <Échelle subjonctive> <Fonction déclarative> Moyenne arithmétique Probabilité	Fonction exponentielle Diviseur Fonction du second degré Équation linéaire Vecteurs Nombre complexe Nombre rationnel Radicaux Polygone Figure isométrique Cosinus Moyenne arithmétique Probabilité	Fonction exponentielle Diviseur Fonction du second degré <Nombre propre> Équation linéaire Vecteurs Nombre complexe Nombre rationnel Figure isométrique Cosinus

Tableau 9 : structure factorielle du modèle 2, CFA.

Le second facteur de ce modèle est également très intéressant, puisqu'il reprend les treize items « vrais » de la question ST62. Il peut donc être utilisé comme indice OTL épuré, sans biais de styles de réponse. Quant au troisième facteur, il est plus difficile à expliquer, puisqu'il fait saturer tant des items « vrais » que des pièges. Peut-être la présence de « nombre propre » sur ce facteur est-elle due, comme avancé précédemment, à sa ressemblance avec « nombre premier ».

Ce modèle 2 s'avère particulièrement prometteur et pourrait être utilisé pour mettre les biais de styles de réponse sous contrôle, à l'échelle de la Belgique. Cependant, comme le but de notre recherche est d'apporter des résultats spécifiques à la Fédération Wallonie-Bruxelles, nous n'en ferons rien. Néanmoins, cela pourrait faire l'objet d'une recherche ultérieure.

Le modèle 2 ne fonctionne hélas pas sur la seule Fédération Wallonie-Bruxelles. C'est le modèle 1 qui reste le meilleur modèle que nous avons trouvé jusqu'à présent pour la FWB.

3.4.3 Modèle 3

Au vu de la difficulté d'établir un modèle valide pour la Fédération Wallonie-Bruxelles, compte tenu de la distribution inégale qui y existe quant aux opportunités d'apprentissage, mais aussi de la corrélation entre les styles de réponse et le parcours des élèves, nous avons envisagé pour ce troisième modèle un fonctionnement différentiel des items de la question ST62. En effet, selon que les élèves sont à l'heure ou ont déjà connu le redoublement, un même concept mathématique devrait leur être familier (il serait donc un indicateur d'OTL) ou non (il serait alors un indicateur de styles de réponse).

C'est, par exemple, le cas de l'item ST62Q03 « fonction du second degré ». Ce concept n'est abordé dans les classes de la FWB qu'à partir de la 4^e année de l'enseignement secondaire. Les élèves à l'heure sont donc susceptibles de la connaître. En revanche, un élève en retard qui se déclarerait familier avec ce point de contenu serait plutôt en train d'acquiescer.

Il en va de même pour les filières d'enseignement. Les élèves de l'enseignement de transition reçoivent davantage d'OTL que ceux de l'enseignement de qualification. Dès lors, ce qui est une OTL vraie pour les uns peut être un indice de styles de réponse pour les autres.

Nous avons donc essayé de modéliser les styles de réponse des élèves de la Fédération Wallonie-Bruxelles en les séparant en trois groupes distincts :

- le groupe 1 se compose des élèves à l'heure (ST01Q01=10 ; 11 ; 12) et inscrits dans l'enseignement de transition (ISCEDO=1),
- le groupe 2 comprend les élèves en retard (ST01Q01=7 ; 8 ; 9) qui fréquentent l'enseignement de transition (ISCEDO=1),
- le groupe 3 est composé des élèves de l'enseignement de qualification (ISCEDO=2 ; 3).

Dans le groupe 1, les données de 821 élèves ont été importées dans un fichier texte grâce au logiciel SAS. Ces données ont ensuite été utilisées dans le logiciel Mplus afin de réaliser les analyses factorielles. Dans un premier temps, une EFA a permis de faire émerger une structure en trois facteurs. De là, le passage en analyse confirmatoire a permis de créer un modèle valide au regard des indices d'ajustement (figure 19), avec un RMSEA particulièrement satisfaisant (=0,046).

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	154.136*
Degrees of Freedom	57**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	2291.839
Degrees of Freedom	33
P-Value	0.0000
CFI/TLI	
CFI	0.957
TLI	0.975
Number of Free Parameters	92
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.046

Figure 19 : indices d'ajustement du modèle 3, groupe 1, CFA.

La structure factorielle (tableau 10) de ce modèle 3, pour le groupe 1, est également satisfaisante. Un premier facteur rassemble la totalité des seize items de la question ST62. Il s'agit donc d'un indice OTL brut, touché par les styles de réponse des élèves.

Le second facteur, celui qui nous intéresse le plus pour notre recherche, regroupe neuf items. Parmi ceux-ci, on retrouve les huit items qui étaient déjà indicateurs d'un biais de styles de réponse dans les modèles 1 et 2. À ceux-ci vient s'ajouter l'item h « nombre rationnel ». Ce sont donc ces neuf items qui sont révélateurs du biais de styles de réponse, chez les élèves à l'heure, inscrits dans l'enseignement de transition, en Fédération Wallonie-Bruxelles.

La présence des nombres rationnels dans l'indice de styles de réponse est assez compréhensible. Bien qu'ils soient enseignés dès l'école primaire, l'appellation « nombres décimaux » ou « nombres fractionnaires » est souvent privilégiée. En outre, il n'est fait aucune mention des nombres rationnels dans les socles de compétences (Stegen, Géron, Daro, & Desmet, n.d.). Ainsi, bien que les élèves les manipulent lors des cours de mathématiques, ils pourraient ne pas être familiers du terme en lui-même.

La troisième dimension du modèle, quant à elle, n'est expliquée que par trois items. Il s'agit de trois concepts qui n'apparaissent que tardivement dans la scolarité des élèves de la FWB. Il pourrait donc s'agir d'un indice d'avancement, qui reflète dans quelle mesure les élèves sont plus avancés dans le curriculum que ce qui est normalement attendu à quinze ans.

FACTOR 1	FACTOR 2	FACTOR 3
Fonction exponentielle Diviseur Fonction du second degré <Nombre propre> Équation linéaire Vecteurs Nombre complexe Nombre rationnel Radicaux <Échelle subjonctive> Polygone <Fonction déclarative> Figure isométrique Cosinus Moyenne arithmétique Probabilité	Fonction exponentielle <Nombre propre> Équation linéaire Nombre complexe Nombre rationnel <Échelle subjonctive> <Fonction déclarative> Moyenne arithmétique Probabilité	Fonction exponentielle Moyenne arithmétique Probabilité

Tableau 10 : structure factorielle du modèle 3, groupe 1, CFA.

Nous avons donc exporté les scores des 821 élèves sur chacune de ces trois dimensions dans un fichier séparé afin de pouvoir les utiliser dans le logiciel SAS. Grâce à ces scores, nous avons pu calculer un ensemble de corrélations présentées dans le tableau 11.

Le facteur RS_BIAS, calculé grâce à l'analyse factorielle, est positivement corrélé avec FAMCON et négativement avec FAMCONC. La corrélation entre RS_BIAS et la performance en culture mathématique ne diffère statistiquement pas de 0.

ρ	Performance	FAMCON	FAMCONC	RS_BIAS
Performance	1	0,20* (0,04)	0,21* (0,04)	-0,01 (0,04)
FAMCON		1	0,29* (0,04)	0,51* (0,04)
FAMCONC			1	-0,52* (0,03)
RS_BIAS				1

Tableau 11 : matrice de corrélations dans le groupe 1.

Il est intéressant de noter que l'indice RS_BIAS est mieux corrélé avec FAMCON et FAMCONC qu'ils ne le sont entre eux. La corrélation positive entre FAMCON et RS_BIAS suggère que les élèves les plus marqués par des styles de réponse obtiennent les scores les plus élevés sur l'échelle de familiarité avec des concepts mathématiques. Cela paraît faire sens puisque cette échelle ne prend pas les biais de styles de réponse en considération.

En revanche, la corrélation négative de -0,52 entre FAMCONC et RS_BIAS indique que les élèves qui obtiennent les scores les moins élevés sur l'échelle FAMCONC sont ceux qui sont les plus touchés par les styles de réponse. Cela se comprend, puisque l'échelle FAMCONC est construite en soustrayant les items pièges à l'échelle FAMCON.

La corrélation non significative entre la performance en culture mathématique et RS_BIAS indique que, lorsque les élèves sont à l'heure et fréquentent l'enseignement de transition, leur performance académique n'a plus de lien avec la manière dont ils sont affectés par les biais de styles de réponse.

Afin de voir si la prise en considération des styles de réponse permet d'améliorer le lien entre les OTL et la performance en culture mathématique, nous avons réalisé trois analyses de régression linéaire (tableau 12). La première explique la performance par l'indice FAMCON, c'est-à-dire sans tenir compte des styles de réponse. La seconde s'appuie sur FAMCONC, donc

en prenant en considération les styles de réponse comme cela est envisagé dans PISA 2012. La troisième explique la performance par FAMCON et RS_BIAS. Cela permet de voir l'effet des FAMCON sous contrôle des styles de réponse calculés grâce à une analyse factorielle.

$$Performance = \alpha + \beta(FAMCON) + \varepsilon$$

$$Performance = \alpha + \beta(FAMCONC) + \varepsilon$$

$$Performance = \alpha + \beta_1(FAMCON) + \beta_2(RS_BIAS) + \varepsilon$$

Paramètres	FAMCON seul	FAMCONC seul	FAMCON, sous contrôle de RS_BIAS
Intercept	555,89* (3,67)	556,76* (3,65)	553,04* (3,81)
$\beta(FAMCON)$	18,62* (3,85)	/	25,71* (4,34)
$\beta(FAMCONC)$	/	13,48* (2,49)	/
$\beta(RS_BIAS)$	/	/	-32,50* (9,14)
r^2	0,04* (0,02)	0,04* (0,02)	0,06* (0,02)

Tableau 12 : paramètres des analyses de régression réalisées sur le groupe 1.

On remarque que le coefficient de régression des OTL mesuré au travers de l'indice FAMCON (18,62) tend à diminuer lorsqu'on prend les biais de styles de réponse par l'indice FAMCONC (13,44). À l'inverse, la modélisation factorielle permet d'augmenter ce coefficient. Sous contrôle des styles de réponse des élèves à l'heure de l'enseignement de transition, une augmentation de 1 point sur l'échelle FAMCON entraîne une augmentation de la performance de 25,71 points. De plus, le pourcentage d'explication augmente très légèrement lorsqu'on prend ainsi les styles de réponse en compte (de 4% à 6%).

Cependant, au vu de la taille réduite de l'échantillon, les différences observées ne sont pas statistiquement valides. Le modèle reste néanmoins une piste intéressante qui pourrait être reproduite à une plus grande échelle, afin de tirer des conclusions fermes et définitives.

Dans le groupe 2, une démarche similaire à celle décrite précédemment a été menée, avec les données de 550 élèves en retard de l'enseignement de transition. Cependant, il ne fut pas possible de faire émerger un modèle pour ce groupe qui soit valide du point de vue des indices d'ajustement (figure 20). Le meilleur modèle trouvé était un modèle à deux dimensions (tableau 12), n'utilisant que treize des seize items de la question ST62, mais son TLI, son CFI et son RMSEA étaient tous trois insatisfaisants.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	272.423*
Degrees of Freedom	32**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	2078.319
Degrees of Freedom	22
P-Value	0.0000
CFI/TLI	
CFI	0.883
TLI	0.920
Number of Free Parameters	72
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.117

Figure 20 : indices d'ajustement du modèle 3, groupe 2, CFA.

La structure factorielle obtenue, si les indices d'ajustement avaient été satisfaisants, aurait été intéressante. En effet, le facteur 1 regroupait l'ensemble des treize items utilisés, ce qui suggère qu'il s'agissait d'un indice de styles de réponse brut, comme dans le groupe 1. Quant aux facteurs 2, il était expliqué par 6 items, qui étaient déjà revenus comme pièges dans les modèles précédents. Cette seconde dimension aurait donc pu être utilisée comme indice de styles de réponse pour les élèves en retard de l'enseignement de transition, si les indices d'ajustement avaient été meilleurs.

FACTOR 1	FACTOR 2
Diviseur Fonction du second degré <Nombre propre> Équation linéaire Vecteurs Nombre complexe Nombre rationnel Radicaux <Échelle subjonctive> Polygone <Fonction déclarative> Figure isométrique Cosinus	<Nombre propre> Équation linéaire Nombre complexe Nombre rationnel <Échelle subjonctive> <Fonction déclarative>

Tableau 13 : structure factorielle du modèle 3, groupe 2, CFA.

Au vu de la non-validité de ce modèle, les scores n'ont pas été utilisés dans le logiciel SAS comme pour le groupe 1.

Dans le groupe 3, qui comptait 486 élèves, l'approche exploratoire a suggéré une structure en trois facteurs. Cependant, comme pour le groupe 2, il n'a pas été possible d'obtenir des indices d'ajustements satisfaisants (figure 21) permettant d'attester de la validité du modèle. Seul le TLI de cette modélisation avait une valeur suffisamment élevée.

La structure factorielle obtenue avec le groupe 3 était pourtant intéressante (tableau 13). Comme pour le groupe 1, un premier facteur reprenait l'ensemble des items de la question ST62, ce qui permettait de l'assimiler à un indice OTL brut. Un second facteur, qui témoignait des styles de réponse, était expliqué par neuf items. On y retrouvait les trois items pièges originaux (d, j et l), ainsi que six autres, avec lesquels les élèves du qualifiant ont en effet peu de chances d'être familiers. En particulier, le concept de cosinus revient pour ces élèves dans le facteur de styles de réponse, alors que ce n'était pas le cas pour les élèves à l'heure de l'enseignement de transition. À l'inverse, «équation linéaire» n'apparaît plus dans la dimension de biais.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	262.275*
Degrees of Freedom	52**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	1944.977
Degrees of Freedom	22
P-Value	0.0000
CFI/TLI	
CFI	0.891
TLI	0.954
Number of Free Parameters	97
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.091

Figure 21 : indices d'ajustement du modèle 3, groupe 3, CFA.

Quant au troisième facteur, il faisait intervenir tant des items avec lesquels les élèves de l'enseignement qualifiant sont susceptibles d'être familiers (diviseur, polygone) que d'autres qui ne devraient pas leur évoquer grand-chose (fonction exponentielle, cosinus). Cela rend l'interprétation de cette dimension en termes de variable latente difficile.

Compte tenu des indices d'ajustement insatisfaisants, les scores des élèves de l'enseignement de transition n'ont pas non plus été exportés afin d'être utilisés dans le logiciel SAS.

FACTOR 1	FACTOR 2	FACTOR 3
Fonction exponentielle Diviseur Fonction du second degré <Nombre propre> Équation linéaire Vecteurs Nombre complexe Nombre rationnel Radicaux <Échelle subjonctive> Polygone <Fonction déclarative> Figure isométrique Cosinus Moyenne arithmétique Probabilité	Fonction exponentielle <Nombre propre> Nombre complexe Nombre rationnel <Échelle subjonctive> <Fonction déclarative> Cosinus Moyenne arithmétique Probabilité	Fonction exponentielle Diviseur Fonction du second degré Équation linéaire Polygone Cosinus Moyenne arithmétique Probabilité

Tableau 14 : structure factorielle du modèle 3, groupe 3, CFA.

Ce troisième modèle, en conclusion, même s'il n'a pas permis d'apporter les réponses attendues, reste intéressant. En particulier, il met en lumière que, dans un système éducatif qui prône un modèle de la séparation (Mons, 2007), le parcours scolaire des élèves, défini en termes de grade et de filière, influence les biais de styles de réponse. Ce qui constitue un indicateur de ce biais pour les élèves en retard n'en est pas forcément un pour les élèves à l'heure. La construction d'un indice de styles de réponse valide dans ces systèmes éducatifs pourrait donc être plus difficile que dans les systèmes intégratifs, comme l'Islande, par exemple. En effet, nous avons sans grande difficulté, fait apparaître pour ce pays une modélisation en trois dimensions qui donne des indices d'ajustements (figure 22) satisfaisants.

Les pistes de réponses obtenues sur les élèves à l'heure de l'enseignement de transition sont également intéressantes et suggèrent que la prise en considération des styles de réponse en ayant recours aux analyses factorielles permet en effet d'améliorer des opportunités d'apprentissage de contenu.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	627.198*
Degrees of Freedom	53**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	12099.053
Degrees of Freedom	13
P-Value	0.0000
CFI/TLI	
CFI	0.952
TLI	0.988
Number of Free Parameters	93
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.070

Figure 22 : indices d'ajustement du modèle 3, Islande, CFA.

3.4.4 Modèle 4

Puisqu'il n'a pas été possible d'établir un indice de styles de réponse valide pour la Fédération Wallonie-Bruxelles au départ des items de la question ST62, nous avons voulu vérifier si cela était envisageable au départ d'autres items. Nous nous sommes inspiré pour ce faire de la méthodologie développée par Van de Gaer (2010), que nous avons essayé de reproduire, dans l'objectif de pouvoir l'utiliser pour améliorer la mesure des OTL en FWB.

Dans un premier temps, nous avons donc tenté de reproduire le plus fidèlement possible le modèle développé par Van de Gaer (voir figure 12 p. 40) avec les données issues cette fois de PISA 2012. Pour rappel, Van de Gaer basait sa recherche sur les données de PISA 2003. Nous avons réalisé cette étape avec les données de l'OCDE dans son ensemble, mais uniquement pour les élèves qui avaient reçu la deuxième forme du questionnaire contextuel (QUESTID=2). En effet, comme pour le questionnaire cognitif, plusieurs formes (trois) du questionnaire contextuel ont été adressées aux élèves, et seule la seconde contenait à la fois les items ST29, ST37 et ST42 (tableau15).

QESTID	ST29	ST37	ST42
1	Oui	Oui	Non
2	Oui	Oui	Oui
3	Non	Non	Non

Tableau 15 : répartition des items ST29, ST37 et ST42 dans les différentes formes des questionnaires contextuels adressés aux élèves.

Grâce au logiciel SAS, nous avons inversé les scores de tous les items, exception faite de l’item ST42Q02. En effet, l’échelle de Likert sur tous les items était construite à l’envers, de « tout à fait d’accord » à « pas du tout d’accord ». Il a donc fallu inverser les scores, afin qu’un score plus élevé corresponde effectivement, par exemple, à un intérêt pour les mathématiques plus grand. Cependant, comme l’item ST42Q42, en plus de l’échelle de Likert inversée, était rédigé négativement, il n’avait pas besoin d’être inversé.

Ce sont donc les données de quelque 96 895 élèves que nous avons importées dans le logiciel Mplus afin de trouver une structure factorielle satisfaisante. Cette étape n’a pas posé de grosses difficultés et nous avons rapidement pu obtenir des indices d’ajustement convenables (figure 23), et une structure factorielle identique à celle obtenue au départ des données PISA 2003, à savoir en quatre variables latentes (tableau 16).

FACTOR 1 (INTMAT)	FACTOR 2 (MATHEFF)	FACTOR 3 (SCMAT)	FACTOR 4 (RS_BIAS)
ST29Q02 ST29Q05 ST29Q07 ST29Q08	ST37Q01 ST37Q02 ST37Q03 ST37Q04 ST37Q05 ST37Q06 ST37Q07 ST37Q08	ST42Q02 ST42Q04 ST42Q06 ST42Q07 ST42Q09	ST29Q02 ST29Q05 ST29Q07 ST29Q08 ST42Q02 ST42Q04 ST42Q06 ST42Q07 ST42Q09

Tableau 16 : structure factorielle du modèle reproduit d’après Van de Gaer (2010).

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	23887.150*
Degrees of Freedom	50**
P-Value	0.0000
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference tests. MLM, MLR and WLSM chi-square difference testing is described in the Mplus Technical Appendices at www.statmodel.com . See chi-square difference testing in the index of the Mplus User's Guide.	
** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to a formula given in the Mplus Technical Appendices at www.statmodel.com . See degrees of freedom in the index of the Mplus User's Guide.	
Chi-Square Test of Model Fit for the Baseline Model	
Value	824849.878
Degrees of Freedom	18
P-Value	0.0000
CFI/TLI	
CFI	0.971
TLI	0.990
Number of Free Parameters	81
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.070

Figure 23 : indices d'ajustement du modèle reproduit d'après Van de Gaer (2010).

Les trois premières variables sont chacune expliquées par les items qui correspondent à un des indices utilisés (SCMAT, INMAT et MATHEFF). Quant à la quatrième, comme prévu, elle est expliquée à la fois par les items relatifs à l'intérêt pour les mathématiques et à ceux qui traitent du concept de soi en mathématiques, mais pas par les items qui mesurent le sentiment d'auto-efficacité en mathématiques. Cette dernière dimension est donc bien l'indice de biais de styles de réponse tel que défini par Van de Gaer.

L'indice de styles de réponse ainsi calculé est négativement et significativement corrélé avec la performance moyenne de tous les pays de l'OCDE, à concurrence de -0,15 (tableau 17). En revanche, à l'intérieur des pays, cette corrélation est généralement plus faible, ce qui va dans le sens selon lequel la désirabilité est bien un facteur culturel. En FWB, cette corrélation ne diffère pas statistiquement de 0. Il n'y a donc pas de lien chez nous entre la désirabilité sociale au niveau de l'élève et la performance académique.

CNT	CORR	SE_CORR
AUS	-0,04*	0,02
AUT	-0,07*	0,03
BFL	-0,03	0,03
BFR	-0,07	0,04
BGE	0,11	0,07
CAN	-0,03	0,02
CHE	-0,14*	0,02
CHL	-0,05*	0,03
CZE	-0,03	0,03
DEU	-0,06	0,03
DNK	0,02	0,03
ESP	-0,06*	0,02
EST	-0,07*	0,03
FIN	-0,01	0,02
FRA	-0,11*	0,03
GBR	-0,06*	0,03
GRC	0,06	0,03
HUN	-0,17*	0,03
IRL	-0,05	0,03
ISL	0,00	0,04
ISR	-0,25*	0,03
ITA	-0,05*	0,02
JPN	0,11*	0,02
KOR	0,12*	0,03
LUX	-0,04*	0,02
MEX	-0,18*	0,02
NLD	-0,11*	0,03
NOR	0,03	0,03
NZL	-0,15*	0,03
POL	-0,12*	0,03
PRT	-0,11*	0,03
SVK	-0,20*	0,03
SVN	-0,05	0,03
SWE	0,00	0,03
TUR	-0,07*	0,03
USA	-0,12*	0,03
OCDE	-0,15*	0,01

Tableau 17 : corrélation entre RS_BIAS_VDG et la performance en culture mathématique pour les pays de l'OCDE.

Comme dans la recherche initiale menée par Van de Gaer (2010), la prise en considération des styles de réponse a permis d'améliorer les corrélations entre les trois variables (INTAMT, MATHEFF et SCMAT) et la performance en culture mathématique, tant entre les pays qu'à l'intérieur de ceux-ci. Les différences les plus fortes s'observent, comme prévu, sur INTMAT (tableau 18).

Ainsi, la corrélation entre INTMAT et la performance en culture mathématique, pour l'ensemble des pays de l'OCDE passe de 0,07 à 0,37 lorsqu'on applique la modélisation factorielle. Les différences sont également très marquées à l'intérieur des pays. Ainsi, les augmentations de corrélations à l'intérieur des pays varient de 0,16 (Japon, Corée) à 0,37 (Slovaquie, Pologne). De plus, les rares corrélations qui n'étaient pas significatives (Israël, Mexique, Slovaquie) le deviennent.

Des résultats similaires, mais plus modestes s'observent sur la variable SCMAT (voir annexe P, p. XXXIII), dont la corrélation entre les pays passe de 0,30 à 0,37, avec des augmentations au sein des pays allant de 0,03 à 0,09. Quant à la variable MATHEFF, conformément aux résultats présentés par Van de Gaer, c'est celle pour laquelle les différences sont les moins marquées (voir annexe O, p. XXXII).

Ce modèle corrobore donc les recherches menées sur les données de PISA 2003. Il existe bien un facteur de biais de réponse (plus spécifiquement, de désirabilité sociale) sur les items qui expliquent l'intérêt pour les mathématiques et le concept de soi en mathématiques. En revanche, la mesure du sentiment d'auto-efficacité en mathématiques n'est pas entachée par la désirabilité sociale.

CNT	corr_intmat	se_corr_intmat	corr_intmat_vdg	se_corr_intmat_vdg
AUS	0,23*	0,02	0,52*	0,01
AUT	0,18*	0,02	0,46*	0,02
BFL	0,17*	0,03	0,37*	0,02
BFR	0,20*	0,04	0,46*	0,03
BGE	0,29*	0,07	0,44*	0,05
CAN	0,24*	0,02	0,53*	0,01
CHE	0,11*	0,02	0,43*	0,02
CHL	0,13*	0,03	0,36*	0,03
CZE	0,23*	0,03	0,49*	0,02
DEU	0,17*	0,03	0,44*	0,02
DNK	0,31*	0,02	0,58*	0,02

ESP	0,19*	0,02	0,46*	0,02
EST	0,21*	0,02	0,51*	0,02
FIN	0,33*	0,02	0,60*	0,01
FRA	0,18*	0,03	0,48*	0,02
GBR	0,20*	0,02	0,48*	0,02
GRC	0,28*	0,03	0,49*	0,02
HUN	0,15*	0,03	0,51*	0,03
IRL	0,20*	0,03	0,48*	0,02
ISL	0,30*	0,03	0,56*	0,02
ISR	-0,04	0,03	0,30*	0,02
ITA	0,16*	0,01	0,41*	0,01
JPN	0,32*	0,02	0,48*	0,02
KOR	0,41*	0,03	0,60*	0,02
LUX	0,18*	0,02	0,43*	0,02
MEX	0,02	0,02	0,30*	0,01
NLD	0,12*	0,03	0,34*	0,03
NOR	0,37*	0,03	0,64*	0,02
NZL	0,11*	0,03	0,47*	0,02
POL	0,26*	0,02	0,62*	0,02
PRT	0,19*	0,03	0,53*	0,02
SVK	0,05	0,03	0,42*	0,03
SVN	0,17*	0,03	0,44*	0,02
SWE	0,28*	0,03	0,54*	0,02
TUR	0,09*	0,03	0,31*	0,03
USA	0,14*	0,03	0,48*	0,02
OCDE	0,07*	0,01	0,37	0,01

Tableau 18 : corrélations entre INTMAT et la performance en culture mathématique dans 37 systèmes éducatifs, sans avec prise en compte des styles de réponse.

L'étape suivante de notre démarche consistait à reprendre le modèle de Van de Gaer, et à le compléter en y ajoutant les scores des élèves sur les items de la question ST62. Nous avons, hélas, été confronté à un obstacle majeur : la forme 2 du questionnaire contextuel élève ne contenait pas la question ST62. Celle-ci n'était présente que dans les formes 1 et 3. Il n'y a donc aucun élève dans la base de données PISA 2012 qui ait répondu à tous les items dont nous avons besoin pour réaliser nos analyses factorielles. Il est donc impossible d'ajouter les OTL directement dans le modèle de Van de Gaer.

De la même manière, il n'était donc pas possible d'utiliser l'indice de biais de styles de réponse construit à partir du modèle de Van de Gaer comme variable de contrôle des opportunités d'apprentissage en FWB, puisque les élèves ayant répondu à la forme 2 n'ont aucune donnée sur les variables FAMCON et FAMCONC, tous deux construits à partir des items ST62.

Il nous est donc impossible d'investiguer, au départ de ce modèle, la manière dont la prise en considération des styles de réponse pourrait améliorer la mesure des opportunités d'apprentissage en Fédération Wallonie-Bruxelles.

À titre indicatif, nous avons calculé les corrélations entre l'indice de biais de styles de réponse obtenus en reproduisant la méthodologie de Van de Gaer et un ensemble de variables contextuelles (tableau 19).

Ces corrélations indiquent qu'il existe un lien significatif entre le biais de styles de réponse et les trois variables contextuelles FAILMAT (attributions causales de l'échec en mathématiques), ANXMAT (anxiété à l'égard des mathématiques) et INSTMOT (motivation instrumentale en mathématiques). Cela indique que la mesure de ces variables est entachée par ces styles de réponse, plus particulièrement, par la désirabilité sociale.

Nous notons aussi que la variable « positive » (en termes de relation avec la performance) est plus fortement reliée au biais de styles de réponse que les deux autres. Cela va dans le sens de la recherche menée par Lafontaine et al. (2019) qui concluaient à l'existence d'un paradoxe attitude-performance sur une variable « positive », mais pas sur son pendant négatif.

Au départ de ces corrélations, nous ne pouvons qu'extrapoler, et supposer que la mesure des OTL est, comme pour les trois variables présentées ici, biaisée par les styles de réponse.

Variabes	CORR RS_BIAS_VDG	SE_CORR RS_BIAS_VDG
FAILMAT	-0,21*	0,05
ANXMAT	-0,11*	0,05
INSTMOT	0,51*	0,03
n = 923		

Tableau 19 : corrélations entre RS_BIAS_VDG et FAILMAT, ANXMAT et INSTMOT.

4 Conclusions

Est-il possible d'améliorer la mesure des opportunités d'apprentissage en mathématiques en Fédération Wallonie-Bruxelles, en prenant en considération les styles de réponse des élèves, afin d'établir un lien entre les OTL et la performance en culture mathématique des élèves ?

Au regard des recherches que nous avons entreprises, il est difficile de répondre de façon claire à notre question de départ. Cependant, nous pouvons apporter un certain nombre de pistes de réponse, en lien avec chacune de nos quatre hypothèses.

La première hypothèse que nous avons formulée était relative à la distribution des opportunités d'apprentissage en Fédération Wallonie-Bruxelles, que nous supposions inégale. Cette hypothèse s'est vue confirmée : **en Fédération Wallonie-Bruxelles, les élèves ne reçoivent pas tous les mêmes chances d'apprendre selon qu'ils sont à l'heure ou en retard, selon la filière qu'ils fréquentent et selon leur statut socioéconomique.**

Il existe, en FWB, des corrélations statistiquement significatives, de l'ordre de 0,35, entre les OTL que les élèves déclarent recevoir dans le cadre des cours de mathématiques, et le fait qu'ils aient connu ou non le redoublement, ou qu'ils fréquentent l'enseignement de transition ou de qualification. De la même façon, il existe des différences significatives entre les OTL moyennes déclarées par les élèves à l'heure ou en retard, comme entre les élèves inscrits dans le général et ceux qui fréquentent l'enseignement technique ou professionnel. Cela indique clairement que les élèves à l'heure déclarent recevoir davantage de chances d'apprendre des contenus mathématiques que ceux qui ont fait l'expérience du redoublement durant leur scolarité. Il en va de même pour les élèves de l'enseignement de transition, qui déclarent recevoir davantage d'OTL que ceux de l'enseignement qualifiant.

Ces différences s'expliquent en partie par la manière dont est construit l'indice OTL utilisé pour vérifier cette première hypothèse. En effet, il s'agit d'un indice de familiarité avec des concepts mathématiques. Or, en FWB, nombre de ces concepts ne sont développés qu'à partir de la quatrième année de l'enseignement secondaire, et certains ne sont abordés que dans l'enseignement de transition. On comprend dès lors les différences observées entre les différents groupes. Néanmoins, ces différences sont révélatrices de la politique de séparation entre les élèves qui a cours en Fédération Wallonie-Bruxelles.

Ces différences quant aux contenus enseignés sont une piste qui explique les grandes différences de performance sur l'échelle PISA de culture mathématique entre ces groupes. Or, l'évaluation de cette culture mathématique se fait dans une visée prospective, afin de mesurer si les jeunes de quinze ans seront à même de s'insérer dans la société. Quand on prend ce fait en considération, on peut se demander si la Fédération Wallonie-Bruxelles remplit les missions prioritaires qu'elle a fixées pour son enseignement, en particulier celles consistant à « préparer tous les élèves à être des citoyens responsables » et « assurer à tous les élèves des chances égales d'émancipation sociale » (FWB, 1997).

Pour ce qui est de l'effet du statut socioéconomique de l'élève sur les opportunités d'apprentissage qu'il déclare recevoir, la corrélation, de l'ordre de 0,25 est plus faible que pour les deux variables précédentes. De même, la différence de moyenne entre les OTL déclarées par les 25% d'élèves les plus favorisés et les 25% d'élèves les plus défavorisés est plus faible, mais elle reste significative, tout comme la corrélation. Les élèves les plus défavorisés sont donc ceux qui déclarent recevoir le moins d'opportunités d'apprentissage, ce qui indique une forme de ghettoïsation de l'enseignement en Fédération Wallonie-Bruxelles.

Néanmoins, dans la mesure où les élèves défavorisés sont souvent ceux qui fréquentent l'enseignement de qualification et/ou qui accusent un retard scolaire, les coefficients de corrélation ne reflètent pas l'effet propre de chacune des variables sur les OTL déclarées. Il conviendrait pour cela de calculer les effets nets et joints de ces variables, afin de différencier la part de chacune.

Notre seconde hypothèse portait sur les liens qui peuvent être établis entre les opportunités d'apprentissage et la performance en mathématiques, lorsqu'on prend en considération les biais de styles de réponse comme c'est le cas dans l'enquête PISA 2012. Notre postulat était que la prise en compte des styles de réponse devait améliorer la mesure des opportunités d'apprentissage entre les pays et au sein de ceux-ci.

Cette hypothèse n'a pas été complètement confirmée. En effet, la façon dont les styles de réponse sont contrôlés dans l'enquête PISA 2012 (en soustrayant la moyenne des trois items pièges à l'indice FAMCON pour construire l'indice FAMCONC) a des résultats très variables d'un pays à l'autre. Pour certains pays, la différence est nulle, alors que dans d'autres, elle est soit positive, soit négative. En revanche, la corrélation, pour l'ensemble des 41 pays du groupe, entre FAMCONC et la performance en mathématiques est plus élevée que celle calculée avec FAMCON.

Il semble donc que l'indice FAMCONC capte bien une partie des styles de réponse qui affectent les élèves. Cependant, vu les faibles différences observées, il semble que cette méthode ne soit pas complètement efficace. C'est pourquoi nous avons eu recours aux analyses factorielles pour améliorer davantage la mesure des opportunités d'apprentissage.

Néanmoins, dans la mesure où nous ne sommes pas parvenu à trouver une modélisation factorielle satisfaisante qui permette de mettre sous contrôle les biais de styles de réponse pour les élèves de la Fédération Wallonie-Bruxelles, **l'indice FAMCONC demeure la meilleure façon d'y arriver, faisant passer la corrélation entre les OTL et la performance de 0,42 à 0,50, en FWB.**

En troisième lieu, nous nous sommes interrogé sur la façon dont la performance académique des élèves peut être reliée à leur propension à être affectés par les styles de réponse. Nous nous attendions, conformément aux travaux déjà menés dans ce sens (Grisay, 2009), à pouvoir mettre en évidence une relation négative entre la performance académique moyenne des pays et la tendance des élèves de ce pays à répondre pour autre chose que ce que la question prétend mesurer. Tel fut en effet le cas.

Entre les pays, il existe une corrélation statistiquement significative, bien que faible (-0,13) entre la performance en culture mathématique et le score sur les items pièges de la question ST62. **Les élèves des pays les moins performants sont ceux qui recourent le plus aux styles de réponse.** Cependant, cette corrélation est peu consistante et les corrélations à l'intérieur des pays diffèrent fortement.

Dans la plupart des pays, les corrélations entre la performance et les styles de réponse sont également négatives. C'est notamment le cas en Fédération Wallonie-Bruxelles. Cela signifie que, dans un pays donné, ce sont bien les élèves les moins performants qui sont le plus affectés par le biais de styles de réponse. Quelques pays (8 sur 41) font exception à cette règle et présentent des corrélations positives. Ces pays sont en outre ceux qui ont les performances moyennes les plus faibles du groupe de référence. Les facteurs permettant d'expliquer cette variabilité dans les coefficients de corrélation à l'intérieur des pays mériteraient, selon nous, de faire l'objet de plus d'approfondissements.

En outre, toujours au sein de la Fédération Wallonie-Bruxelles, il existe également des corrélations significatives entre le biais de styles de réponse qui affecte un élève et le fait qu'il ait redoublé ou qu'il fréquente l'enseignement de qualification. Les élèves ayant connu le

redoublement, tout comme ceux de l'enseignement de qualification, sont davantage touchés par les styles de réponse. Cela peut témoigner chez ces élèves d'une forme de comportement satisfaisant, qui les conduit à moins s'impliquer dans les questionnaires contextuels comme ceux de PISA. Cela peut aussi être relié avec leur performance académique, puisque ce sont ces élèves qui performant le moins bien en culture mathématique. À nouveau, comme pour la distribution des OTL, il serait intéressant de calculer les effets nets et joints de chaque variable sur les styles de réponse.

La quatrième et dernière partie de notre démarche de recherche consistait à essayer d'expliquer les biais de styles de réponse grâce à une analyse factorielle. L'hypothèse sous-jacente était qu'en faisant cela, il serait possible d'améliorer la mesure des opportunités d'apprentissage en FWB. Cela ne fut hélas pas le cas. **Aucun des modèles éprouvés n'a permis de construire un indice de styles de réponse qui puisse améliorer la mesure des OTL en Fédération Wallonie-Bruxelles.** Néanmoins, les modèles que nous avons testés se sont tout de même révélés prometteurs et pourraient être réutilisés.

En particulier, notre modèle 2 indique qu'il serait possible de contrôler les styles de réponse à l'échelle de la Belgique afin d'améliorer, peut-être, la mesure des opportunités d'apprentissage. Il nous paraîtrait dès lors intéressant de pousser les recherches plus loin en ce sens, afin d'apporter des réponses aux questions que notre recherche laisse en suspens.

De la même manière, notre troisième modèle a mis en lumière le fonctionnement différentiel des items selon la filière et le caractère redoublant de l'élève. Cela fait écho à notre troisième hypothèse, qui a également pointé que les élèves ne sont pas tous égaux face aux styles de réponse. Ce modèle suggère donc que, **dans des systèmes éducatifs qui pratiquent un modèle de la séparation, la prise en considération des styles de réponse, pour qu'elle soit efficace, devrait tenir compte des filières d'enseignement et du redoublement.** En revanche, comme l'a montré notre exemple avec l'Islande, cela ne semble pas nécessaire dans les modèles intégratifs.

Enfin, notre quatrième modèle, bien qu'il n'ait pas permis de répondre à nos questions, a permis de répliquer la recherche menée par Eva Van de Gaer (2010). Cela confirme une fois de plus l'impact des styles de réponse (ici, de la désirabilité sociale) sur les variables contextuelles. Il est regrettable de ne pas avoir pu faire coïncider ce modèle avec les opportunités d'apprentissage à cause d'un obstacle logistique, mais cela ouvre la voie à de futures recherches.

En définitive, nous retenons de cette recherche les éléments suivants.

- Les opportunités d'apprentissage de contenu sont très inégalement distribuées entre les élèves de la Fédération Wallonie-Bruxelles, ce qui reflète bien l'iniquité de notre système éducatif.
- Les opportunités d'apprentissage, comme les autres variables contextuelles, sont entachées par un biais de styles de réponse qu'il convient de mettre sous contrôle afin d'en garantir une mesure valide.
- Faute de mieux, la manière dont les styles de réponse qui touchent les opportunités d'apprentissage sont envisagés dans PISA 2012 est la meilleure dont nous disposons.
- Les élèves ne sont pas tous égaux face aux styles de réponse : les élèves les plus faibles y sont davantage exposés que les élèves plus performants.
- Il n'a pas été possible de modéliser les styles de réponse portant sur les opportunités d'apprentissage à l'échelle de la Fédération Wallonie-Bruxelles.
- Une modélisation efficace des styles de réponse devrait tenir compte des filières et du redoublement, car ils affectent la manière dont les élèves peuvent répondre pour autre chose que ce que la question prétend mesurer.

5 Limites et perspectives

La première et plus importante limite de ce travail de recherche est qu'il ne permet pas de répondre à la question que nous nous étions fixée. En effet, nous ne sommes pas parvenu à améliorer la mesure des opportunités d'apprentissage en Fédération Wallonie-Bruxelles grâce à l'analyse factorielle. Cependant, nous restons confiant sur la possibilité d'y arriver et nous espérons que cela pourra être accompli dans le futur, soit à une autre échelle, soit sur des données issues d'une autre enquête que PISA 2012. De plus, cette absence de résultats ouvre la voie à de nouvelles perspectives de recherche.

La seconde limite que nous pointons dans notre recherche est qu'elle s'est attachée à ne présenter des résultats que pour la Fédération Wallonie-Bruxelles. Bien que cela en diminue la portée, c'était une volonté de notre part. Nous l'avons fait afin d'apporter des éléments de réponse propres à notre système éducatif, d'une part, mais aussi pour ouvrir la voie à une recherche plus élargie, qui investiguerait les liens entre opportunités d'apprentissage, styles de réponse et performance en culture mathématique à l'échelle internationale.

Une troisième limite dans ce travail apparaît dans la manière dont nous avons étudié les liens entre la distribution des opportunités d'apprentissage en Fédération Wallonie-Bruxelles et les filières d'enseignement, le redoublement et le statut socioéconomique de l'élève. En effet, bien que nous ayons mis en lumière que ces liens existent, nous n'avons pas calculé les effets nets et joints de chacune de ces variables. Or, comme nous l'avons déjà signalé, elles connaissent sans doute une certaine redondance.

Dès lors, il pourrait être intéressant de parler « toute chose étant égale par ailleurs », afin de savoir, par exemple, si le statut socioéconomique de l'élève a encore une influence sur les OTL qu'il déclare recevoir, lorsque l'on place le redoublement et les filières sous contrôle. Un raisonnement similaire est en outre possible pour les liens qui existent entre le biais de styles de réponse qui affecte les élèves, et leur caractère redoublant, leur fréquentation de l'enseignement qualifiant, et leur performance académique.

Notre recherche offre donc un certain nombre de pistes de prolongation, qui sont autant d'opportunités pour de futurs travaux de recherche. Une possibilité serait de repartir de notre second modèle d'analyse factorielle. En effet, ce modèle suggère qu'il est possible de mettre sous contrôle les styles de réponse afin d'améliorer la mesure des opportunités d'apprentissage

à l'échelle de la Belgique. Il serait donc envisageable de repartir de notre travail pour répondre à notre question, mais sans différencier les trois Communautés linguistiques du pays.

Une recherche ultérieure qui voudrait apporter des réponses à nos questions pourrait également s'inspirer du *superconstruct* développé par Lie et Turmo (2005, cités par Lafontaine, 2017). Cette idée de faire saturer un grand nombre de variables contextuelles sur un seul et même facteur, qui reflèterait les styles de réponse des élèves, nous semblait initialement très intéressante. Cependant, nous ne l'avons pas développée dans notre méthodologie. Il pourrait donc être judicieux de vérifier dans quelle mesure la logique de Lie et Turmo pourrait être appliquée aux opportunités d'apprentissage.

La reproduction du Modèle de Van de Gaer (2010) constitue également une piste d'ouverture à de futurs travaux. Nous avons en effet reproduit cette recherche au départ des données issues de PISA 2012, confirmant bien que la désirabilité sociale affecte certaines variables de contexte. Cependant, nous n'avons pas pu étendre cette recherche aux opportunités d'apprentissage de contenu, définies au départ de la question ST62. Il pourrait dès lors être proposé d'essayer de le faire avec un des autres indices d'OTL présents dans le questionnaire contextuel de PISA 2012, pour autant qu'il apparaisse dans la seconde forme dudit questionnaire.

Enfin, notre recherche pourra être répliquée au départ des données issues de PISA 2021. En effet, lors de ce cycle, la culture mathématique constituera à nouveau le domaine majeur. Il y a dès lors de fortes chances que les OTL en mathématiques fassent à nouveau partie des variables mesurées dans les questionnaires contextuels. Il reste à espérer que la distribution des items entre les différentes formes du questionnaire permettra de comparer les données des variables FAMCON, INTMAT, MATHEFF et SCMAT. Cela constituerait en effet une opportunité d'ajouter les OTL au modèle développé par Van de Gaer (2010) et que nous avons reproduit.

Bibliographie

- Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English Language Learners and Math Achievement : A Study of Opportunity to Learn and Language Accommodation*. (CSE Report 702, 2006). Los Angeles, CA : Université de Californie, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.
- Baye, A., Quittre, V., Hindryckx, G., Fagnant, A., & Lafontaine, D. (2007). *Les acquis des élèves en culture scientifique. Premiers résultats de PISA 2006*. Université de Liège, Service d'analyse des Systèmes et des Pratiques d'enseignement (aSPe).
- Byrne, B. (2012). *Structural equation modeling with Mplus : Basic concepts, applications, and programming*. (Multivariate applications series). New York : Routledge Academic.
- Crahay, M. (1999). *Psychologie de l'éducation*. Paris : Presses Universitaires de France.
- Danhier, J., Jacobs, D., Devleeshouwer, P., Martin, É., & Alarcon, A. (2014). *Vers des écoles de qualité pour tous ? Analyse des résultats à l'enquête PISA 2012 en Flandre et en Fédération Wallonie-Bruxelles*. Bruxelles : Fondation Roi Baudouin.
- Demonty, I., Blondin, C., Matoul, A., Baye, A., & Lafontaine, D. (2013). La culture mathématique à 15 ans Premiers résultats de PISA 2012 en Fédération Wallonie-Bruxelles. *Cahiers des Sciences de l'Éducation – Université de Liège (aSPe)*, 34, pp. 1-26.
- Fagnant, A., & Demonty, I. (2018). *Enseignement et apprentissage des mathématiques dans l'enseignement fondamental et secondaire inférieur, Partim 1*. Notes de cours non publiées, Université de Liège, Liège.
- Fédération Wallonie-Bruxelles. (1997). *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. Bruxelles : Communauté française de Belgique, Ministère de l'Éducation, de la Recherche et de la Formation.
- Fédération Wallonie-Bruxelles. (2013). *Les indicateurs de l'enseignement*. (8è éd.). Bruxelles : Administration générale de l'Enseignement.

- Fédération Wallonie-Bruxelles. (2017). *Pacte pour un Enseignement d'Excellence. Avis n° 3 du Groupe central*. Bruxelles : Administration générale de l'Enseignement.
- Fédération Wallonie-Bruxelles. (2019). *Les indicateurs de l'enseignement*. (13^e éd.). Bruxelles : Administration générale de l'Enseignement.
- Grisay, A. (2009). *Draft on social desirability in PISA 2006*. Document non publié, Université de Liège, Liège.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study : investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Siedel (Eds), *The power of Video studies in investigating teaching and learning in the classrooms* (pp. 137-160). Waxmann : Münster.
- Lafontaine, D. (2016). *Approche expérimentale des pratiques et des politiques éducatives*. Notes de cours non publiées, Université de Liège, Liège.
- Lafontaine, D. (2017). Évaluations à large échelle : Prendre la juste mesure des effets de contexte. In Detroz, P., Crahay, M. & Fagnant, A. (Eds.), *L'évaluation à la lumière des contextes et des disciplines* (pp. 21-51). Belgique, Louvain-La-Neuve : De Boeck.
- Lafontaine, D., Baye, A., Vieluf, S., & Monseur, C. (2015). Equity in opportunity-to-learn and achievement in reading : A secondary analysis of PISA 2009 data. *Studies in Educational Evaluation*, 47, pp. 1-11. doi : 10.1016/j.stueduc.2015.05.001
- Marcoux, G. (2014). Résolution de problèmes arithmétiques dans le cadre d'une approche par compétences : ordre de grandeur des tâches et parts d'influence de quelques facteurs cognitifs et motivationnels. *Cahiers des Sciences de l'Éducation – Université de Liège (aSPe)*, 36, pp. 67-114.
- McDonnell, Lorraine M. (1995). Opportunity to Learn as a Research Concept and a Policy Instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305-322.
- Mons, N. (2007). *Les nouvelles politiques éducatives : la France fait-elle les bons choix ?* Paris : Presses Universitaires de France.
- Monseur, C. (2016). *Questions d'évaluation*. Notes de cours non publiées, Université de Liège, Liège.

- Monseur, C. (2017). *Évaluation des systèmes éducatifs par les enquêtes internationales*. Notes de cours non publiées, Université de Liège, Liège.
- OCDE. (2013). *Cadre d'évaluation et d'analyse du cycle PISA 2012. Compétences en mathématiques, compréhension de l'écrit, en sciences, en résolution de problèmes et en matières financières*. Paris : OCDE.
- OCDE. (2014). *PISA 2012 Technical Report*. Paris : OCDE.
- Scherff, L., & Piazza, C. L. (2008). Why Now, More Than Ever, We Need to Talk About Opportunity to Learn. *Journal of Adolescent & Adult Literacy*, 52(4), pp. 343-352. doi : 10.1598/JAAL.52.4.7
- Schmidt, W. H., & Maier, A. (2006). Opportunity to Learn. In G. Sykes, B. Schneider & D. N. Plank, *Handbook of Education Policy Research* (pp. 541-559). New York, NY : Routledge.
- Schmidt, W. H., Zoido, P., & Cogan, L. (2014). *Schooling matters : Opportunity to Learn in PISA 2012. OECD Education Working Papers No. 95*. Paris : OCDE.
- Stegen, P., Géron, C., Daro, S., & Desmet, L. (n.d.). *Liaison primaire-secondaire. L'enseignement des rationnels*. ASBL Hypothèse, Liège.
- Van de Gaer, E. (2010). *The Modeling of Response Style Bias : an Answer to the Attitude-Achievement Paradox ?* Technical Advisory Group Meeting. Australian Council for Educational Research, Melbourne.
- Van de Gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The Reference Group Effect : An Explanation of the Paradoxical Relationship between Academic Achievement and Self-Confidence across Countries. *Journal of Cross-cultural Psychology*, 43, pp. 1205-1228. doi : 10.1177/0022022111428083
- Wang, J. (1998). Opportunity to Learn : The Impacts and Policy Implications. *Educational Evaluation and Policy Analysis*, 20(3), 137–156.
- Weijters, B. (2006). *Response Styles in Consumer Research* (Doctoral dissertation). Université de Gand, Gand.

Table des illustrations

Figures

Figure 1 : cycle de modélisation mathématique dans PISA2012.	6
Figure 2 : diagramme de modélisation mathématique de Verschaffel.	6
Figure 3 : différences de scores moyens en culture mathématique entre différentes catégories d'élèves à PISA 2012	12
Figure 4 : items de la question ST61 — expérience avec des tâches mathématiques.....	19
Figure 5 : items de la question ST62 — familiarité avec des concepts mathématiques.	20
Figure 6 : items de la question ST74 — expérience de différents types de problèmes mathématiques, tâches procédurales.	21
Figure 7 : exemple d'items de la question ST79 — comportement de l'enseignant	21
Figure 8 : items de la question ST80 — activation cognitive.	22
Figure 9 : taux de correspondance et corrélation entre les différentes sources de données de l'étude SII 2004.	25
Figure 10 : corrélation entre le concept de soi et les performances en lecture au niveau pays à PIRLS 2011.	36
Figure 11 : analyse de composant principal sur dix variables contextuelles de PISA 2006. ...	38
Figure 12 : structure factorielle utilisée par Van de Gaer (2010).....	40
Figure 13 : pourcentages de réponses aux différentes modalités des items ST62Q04, ST62Q11 et ST62Q13 en FWB.	59
Figure 14 : pourcentages de réponses aux différentes modalités des items ST62Q04, ST62Q11 et ST62Q13 à l'échelle internationale.	59
Figure 15 : corrélation entre BIAS et la performance en culture mathématique.	60
Figure 16 : saturations sur les trois facteurs du modèle 1, EFA.....	63
Figure 17 : indices d'ajustement du modèle 1, CFA.....	64
Figure 18 : indices d'ajustement du modèle 2, CFA.....	66
Figure 19 : indices d'ajustement du modèle 3, groupe 1, CFA.....	68
Figure 20 : indices d'ajustement du modèle 3, groupe 2, CFA.....	72
Figure 21 : indices d'ajustement du modèle 3, groupe 3, CFA.....	74
Figure 22 : indices d'ajustement du modèle 3, Islande, CFA.	76
Figure 23 : indices d'ajustement du modèle reproduit d'après Van de Gaer (2010).	78

Tableaux

Tableau 1 : design du plan d'évaluation incomplet de PISA 2012	11
Tableau 2 : comparaison entre l'échantillon PISA 2012 en FWB et les données administratives de la population-cible, par forme et année d'enseignement.	47
Tableau 3 : signification des variables utilisées	50
Tableau 4 : corrélations et différences de moyennes entre FAMCONC et QUALIFIANT, REDOUBLE et ESCS	51
Tableau 5 : corrélations entre la performance des élèves et FAMCON et FAMCONC dans 41 pays.....	54
Tableau 6 : performance en culture mathématique et biais de styles de réponse dans 41pays.	58
Tableau 7 : corrélations entre BIAS et QUALIFIANT, REDOUBLE et la performance en culture mathématique en FWB.	61
Tableau 8 : structure factorielle du modèle 1, CFA.	65
Tableau 9 : structure factorielle du modèle 2, CFA.	66
Tableau 10 : structure factorielle du modèle 3, groupe 1, CFA.	69
Tableau 11 : matrice de corrélations dans le groupe 1.	70
Tableau 12 : paramètres des analyses de régression réalisées sur le groupe 1.....	71
Tableau 13 : structure factorielle du modèle 3, groupe 2, CFA.	73
Tableau 14 : structure factorielle du modèle 3, groupe 3, CFA.	75
Tableau 15 : répartition des items ST29, ST37 et ST42 dans les différentes formes des questionnaires contextuels adressés aux élèves.....	77
Tableau 16 : structure factorielle du modèle reproduit d'après Van de Gaer (2010).....	77
Tableau 17 : corrélation entre RS_BIAS_VDG et la performance en culture mathématique pour les pays de l'OCDÉ.	79
Tableau 18 : corrélations entre INTMAT et la performance en culture mathématique dans 37 systèmes éducatifs, sans avec prise en compte des styles de réponse.	81
Tableau 19 : corrélations entre RS_BIAS_VDG et FAILMAT, ANXMAT et INSTMOT. ...	82

Résumé de la recherche

Le présent mémoire investigate les liens qui existent entre la performance en mathématiques des élèves de quinze ans en Fédération Wallonie-Bruxelles à l'enquête PISA 2012, les opportunités d'apprentissage qui leur sont données dans le cadre des cours de mathématiques et les biais de styles de réponse qui découlent de la façon dont les données de contexte sont récoltées dans les enquêtes à large échelle.

Mesurer la performance des élèves est le souci des enquêtes internationales depuis les années 1960. Cette recherche se base sur les données issues de PISA 2012, cinquième cycle de cette enquête à large échelle menée par l'OCDE, dans laquelle la culture mathématique constituait le domaine majeur.

Les opportunités d'apprentissage sont envisagées selon l'axe du contenu dans cette recherche, et mesurées dans le questionnaire contextuel adressé aux élèves en termes de familiarité avec un ensemble de concepts mathématiques variés (cosinus, fonction du second degré, polygone, radicaux, nombres complexes...). Il s'agit donc de mesurer si les élèves de quinze ans ont eu l'occasion, lors des cours de mathématiques, de se familiariser avec ces concepts, donc d'être plus à même de les maîtriser.

Les styles de réponse sont un biais méthodologique qui naît de la manière dont les mesures de variables de contexte sont réalisées, au travers d'échelles de Likert. Ils sont une tendance des individus à répondre pour autre chose que ce que la question prétend mesurer, sur base de facteurs individuels ou culturels, par manque d'investissement ou à cause d'une mauvaise maîtrise de la langue de la question. Ils sont une cause bien établie du paradoxe attitude-performance. Vouloir les contourner nécessite d'avoir recours à des traitements statistiques adaptés.

La partie pratique de cette recherche s'applique à mettre en évidence comment les opportunités d'apprentissage peuvent être reliées à la performance des élèves de la Fédération Wallonie-Bruxelles et comment la prise en compte des styles de réponse peut améliorer la mesure de ces opportunités d'apprentissage.