

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRODUCTION AUTOMATISÉE DE TÂCHES D'ÉVALUATION EN ANGLAIS,  
LANGUE SECONDE

RAPPORT

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN ÉDUCATION, PROFIL INTERVENTION

PAR  
CRISTINA CARLOMUSTO

FÉVRIER 2013

J'aimerais dédier ce rapport d'activités à ma mère Marcella.  
Sans toi, je n'aurais pu compléter ces études. Du fond de mon cœur, merci.

## **REMERCIEMENTS**

J'aimerais tout d'abord remercier mon directeur de recherche, le professeur Gilles Raïche et mon codirecteur de recherche, le professeur Martin Riopel sans qui la réalisation de ce projet n'aurait pu avoir lieu. Plus particulièrement, je tiens à remercier profondément mon directeur de recherche qui a su m'aider et me guider tout au long de mes études. Ses conseils, son écoute et sa précieuse collaboration ont été un atout durant la réalisation de ce projet d'intervention.

Par ailleurs, j'aimerais remercier Sébastien Béland, qui m'a été d'une grande aide lors de mon analyse de données. Merci énormément pour ta patience, ton aide et ton support durant ce moment. Je tiens aussi à remercier Catherine Croisetière, qui m'a aidée lors de la correction de ce rapport.

Je voudrais également remercier les deux enseignantes qui m'ont permis de faire l'intervention dans leur classe, Christine Lafleur et Nathalie l'Écuyer, ainsi que les directrices des deux écoles, Manon Laporte et Marie Peloquin.

Je n'oublie pas de remercier sincèrement ma famille pour leur soutien constant et leurs encouragements tout au long de mon projet d'intervention.

## TABLE DES MATIÈRES

REMERCIEMENTS.....	III
TABLEAU DES MATIÈRES.....	IV
LISTE DES TABLEAUX.....	VI
RÉSUMÉ.....	VII
INTRODUCTION.....	1
CHAPITRE I	
PROBLÉMATIQUE.....	3
1.1 L'enseignement de l'anglais langue seconde.....	3
1.2 Les composantes de la lecture.....	4
1.3 La compréhension de textes et l'anglais langue seconde.....	5
1.4 But et pertinence du projet d'intervention.....	7
CHAPITRE II	
CONTEXTE THÉORIQUE.....	8
2.1 Le modèle de Bormuth (1970).....	8
2.2 Concepts sur la production d'items.....	10
2.3 Études sur la production d'items.....	12
2.4 Exemple d'application sur la production d'items : le modèle de Cobb.....	15
2.5 Exemple d'application sur la production d'items : le modèle de Stenner.....	18
2.6 Forces et limites des articles présentés.....	19
2.7 Objectifs spécifiques.....	23
CHAPITRE III	
MÉTHODOLOGIE.....	24
3.1 Sujets.....	24
3.2 Déroulement.....	26
3.3 Description de l'instrument.....	26
3.3.1 Élaboration de l'instrument.....	26
3.3.2 Choix des textes en compréhension de lecture.....	27
3.3.3 Choix des items.....	28
3.4 Méthode d'analyse choisie.....	28
3.5 Considérations éthiques.....	31

CHAPITRE IV	
RÉSULTATS.....	32
4.1 Analyses descriptives.....	32
4.2 Analyse de la précision de mesure et des coefficients de détermination.....	33
4.3 Analyses de régression linéaire.....	34
4.3.1 Analyse de régression linéaire multiple pour les textes.....	35
4.3.2 Analyse de régression linéaire multiple pour les questions.....	37
4.3.3 Analyse de régression linéaire multiple pour les choix de réponses.....	39
CHAPITRE V	
DISCUSSION.....	42
5.1 Explications concernant les textes.....	42
5.2 Explications concernant les questions.....	44
5.3 Explications concernant les choix de réponses.....	46
CONCLUSION.....	49
RÉFÉRENCES.....	53
ANNEXE A : Formulaire de consentement pour les directions d`école.....	55
ANNEXE B : Formulaire de consentement pour les élèves.....	59
ANNEXE C : Informations sur les calculs pour l`échelle de mesure et exemple d'application.....	63

## **LISTE DES TABLEAUX**

Tableau 1.	Distribution des sujets selon l'école.....	24
Tableau 2.	Distribution des sujets selon le genre.....	25
Tableau 3.	Distribution des sujets selon le moment de suivre le programme d'anglais intensif.....	25
Tableau 4.	Fréquence des élèves qui ont réussi chacun des items et proportion des 40 items, indice de difficulté, pour l'évaluation en lecture.....	32
Tableau 5.	Valeur de la corrélation item-total, indice de discrimination pour chacun des items utilisés lors de l'évaluation en lecture.....	34
Tableau 6.	Coefficients de corrélation et de détermination pour la régression multiple des textes.....	35
Tableau 7.	Coefficients de corrélation et de détermination pour la régression multiple pour prédire l'indice de difficulté des textes.....	37
Tableau 8.	Coefficients de corrélation et de détermination pour la régression multiple des questions.....	38
Tableau 9.	Coefficients de corrélation et de détermination pour la régression multiple pour prédire l'indice de difficulté des questions.....	39
Tableau 10.	Coefficients de corrélation et de détermination pour la régression multiple des choix de réponses.....	39
Tableau 11.	Coefficients de corrélation et de détermination pour la régression multiple pour prédire l'indice de difficulté des choix de réponses.....	40
Tableau 12.	Exemple de trois items pour la mise en œuvre du modèle de prédiction...	66
Tableau 13.	Valeur de chacun des trois items pour les différents aspects du modèle de prédiction.....	66

## RÉSUMÉ

Ce projet d'intervention s'intéresse à l'identification de variables qui peuvent être liées au niveau de difficulté de tâches en compréhension de lecture pour les élèves du troisième cycle du primaire en anglais langue seconde. Plus précisément, ce projet d'intervention propose un modèle qui permet de prédire le niveau de difficulté de tâches présentées aux élèves.

Afin d'atteindre notre objectif, un outil d'évaluation en compréhension de textes a été mis sur pied et soumis à des élèves du troisième cycle du primaire provenant de différents quartiers de l'île de Montréal. À partir des données recueillies, différentes composantes d'une tâche en compréhension de textes ont été retenues comme variables de prédiction du niveau de difficulté. Ensuite, nous avons appliqué un modèle de régression multiple sur les variables retenues afin d'identifier quelles variables pouvaient prédire le niveau de difficulté d'une tâche.

Les résultats obtenus démontrent que certaines variables semblent en lien avec le niveau de difficulté de la tâche. Premièrement, l'ordre dans lequel nos questions sont posées dans le texte influence la proportion de réussite. De plus, lorsque les élèves possèdent suffisamment d'information provenant du texte, ils ont plus de facilité à compléter les choix de réponses. Troisièmement, la difficulté des choix de réponses influence la réussite aux diverses compréhensions de textes présentés dans le texte. Ensuite, lorsque les élèves doivent faire preuve d'inférence, ceci vient augmenter le niveau de difficulté. Enfin, en ce qui concerne les textes, aucune variable retenue n'a pu être démontrée statistiquement significative et utile pour déterminer l'indice de difficulté des items.

Au niveau pratique, les enseignants pourraient utiliser le modèle développé dans leur salle de classe afin de prédire le niveau de difficulté de leurs propres tâches en compréhension de textes. Ils pourraient par la suite choisir les meilleures tâches afin de présenter une évaluation qui sera mieux adaptée au niveau scolaire de leurs élèves.

Mots clés: élaboration de tâche d'évaluation, compréhension de textes, niveau de difficulté

## INTRODUCTION

Dans notre société, la lecture est une habileté à laquelle nous ne pouvons échapper. Dans un texte de Nichols (2007), on nous explique qu'une des fonctions premières de la lecture est de permettre à chaque citoyen de faire face aux diverses situations qui vont se présenter dans son environnement, surtout lorsque nous apprenons une langue seconde. Il est donc essentiel que la compréhension de textes prenne une place importante dans le curriculum puisque la lecture n'est pas seulement utilisée à des fins scolaires.

Lorsqu'un étudiant fait face à un texte, plusieurs variables peuvent influencer la façon de l'interpréter, ce qui aura un impact sur sa compréhension et donc, du même coup, sur ses réponses. Lorsqu'on désire observer les variables qui peuvent influencer la compréhension d'un texte, il est important de considérer tous les aspects qui englobent la compréhension de la lecture c'est-à-dire le texte, les questions et les choix de réponses, afin d'obtenir un bon aperçu de ce qui se produit chez les élèves.

Le projet d'intervention qui suit vise à observer ces différentes variables et identifier lesquelles peuvent influencer la compréhension de textes. Tout ceci se fera dans le but de proposer un modèle qui permettra de prédire le niveau de difficulté des tâches en compréhension de textes. Le modèle obtenu pourra être utilisé par les enseignants du milieu scolaire lors de la réalisation de tâches d'évaluations avec leurs élèves.



Pour ce faire, toutes nos analyses seront basées sur la théorie classique des tests. Dans un premier temps, des calculs des paramètres des items pour connaître les indices de discrimination et de difficulté seront effectués. Ensuite, nous procéderons à différentes analyses de régression multiple afin de savoir quelles variables peuvent être utilisées pour la prédiction du niveau de difficulté.

Dans le rapport d'intervention qui suit, nous tenterons d'une part, au chapitre I, d'expliquer l'origine du problème ainsi que sa pertinence. Ensuite, le chapitre II permettra de comprendre quels sont les concepts reliés à la production d'items et quels exemples d'application existent dans ce domaine. Une synthèse sera enfin présentée pour introduire nos objectifs de recherche. Le chapitre III présentera les aspects méthodologiques de ce projet d'intervention : les sujets, le déroulement, l'instrumentation, les méthodes d'analyses ainsi que les considérations éthiques. Les résultats de notre intervention ainsi que la discussion sont finalement présentés aux chapitres IV et V.

En guise de conclusion, nous ferons une synthèse de notre rapport d'intervention afin de proposer des recommandations pour des recherches ultérieures. À l'annexe A, on retrouve le formulaire de consentement pour les directions d'écoles. Le formulaire de consentement pour les élèves est présenté à l'annexe B. Pour terminer, à l'annexe C, on trouve les informations sur les calculs de l'échelle de mesure ainsi qu'un exemple d'application de cette échelle.

## CHAPITRE I

### PROBLÉMATIQUE

L'éducation et l'enseignement sont deux aspects qui sont constamment en évolution. D'une année à l'autre, les élèves qui sont présents dans les écoles ne sont pas les mêmes et n'ont forcément pas les mêmes besoins. Il est ainsi important pour le milieu de l'enseignement de se renouveler constamment afin de bien répondre aux besoins de sa clientèle. Depuis quelques années, l'enseignement de l'anglais langue seconde au Québec a beaucoup progressé. Avec le *Programme de formation de l'école québécoise* (ministère de l'Éducation, des Loisirs et du Sport, 2001), l'enseignement de cette matière prend de plus en plus de place dans nos écoles primaires. Plus d'heures d'enseignement y sont accordées, notamment par l'ajout de l'enseignement de cette discipline au premier cycle et par la mise en place du programme d'anglais intensif qui aborde la matière de façon plus concentrée.

#### 1.1 L'enseignement de l'anglais, langue seconde, au primaire

Au troisième cycle du primaire, les élèves se retrouvent avec un programme d'anglais riche et complexe qui les aidera à développer les trois compétences du programme qui seront utiles pour leurs études futures. Ces trois compétences sont : communiquer à l'oral en anglais (*to interact orally*), réinvestir la compréhension des textes écrits et des textes entendus (*to reinvest the understanding of oral and written text*) et écrire des textes en anglais (*to write text*) (ministère de l'Éducation, des Loisirs et du Sport, 2001, p. 99-105). De ces trois compétences, c'est la deuxième qui est la plus problématique pour nos élèves.

Le fait de lire des textes peut être considéré comme une tâche assez facile à exécuter pour certains individus, mais c'est surtout l'interprétation des mots qui peut devenir très difficile pour certaines personnes. Les élèves en question font habituellement l'apprentissage d'une deuxième langue et, parfois, d'une troisième ou même d'une quatrième langue. De plus, si ces élèves font face à des problèmes concernant la compréhension de textes en français, il y a de fortes chances que ces difficultés se manifestent également en anglais.

## 1.2 Les composantes de la lecture

On considère souvent la lecture comme étant quelque chose d'assez simple et on y porte peu attention. Cependant, lorsqu'on se met à regarder tous les aspects qui entourent celle-ci, on constate qu'elle est beaucoup plus complexe qu'elle ne paraît. Ceci peut être expliqué à l'aide du *Searchlights model* de Clay et Cazden (1990), qui est présenté par Snowling, Stainthorp et Stuart (2008). Ce modèle explique quels processus sont rattachés à la lecture. Ces derniers nous expliquent que pour être en mesure de bien comprendre un texte, il est primordial d'avoir certaines connaissances de base. Tout d'abord, le lecteur doit avoir une certaine connaissance du contexte auquel se réfère le texte afin de comprendre ce dont il est question. À l'aide de ses connaissances syntaxiques, il doit aussi comprendre la structure des phrases et leurs significations. Par la suite, il doit posséder certaines connaissances visuelles à propos de la forme des mots, de l'orthographe, ainsi que de leur position relative dans le texte, ce qui lui permettra de les reconnaître. Finalement, il doit maîtriser des connaissances phonologiques des mots, qui vont lui permettre de les reconnaître oralement. Il est important de souligner qu'avant qu'un enfant ne soit capable de comprendre un texte, il doit tout d'abord être capable de lire et de reconnaître les mots en utilisant une stratégie de décodage. Cette dernière se développe en quatre stades.

Le premier stade, pré-alphabétique, signifie que les enfants sont capables de reconnaître certains mots écrits, mais sans savoir qu'il existe une relation entre leur son et leur écriture.

Ensuite, lorsqu'un enfant est capable d'identifier les premiers et les derniers phonèmes d'un mot et qu'il est capable d'utiliser cette connaissance pour différencier certains mots écrits, on dit qu'il est dans le stade alphabétique partiel. Troisièmement, l'enfant entre dans le stade alphabétique complet où il est capable d'associer les lettres aux sons qui seront utilisés pour reconnaître des mots. Finalement, au dernier stade, l'enfant est capable de décoder des mots inconnus grâce aux acquisitions qu'il a faites jusqu'à présent. Ainsi, une fois que les mots sont reconnus de façon individuelle, il sera capable de faire le lien entre eux et de comprendre le message véhiculé par la phrase ou le passage. La lecture peut servir à plusieurs fonctions et il faut se rappeler qu'il n'existe pas seulement une seule forme de texte. En plus de la forme que peut prendre un texte, on ne lit pas toujours pour les mêmes raisons. Parfois, on lit pour notre propre plaisir et parfois, on lit pour le travail.

### 1.3 La compréhension de textes et l'anglais langue seconde

Comme on peut le voir plus haut, la lecture est un processus de base qui est formé de plusieurs composantes. Le simple fait de décoder un texte n'est pas une tâche complexe, mais c'est interpréter ce qui a été lu qui peut être problématique. De plus, il existe plusieurs formes de textes et chaque forme peut correspondre à un moyen de communication particulier. Par ailleurs, la deuxième compétence du Programme de formation de l'école québécoise (ministère de l'Éducation, des Loisirs et du Sport, 2001), soit réinvestir sa compréhension des textes écrits et entendus, est associée à plusieurs critères d'évaluation.

L'élève doit démontrer des signes de sa compréhension, utiliser les connaissances tirées des textes dans une tâche de réinvestissement et utiliser des stratégies d'apprentissage. Ce sont tous des critères qui font partie de la progression des apprentissages et qui doivent être évalués tout au long du cycle.

De plus, les enseignants manquent parfois de ressources ou d'idées concernant l'évaluation de la compréhension de textes. Certains d'entre eux vont présenter des textes et demander aux élèves de répondre à des questions afin de voir s'ils ont bien compris. Pourtant, ce qui est prescrit par le programme c'est de voir si les élèves ont les moyens de réinvestir ce qu'ils ont appris dans une autre tâche connexe avec l'information qui est tirée d'un texte. Avec le nouveau programme, les enseignants utilisent souvent des situations d'apprentissage et d'évaluation auprès de leurs groupes afin de faciliter la compréhension des connaissances de leurs élèves. Lors de ces situations, les élèves ont la chance d'en apprendre davantage sur un thème à l'aide de textes, d'exercices d'écriture et d'activités de communication. À la fin des diverses tâches, les élèves doivent, par exemple, faire une rédaction à l'aide de l'information qui a été tirée lors de la situation d'apprentissage et d'évaluation. De cette façon, les enseignants pourront voir si les élèves ont bien compris ce qu'ils ont appris et s'ils réinvestissent leurs apprentissages de façon convenable. Pour ce faire, les élèves doivent d'abord être confiants concernant les différentes tâches de lecture et les enseignants doivent concevoir des instruments d'évaluation des apprentissages adaptés à leurs élèves.

#### 1.4 But et pertinence du projet d'intervention

Le but de ce projet d'intervention est de proposer un modèle et un instrument de mesure qui permettent de prédire le niveau de difficulté des tâches en compréhension de lecture. Avec l'élaboration de ce modèle, les enseignants seront en mesure de décider quelles tâches seraient appropriées pour leurs élèves et le niveau de difficulté de celles-ci. La détermination de ce niveau de difficulté permettra d'assurer une meilleure équivalence quant au niveau de difficulté des examens administrés aux étudiants. Le modèle qui sera proposé pourra être utilisé dans le contexte d'une évaluation à fonction formative en début d'année scolaire afin de bien saisir les besoins des élèves et de les aider à progresser. Puisque les enseignants auront un portrait plus adéquat des capacités de leurs élèves, ils pourront les préparer de façon adéquate aux différentes tâches qui leur seront présentées. Le modèle pourra aussi être utilisé dans le contexte d'une évaluation à fonction sommative, qui permettra alors plus facilement de comparer un groupe classe à un autre.

## CHAPITRE II

### CONTEXTE THÉORIQUE

Le chapitre qui suit présente différents aspects qui sont en lien avec notre projet d'intervention. En premier lieu il sera question du modèle de Bormuth (1970), qui porte sur l'élaboration des items. Deuxièmement, des concepts concernant la production d'items et deux exemples d'application sur la production d'items seront présentés. D'une part, il sera question de l'application *Lextutor* créée par Cobb (2004) et de l'échelle de Lexile de Stenner (1996). Le tout sera suivi d'une synthèse ainsi que des objectifs spécifiques de recherche.

#### 2.1 Le modèle de Bormuth (1970)

Le travail de Bormuth (1970) consiste à comprendre le processus qui existe derrière la création d'items. Selon ses recherches, plusieurs items sont écrits avec subjectivité et manquent de rigueur scientifique : ils ne peuvent donc pas permettre de mesurer avec précision les acquis des élèves. Selon lui, afin qu'un élève puisse bien réussir il est primordial que les items qui lui sont présentés soient en lien avec le contenu du cours.

Selon ses recherches, cet auteur suggère aux enseignants de composer des items qui sont en lien avec ce qui sera appris et de s'assurer que la méthode d'enseignement utilisée facilitera l'apprentissage, donc la réussite aux items qui seront présentés. Lorsque vient le temps de produire les items, l'enseignant peut entreprendre différentes démarches afin de faciliter la tâche pour l'apprenant (Bormuth, 1970).

D'une part, lors de la rédaction des items, il est indispensable que ceux-ci soient créés pour mesurer un processus cognitif spécifique et que les items soient regroupés selon les mêmes processus cognitifs choisis. De cette façon, les liens entre les items se feront plus rapidement et ceci augmentera les chances de réussite. De plus, selon cet auteur, les items sont pertinents si ces derniers peuvent être résolus grâce aux différentes connaissances qui découlent en grande partie des enseignements. Donc, on revient au principe présenté plus tôt, qui veut que les cours soient présentés en fonction des items qui sont dans les examens (Bormuth, 1970).

Puisqu'il suggère que les enseignants utilisent leur enseignement comme le fondement des items qui seront créés, l'enseignant doit s'assurer qu'il y aura une hiérarchie entre les différentes connaissances vues en classe afin de respecter une structure logique auprès des élèves. Ceci est suggéré, puisque certains acquis permettent de mieux résoudre ce qui sera présenté ultérieurement. De plus, les enseignants doivent s'assurer de présenter seulement le contenu qui sera assujéti aux évaluations afin que les élèves se retrouvent facilement lors de leurs études et aucun piège ne doit être inclus dans la tâche d'évaluation. Il est aussi déconseillé d'utiliser des items qui sont redondants. Les items doivent permettre de mesurer différents aspects afin de recueillir le plus d'informations possible durant cette tâche d'évaluation. Finalement, il suggère de bien structurer les items, c'est-à-dire qu'il faut s'assurer que le texte, l'item lui-même, et les choix de réponses soient cohérents. Les items doivent être indépendants les uns des autres. Ainsi, le soutien d'autres items ne devrait pas être requis (Bormuth, 1970).



## 2.2 Concepts sur la production d'items

Selon Irvine et Kyllonen (2002), grâce au développement de la technologie, plusieurs ont aujourd'hui l'opportunité d'avoir recours aux ordinateurs pour produire des items et des examens, ce qui a un impact sur la façon dont un examen est conçu et administré. Les concepts et les résultats d'une étude portant sur le sujet sont présentés ci-dessous.

Afin de produire un item, Newstead, Brandon, Handley, Evans et Dennis, (2002), suggèrent d'identifier un algorithme qui permet la production d'un grand échantillon d'items afin que tous les items qui peuvent être dérivés de ces opérations soient produits. Ensuite, les caractéristiques qui permettent d'identifier le niveau de difficulté d'un item doivent être connues et compréhensibles : il faudra s'assurer de ne pas inclure des caractéristiques qui n'ont aucun impact sur le niveau de difficulté de l'item.

Newstead et collab. (2002), suggèrent deux approches pour produire les items : une première empirique, une seconde théorique. Avec la première approche, on perçoit les items comme un ensemble qui est produit par la combinaison de différents facteurs. Cette approche permet d'utiliser au départ différents items préalablement administrés à des groupes d'élèves : dans cette situation le niveau de difficulté des items peut être obtenu. Puisque le niveau de difficulté des items est alors déjà connu, la tâche consiste à identifier les caractéristiques qui différentient les items et d'élaborer un modèle de prédiction du niveau de difficulté à l'aide des caractéristiques identifiées. Avec la deuxième approche, on identifie au départ les caractéristiques qui devraient permettre de prédire le niveau de difficulté des items et on rédige par la suite des items qui répondent à ces caractéristiques : ce n'est qu'après qu'on vérifie le niveau de difficulté des items et qu'on applique un modèle de prédiction de celui-ci en fonction des caractéristiques de ceux-ci.

La première approche utilise donc des items préalablement élaborés sans avoir utilisé un modèle théorique des caractéristiques ciblées. Dans la seconde approche, on détermine au départ, selon un modèle théorique, les caractéristiques qui devraient permettre de prédire le niveau de difficulté des items et on rédige des items selon ce modèle. Comme pour la première approche, on applique ensuite un modèle de prédiction en fonction des caractéristiques théoriques préalablement identifiées. Avec cette seconde approche, un petit répertoire de caractéristiques théoriques suffit afin de rédiger un grand répertoire d'items (Newstead et collab, 2002).

Selon Bejar (2002), il est important de bien connaître les résultats des recherches précédentes et ce qui a été acquis de ces recherches afin de ne pas reproduire les mêmes erreurs. Chaque item doit se concentrer sur un aspect spécifique du construit afin d'obtenir un résultat plus détaillé au test. Il faut toujours garder en tête que la représentation du construit et la précision de la mesure sont intimement liées. De plus, tout ce qui entoure les items, c'est-à-dire, les consignes d'administration, la notation, et dans quelles conditions l'outil sera utilisé, doit être connu dès le départ. Il suggère, entre autres, d'utiliser un bassin d'items avec des niveaux de difficulté qui diffèrent et de formats différents, ce qui nous permettra de décider si l'on désire conserver des items d'un même niveau de difficulté où si l'on préfère varier leur niveau de difficulté.

Ivrine et Kyllonen (2002), croient que la production d'items permet d'accélérer le processus d'écriture de nouveaux items ce qui aura comme conséquence de prédire le niveau de difficulté des items dans différents contextes. Ceci fera en sorte qu'on pourra estimer des paramètres d'items qui, par conséquent, pourront permettre le développement d'examens.

En ce qui concerne le niveau de difficulté des items, Irvine et Kyllonen (2002), ont cru qu'on doit s'appuyer sur le jugement du créateur des items et choisir des variantes qui ressemblent aux items de départ. Les chercheurs notent qu'il faut se restreindre en ce qui a trait aux règles non pertinentes de création des items, puisqu'elles peuvent parfois contribuer à la difficulté des items, mais qu'elles ne sont pas nécessairement associées à la compétence qu'on désire évaluer. Enfin, afin de bien estimer le niveau de difficulté des items, il est nécessaire, selon eux, de décomposer le domaine dans lequel s'inscrivent les items afin d'identifier des sous-classes d'items qui vont probablement nécessiter différentes connaissances et démarches de solution.

### 2.3 Étude sur la production d'items

L'étude de Enright et Sheehan (2002) a eu comme but de mesurer les processus de raisonnement quantitatifs de différents problèmes mathématiques. Afin d'atteindre leurs objectifs, ils s'intéressent aux études appliquant la théorie de la réponse aux items qui expliquent les mécanismes de prédiction du niveau de difficulté.

Leur étude se compose de deux parties. Pour débiter, ils ont utilisé un échantillon de 20 items de problèmes mathématiques, qui proviennent du *Graduate record examination*, pour analyser et vérifier quel était le niveau de difficulté des items. Cela leur a permis d'identifier les caractéristiques des items sans avoir à se baser sur un modèle théorique formel. Ensuite, ils ont étudié quel était l'effet des caractéristiques préalablement établies sur le niveau de difficulté de nouveaux items produit grâce aux données recueillies.

Pour débiter, deux familles de problèmes ont été sélectionnées pour lesquelles une matrice avec trois caractéristiques d'items a été entrecroisée afin de créer huit variations au sein d'une même famille. Au total, six items ont été rédigés pour chaque classe et ceci a produit 48 variations d'items avec cinq choix de réponses. Ensuite, ils ont jugé nécessaire de séparer les items d'une même famille selon leur difficulté. Chaque variation établie dans la matrice a été assignée à l'une des 24 sections du *Graduate record examination*. Les items ont ensuite été sélectionnés au regard de leur niveau de difficulté et calibrés avec un modèle logistique issu de la théorie de la réponse aux items. Ensuite un modèle de régression a été utilisé pour prédire le niveau de difficulté des items.

Les résultats démontrent que 90 % de la variance concernant le niveau de difficulté des items peut être expliquée grâce à la diversité des problèmes. Par ailleurs, cette recherche a démontré que le lien avec le contexte du problème est statistiquement significatif au seuil de  $p \leq 0,05$ , ce qui signifie que le contexte peut aider ou nuire à résoudre le problème.

Pour la deuxième famille d'items, puisque le niveau de difficulté entre les items était très similaire, le pourcentage de variance expliquée était moins élevé. Au sein de cette famille d'items, les chercheurs ont remarqué que plus la complexité du nombre de tâches augmente, plus l'impact sur le niveau de difficulté se faisait sentir. Par ailleurs, les auteurs ont relevé que les items portant sur un contexte de la vie des étudiants étaient non statistiquement significatifs. Il ne serait donc pas nécessaire d'utiliser des problèmes qui font référence au quotidien des élèves.

La deuxième partie de l'étude voulait vérifier si on pouvait prédire le niveau de difficulté et identifier différentes façons de produire les variantes d'items systématiquement.

Ceci pourrait ultérieurement déterminer jusqu'à quel point les données sur un item nous permettent d'obtenir d'autres informations afin de prédire leur niveau de difficulté. Pour ce faire, les chercheurs ont trouvé que lorsqu'on possède préalablement une classification des items concernant le construit qui est mesuré par le test, on obtient de bonnes balises pour la production de ceux-ci. Les items sont classés selon leur contenu, leur format et les détails concernant les aspects mathématiques afin de créer des catégories et comparer leur forme. Pour ce faire, on a utilisé un plus grand échantillon, soit 368 items.

Afin de vérifier leurs objectifs, les chercheurs ont classé les items selon des compétences cognitives. La première sous-classe d'items fait appel aux connaissances procédurales, soit la mesure des connaissances de base des notions mathématiques. La deuxième sous-classe d'items concerne les items qui nécessitent une compréhension conceptuelle, soit ce qui englobe les items qui permettent de mesurer les idées ou les concepts mathématiques. Cela permet de mesurer si les élèves peuvent expliquer comment résoudre une équation mathématique. Finalement, on retrouve tous les problèmes mathématiques qui nécessitent un haut niveau de raisonnement. Les items dans cette catégorie permettent de mesurer l'habileté à résoudre des problèmes faisant appel à un raisonnement ou une analyse particulière. Les items qui se retrouvent dans cette sous-classe utilisent le principe d'inférence où les élèves doivent utiliser des données préalablement présentées.

La classification à l'aide des caractéristiques cognitives permet d'expliquer 20 % de la variance du niveau de difficulté des items. Avec ces résultats, on peut identifier quels sont les raisonnements et processus cognitifs qui sont nécessaires pour résoudre ces types de problèmes. Les chercheurs suggèrent alors de créer un modèle avec les caractéristiques cognitives qui accompagnent les différentes sous-classes d'items.

Ensuite, 154 items ont été ajoutés selon ces caractéristiques cognitives, ce qui permet d'expliquer 34 % de la variance concernant le niveau de difficulté. Ils ont aussi pu voir que les items qui nécessitent plus d'étapes étaient plus complexes et difficiles à résoudre que les items qui étaient plus simples et comportaient moins d'étapes. Une validation croisée a été effectuée entre les deux familles d'items qui ont inclus tous les formats des items : les caractéristiques cognitives permettent ainsi d'expliquer 30 % de la variance concernant la difficulté.

Même si la variance expliquée n'est pas aussi importante qu'avec la première partie de l'étude, ces résultats permettent d'être généralisés à la rédaction d'autres items dans ce domaine de connaissance. Afin de bien créer des items, les principes qui nous permettent de produire les items et la manipulation de différentes caractéristiques doivent être connus. On a pu comprendre que lorsqu'on utilise des classes d'items et que l'on regroupe les items concernant leurs caractéristiques cognitives, on a une meilleure compréhension de ce qui se produit. De plus, lorsqu'on possède une bonne variabilité entre les items, ceci nous permet de mieux cerner le niveau de difficulté existant entre ces derniers.

#### 2.4 Exemples d'application de production d'items : modèle de Cobb (1997)

Il existe plusieurs exemples d'application avec la production d'items. Une de celles-ci, qui se retrouvent sur l'internet, a été conçue à l'aide des écrits de West (1953) est le *Lextutor* ([www.lexutor.ca](http://www.lexutor.ca)). Cet outil a été conçu afin de maximiser la présence des mots de vocabulaire afin que les élèves puissent les assimiler plus rapidement. Selon Folse (2004), lorsqu'on présente un mot de vocabulaire plusieurs fois, les élèves le retiennent plus facilement, ce qui engendre le déplacement des mots de la mémoire à court terme vers la mémoire à long terme.

*Lextutor*, créé par Cobb (2004), est un outil qui permet aux élèves d'étudier un même mot dans différents contextes pour qu'ils puissent comprendre sa signification afin de se l'approprier et de l'utiliser lorsque nécessaire. L'enseignant peut aussi créer des activités pédagogiques avec des mots de vocabulaire qui seront présentées tout au long de l'année afin d'accroître la compréhension des mots dans différents contextes. C'est avec ce principe que Cobb a modélisé le niveau de difficulté de textes afin que ceux-ci soient adéquats pour les élèves. Le principe de départ derrière la création de cet outil est que les élèves apprennent mieux les mots de vocabulaire si ceux-ci sont présentés sous forme de listes. Ces listes sont construites selon la fréquence d'apparition d'un mot à travers la langue. Avec l'utilisation des listes de vocabulaire, on permet ainsi aux apprenants d'une langue seconde de maximiser la compréhension de certains mots qui ne sont pas toujours utilisés dans des contextes naturels (Cobb, 1997).

De plus, comme dans toutes les langues, l'anglais recycle et réutilise les mêmes mots dans divers contextes. Donc, lorsqu'un élève apprend les mots les plus fréquents, il maximise ses chances de compréhension des textes dans lesquels ceux-ci seront utilisés. À l'aide des listes de vocabulaire qui sont basées sur la fréquence des mots, les apprenants peuvent se concentrer sur les mots qui sont les plus pertinents à leurs apprentissages. Par ailleurs, le nombre de mots maîtrisés augmente selon le niveau scolaire (Cobb, 1997).

Les listes de vocabulaire que l'on retrouve sur *Lextutor* (2004) ont été tirées du corpus de Kucera et Francis (1979), corpus qui est l'aboutissement d'un travail d'analyse de différents textes anglophones ainsi que des corpus de textes audio depuis plus de 50 ans.

On retrouve diverses listes : les 1000 mots les plus fréquents et les 2000 mots les plus fréquents. De plus, les mots qui sont d'un calibre plus élevé se retrouvent sur la liste par niveaux scolaires et, finalement, il y a les mots qui sont hors liste, c'est-à-dire des mots qui sont des acronymes ou proviennent d'autres contextes. Grâce à l'évolution de la technologie et à l'accessibilité des textes en ligne, Cobb (1997), a rendu accessibles ses listes ainsi que les outils qui permettront aux étudiants de mieux comprendre les textes. Les élèves et leurs enseignants peuvent avoir accès à des listes qui sont déjà sur le site internet : il est aussi possible de créer ses propres listes.

Lorsque les enseignants désirent présenter un nouveau texte à leurs élèves, ils peuvent se servir de cet outil puisqu'il possède plusieurs avantages. D'une part, il existe une fonction qui nous permet d'obtenir le pourcentage d'apparition des mots à l'intérieur d'un même texte pour juger si le texte est trop complexe pour nos élèves. Les enseignants peuvent utiliser *Lextutor* (2004) afin de vérifier les textes qu'ils utilisent en classe et peuvent décider de modifier certains mots qui sont trop difficiles pour les élèves avec des mots qui proviennent de la même famille que ceux-ci.

Ils peuvent créer des exercices de compréhension de textes, de vocabulaire et de grammaire qui seront adéquats en fonction du niveau scolaire du groupe. De plus, il y a une section pour les élèves qui comportent des exercices pour apprendre à bien maîtriser la langue et, pour les étudiants plus avancés, à évaluer leurs connaissances en vocabulaire et en grammaire. La théorie qui est stipulée par ce chercheur concernant l'apprentissage de mots de vocabulaire grâce à un outil informatisé n'a pas été vérifiée à l'aide de données empiriques. Il reste un travail à accomplir afin de pouvoir affirmer que ces idées sont réellement utilisables et transférables à différents niveaux de compétences des élèves.



## 2.5 Exemples d'application de production d'items : modèle de Stenner (1996)

Au sein de nos groupes, on retrouve des élèves très hétérogènes, tant sur le plan de leurs connaissances que sur leur habileté à utiliser l'anglais. Grâce à des années de recherches et d'études, Stenner (1996) et son équipe chez Metametrics ont été capables de produire un outil qui vise à sélectionner des textes qui sont appropriés pour les capacités de lecture des élèves. Cet outil se nomme le *Lexile Framework for Reading* (Stenner, 1996).

À la suite de nombreuses recherches, Stenner (1996) s'est aperçu que les enseignants utilisaient des textes trop faciles ou trop complexes pour leurs élèves. De plus, ils choisissaient des thèmes et des sujets de textes qui n'étaient pas adéquats pour l'âge des élèves qu'ils avaient dans leur classe. Ceci a fait en sorte que les élèves se décourageaient rapidement et perdaient la motivation envers le sujet choisi. Il en conclut qu'on ne devrait pas choisir des textes ou des livres en fonction de l'âge ou groupe scolaire d'un enfant, mais plutôt par rapport à ses capacités en lecture. De même qu'avec le thème du livre ou du texte, ce ne sont pas tous les élèves d'un niveau de compétence donné qui seront interpellés par un sujet en particulier (Stenner, 1996).

Le *Lexile Framework for Reading* ([www.lexile.com](http://www.lexile.com)) permet d'évaluer les habiletés en lecture des élèves ainsi que le niveau de difficulté d'un texte afin de pouvoir faire une association qui sera adéquate pour chaque élève. De plus, cet outil peut aussi guider les élèves vers des sujets qu'ils trouvent pertinents afin d'accroître leur motivation à lire. De cette façon, chaque élève aura les bons outils afin de bien réussir et se sentira plus apte à le faire. De plus, le *Lexile Framework for Reading* permet de donner des renseignements sur le type d'enseignement qui devrait être privilégié afin d'améliorer le développement de la compréhension pour chacun différents niveaux de lecture et les sujets.

Par ailleurs, il produit un portrait global des capacités des élèves et permet de renseigner sur leur progression, ce qui permet d'aider l'enseignant à guider adéquatement les élèves vers la réussite. Selon Stenner (1996), se pratiquer à lire reste le meilleur moyen pour permettre à un élève d'accroître ses capacités en lecture et le *Lexile Framework for Reading* permet, de plus, de rendre la pratique significative pour l'élève.

Pour ce faire, ils ont sélectionné des items d'un projet pilote où le niveau de difficulté des textes selon la théorie de la réponse aux items avait déjà été calculé. L'analyse par régression linéaire leur a permis d'obtenir une corrélation qui est égale à 0,97 entre les niveaux de difficulté des textes qui étaient directement observés lors du projet et ceux prédits par la théorie de la réponse aux items. Il s'agit d'une valeur très importante, puisque celle-ci indique qu'avec ce type de texte, il n'est plus nécessaire de calibrer le niveau de difficulté des textes avec des données réelles, mais que le modèle de prédiction est suffisant. Ainsi, le *Lexile Framework for Reading* est un outil intéressant à utiliser.

## 2.6 Limites et forces des articles présentés

Le modèle de Bormuth (1970) qui a été présenté dans ce chapitre apporte plusieurs propos intéressants afin de maximiser la réussite aux examens. D'une part, il suggère aux enseignants de préalablement choisir la matière qui sera présentée à leurs cours afin de pouvoir bien orienter les élèves lors des études. Lorsque les enseignants présentent les concepts à acquérir aux élèves, il est primordial de respecter un ordre hiérarchique pour permettre aux élèves de faire le lien entre ce qui leur est présenté et de mettre en pratique plus aisément les concepts acquis. Par ailleurs, il suggère aux enseignants de regrouper les items utilisant le même processus cognitif afin de faciliter la tâche et augmenter la réussite.

Une limite qui est apparue suite à la lecture du modèle de Bormuth (1970) concerne les méthodes d'enseignement qui sont préconisées. Il suggère d'enseigner en fonction des examens qui seront présentés aux apprenants afin de maximiser la réussite des élèves. Toutefois, ceci peut être une tâche ardue auprès des enseignants, puisqu'ils doivent souvent s'adapter aux types de clientèles qui se retrouvent devant leur classe et donc modifier les examens. Si les enseignants doivent effectuer de nouveaux examens à chaque rentrée scolaire, ceci peut devenir fastidieux et la qualité des examens peut en souffrir. De plus, son modèle suggère que seulement la matière qui sera assujettie aux examens devra être présentée. Encore une fois, ceci est très favorable pour la réussite des tests, mais il ne faut pas oublier que certaines connaissances doivent être adaptées à certains niveaux de compétences des élèves afin que ceux-ci puissent maîtriser ce qui sera présenté durant l'année scolaire suivante.

Ensuite, plusieurs éléments sont soulevés afin de bien cerner ce qui est en jeu lors de la production d'items. Newstead et collab. (2002), ont proposé deux approches qui peuvent être utilisées selon les expertises de la personne. D'une part, nous pouvons utiliser des items déjà existants afin d'en créer d'autres, une approche empirique, ou bien, selon une approche théorique, nous pouvons utiliser une équation mathématique qui permet de produire d'autres items. De plus, ils suggèrent de procéder à une bonne analyse des items et de leurs caractéristiques ce qui permet de s'assurer d'évaluer les bonnes compétences.

Bejar (2002) a suggéré de posséder une bonne connaissance du construit et de tout ce qui englobe le contexte d'administration des items. Ceci aura comme effet de maximiser la réussite des items.

De son côté, Irvine et Kyllonen (2002) stipulent que la production d'items permet d'accélérer le processus d'écriture et permet de définir des caractéristiques de différents items qui pourront être utilisées à travers différents contextes. Il est aussi très important de connaître quelles caractéristiques peuvent être déduites des items afin de créer de nouveaux items variés qui pourront être utilisés pour une longue période de temps.

Dans l'étude qui a été présentée, on a pu comprendre que lorsqu'on possède une bonne connaissance des caractéristiques des items, il est plus facile de produire des items dont le niveau de difficulté varie. Par ailleurs, le contexte d'administration est très important, puisque ce dernier peut avoir des effets positifs ou néfastes sur la réussite des élèves. De plus, on a pu observer que lorsqu'on possède des connaissances sur les caractéristiques cognitives des items, ceci contribue à prédire le niveau de difficulté des items. Finalement, l'étude propose de séparer les items selon leurs formes et selon les procédures requises pour les réussir (Enright et Sheehan, 2002).

Ces concepts sont très intéressants, cependant, certaines limites quant à leur utilisation se présentent. D'une part, par manque de temps et de ressources, ce ne sont pas tous les enseignants qui peuvent effectuer une étude préalable pour permettre une production d'items aussi exhaustive parmi celles suggérées dans la recension de ces articles.

De plus, on a pu découvrir que l'exposition fréquente à différents mots permet aux apprenants d'une langue seconde de mieux les retenir et de les utiliser plus fréquemment. Il est préférable d'utiliser comme moyen d'étude des listes de vocabulaire présentées selon leur fréquence d'utilisation.

Le fait d'utiliser des listes de vocabulaire permet aux apprenants d'une langue seconde d'avoir la chance de rencontrer des mots qui sont moins souvent présents et leur permettra d'élargir leur vocabulaire. Ils auront ainsi plus de chance de comprendre les textes qui seront présentés et d'explorer des textes qui n'auraient pas été abordés si ce ne fut pas pour l'acquisition du vocabulaire.

Concernant les limites de l'utilisation de listes de vocabulaire, des études futures qui comparent la performance des élèves qui utilisent les listes de vocabulaire et ceux qui ne les utilisent pas auprès de clientèles jeunes seraient pertinentes. De plus, *Lextutor* (2004) peut être un outil pertinent pour les enseignants afin de bien diriger leurs élèves et créer du matériel pédagogique qui est en lien avec le contenu abordé dans les textes. Par contre, il est un peu complexe pour les élèves du primaire et ceux-ci risquent de se perdre en l'utilisant. Il est plutôt conçu pour les élèves avancés qui ont de l'expérience avec la matière. Par ailleurs, les bienfaits d'un outil comme celui-ci ont été difficiles à vérifier empiriquement. Les idées de départ qui sont le fondement de ce projet sont très intéressantes, mais des études supplémentaires qui nous donnent des données statistiques seront nécessaires afin que ce genre de matériel soit utilisé auprès des élèves et les écoles.

Ensuite, Stenner (1996) s'est aperçu que les enseignants ne faisaient aucune différenciation entre les élèves de leurs classes et présentaient simplement les textes qui étaient prescrits au curriculum. Il a décidé de créer le *Lexile Framework for Reading*, une échelle unique puisqu'elle permet de mesurer le niveau d'habileté en lecture des individus et le niveau de difficulté des textes. Avec ces informations, l'enseignant pourra utiliser les textes qui seront au niveau approprié de chaque élève afin de maximiser leur réussite qui représente une des forces de cet outil.

Une limite concernant l'utilisation du *Lexile Framework for Reading* serait son accessibilité dans nos classes. Ce n'est peut-être pas un outil qui serait accessible dans toutes les écoles et il demanderait un effort de gestion supplémentaire de la part des enseignants en plus de leur travail habituel.

## 2.7 Objectif spécifique

Lors de la réalisation de ce contexte théorique, nous avons pu constater que plusieurs ressources existent dans le domaine de la compréhension de textes afin d'aider les élèves à utiliser les acquis appris en classe dans divers contextes. De plus, plusieurs suggestions sont faites auprès des enseignants afin de maximiser l'apprentissage, ce qui aura comme conséquence une meilleure réussite auprès des élèves. Avec ce qui a été présenté plus haut, nous avons comme objectif spécifique de proposer un modèle adapté aux épreuves administrées aux élèves du primaire au Québec qui permet de prédire le niveau de difficulté des tâches en compréhension de lecture en anglais, langue seconde. Pouvant prédire le niveau de difficulté des tâches, ce modèle devrait permettre aux enseignants de construire des examens dont le niveau de difficulté sera formellement connu et ainsi mieux assurer l'équivalence des résultats d'évaluation d'un groupe classe à un autre ou d'un enseignant à un autre.

## CHAPITRE III

### MÉTHODOLOGIE

Dans cette section, il sera question de la méthodologie utilisée afin de répondre à l'objectif de recherche. Les sujets choisis, l'instrument de mesure, le déroulement et les méthodes d'analyses choisies seront présentés. Finalement, les considérations éthiques qui ont dû être respectées seront expliquées.

#### 3.1 Sujets

Les données qui sont utilisées pour ce projet d'intervention proviennent des élèves du primaire de l'île de Montréal. Si on regarde l'âge de ces étudiants, celle-ci varie de 10 à 12 ans. Les participants proviennent de deux écoles situées dans des quartiers résidentiels différents. La première école se situe dans le nord de la ville et la seconde dans l'est. Les élèves de la première école sont tous des élèves de sixième année qui suivent le programme d'anglais intensif ( $n = 33$ , soit 29,2 % de tous les élèves ayant participé au projet). Pour les élèves de la deuxième école, on retrouve des élèves de sixième année qui suivent le programme régulier et intensif d'anglais langue seconde, ainsi que des élèves de la cinquième année régulière ( $n = 80$ , soit 70,8 % de tous les élèves ayant participé au projet). Ces données sont présentées au tableau 1.

Tableau 1. Distribution des sujets selon l'école

ÉCOLE FRÉQUENTÉE	FRÉQUENCE	POURCENTAGE
Nord de la ville	33	29,2
Est de la ville	80	70,8
Total	113	100,00

La répartition selon le sexe et le niveau d'enseignement de ces élèves est présentée au tableau 2. On peut y remarquer que le nombre de garçons ( $n = 61$ , soit 54 %) est légèrement supérieur à ceux des filles ( $n = 52$ , soit 46 %).

Tableau 2. Distribution des sujets selon le genre

GENRE	FRÉQUENCE	POURCENTAGE
Féminin	52	46,0
Masculin	61	54,0
Total	113	100,0

De ces élèves, 36 se retrouvent en cinquième année, 9 en sixième année régulière et 68 dans le programme de sixième année intensif. En ce qui concerne les élèves qui font partie du programme intensif, deux options sont possibles : 1) les élèves commencent l'année scolaire avec le programme d'anglais et terminent avec le programme scolaire ou 2) les élèves débutent l'année avec le programme scolaire régulier et ils terminent avec le programme d'anglais intensif. La répartition selon le moment où les élèves ont suivi le programme d'anglais est présentée au tableau 3. On peut y remarquer que le nombre des élèves qui commencent en anglais ( $n = 33$  soit, 22,8 %) est semblable à celui de ceux qui terminent en anglais ( $n = 35$  soit, 31 %).

Tableau 3. Distribution des sujets selon le moment de suivre le programme d'anglais intensif

MOMENT OÙ LE PROGRAMME EST SUIVI	FRÉQUENCE	POURCENTAGE
Début	33	29,2
Termine	35	31,0
Élèves au régulier	45	39,8
Total	113	100,00



### 3.2 Dérroulement

Afin de réaliser ce projet d'intervention, les données sont recueillies à deux moments distincts, soit à l'hiver 2010 et au printemps 2011. Tout d'abord, les élèves qui ont commencé avec le programme d'anglais langue seconde dans les deux écoles ont répondu aux questions un peu avant la fin du programme, c'est-à-dire au mois de décembre 2010. Ensuite, les élèves qui ont entamé le bain linguistique, programme d'anglais intensif, au mois de janvier 2011 ont répondu à l'examen au printemps, soit au mois de mars 2011, peu après le début du bain linguistique. En ce qui concerne les élèves qui suivent le programme d'anglais régulier de la cinquième et de la sixième année, les deux groupes ont répondu à l'examen au mois de mars 2011. Lors de la passation de l'examen, les élèves n'ont eu droit à aucun matériel pédagogique et n'ont pu poser aucune question à l'enseignant présent.

### 3.3 Description de l'instrument de mesure

L'examen qui a été utilisé pour ce projet d'intervention identifie les habiletés en lecture des élèves du troisième cycle du primaire en anglais, langue seconde. On y retrouve des textes et des questions variés afin d'avoir un meilleur aperçu du niveau de compétence des étudiants.

#### 3.3.1 Élaboration de l'instrument de mesure

Une première version de l'examen a été mise sur pied en septembre 2010. Certains textes ont été choisis à l'intérieur du manuel scolaire des élèves de la sixième année, *A New TWIST to English : a competency based approach*, (Gunning, Lalonde, Schinck et Watts, 2001).

Par la suite, des questions ont été composées pour accompagner ces textes. Cette première version a été présentée à deux enseignantes d'anglais langue seconde. Suite à leur recommandation, une deuxième version a été composée en novembre 2010 en ajoutant des textes de deux manuels distincts, soit celui de Flores, Moran, et Orzo (2004) et celui de Black, Marsha et McAuliffe (2001). Cette deuxième version a été soumise à la conseillère pédagogique de l'école ainsi qu'aux enseignantes d'anglais. Par la suite, cette version a été présentée à la conseillère pédagogique du primaire en anglais, langue seconde de la Commission Scolaire de Montréal. Les dernières modifications ont été apportées à cette deuxième version et le produit final était prêt à être utilisé en décembre 2010. Au total, on retrouve 113 élèves qui ont répondu à l'examen avec 67 données manquantes, c'est-à-dire des questions non répondues.

### 3.3.2 Choix des textes de compréhension en lecture

En ce qui concerne les textes de lecture sélectionnés pour l'examen, certains possèdent une structure que les élèves connaissent déjà. Au total, on retrouve 10 textes. Certains de ces textes sont très simples et comportent peu de phrases et un vocabulaire simple pour les élèves. Les autres sont composés de phrases plus longues et d'un vocabulaire plus complexe. De ces textes, 4 proviennent du manuel scolaire utilisé par les élèves. À l'opposé, 5 possèdent une structure un peu différente et pour laquelle les élèves ne sont pas familiers. Tous ces textes ont été analysés à l'aide du logiciel *Lextutor* (Cobb, 2004). Ce logiciel nous permet de voir combien de mots se retrouvent dans le texte, la fréquence, ainsi que la difficulté de ces mots.

### 3.3.3 Choix des items

Chaque texte est associé à 4 questions à choix de réponses. Certains de ces choix de réponses sont des phrases à compléter, des vrais ou faux ou de simples énoncés. Les items ont été composés à la suite des recommandations de la conseillère pédagogique en anglais, langue seconde de la CSDM. Puisque cette dernière produit les examens de fin de cycle pour les élèves et qu'elle est une enseignante avec plusieurs années d'expérience dans le milieu, ses conseils se sont avérés pertinents et très utiles.

### 3.4 Méthode d'analyse choisie

Dans notre cas, pour avoir une idée de ce qui se produit à l'aide de cet instrument de mesure, chaque item sera analysé individuellement. Afin de réaliser nos analyses concernant l'instrument de mesure présenté plus haut, il faut tout d'abord procéder à une analyse descriptive des données afin d'avoir une idée des données qui seront explorées ultérieurement. Dans le cadre de l'analyse descriptive, nous observerons dans un premier temps la distribution des données qui a été faite pour chaque item afin de voir combien d'étudiants ont réussi chacun d'eux, ce qui correspond à notre indice de difficulté des items mis à l'épreuve. Il est important d'indiquer que plus l'indice a une petite valeur, plus il sera jugé difficile et plus l'indice a une grande valeur, plus il sera jugé facile. L'indice de difficulté sera jugé très petit s'il est égal ou inférieur à 0,10. Il sera jugé élevé s'il est égal ou supérieur à 0,90. Ensuite, nous regarderons l'indice de discrimination afin de voir si les items en lecture que nous avons élaborés peuvent faire la différence entre les élèves forts et faibles : une valeur minimale de 0,30 est espérée. Finalement, la consistance interne du test sera vérifiée. Pour ce faire, le calcul de la fidélité de l'instrument de mesure sera effectué.

Pour qu'un examen présente une bonne fidélité, un minimum de 0,80 doit être atteint, seuil que nous avons fixé. Ceci sera vérifié à l'aide du coefficient alpha de Cronbach (Bertrand et Blais, 2004).

Nous procéderons ensuite à trois analyses de régressions multiples pour prédire la difficulté selon la théorie classique des tests : pour les textes, les questions et les choix de réponses. Pour ces analyses, plusieurs coefficients seront utilisés. Premièrement, nous analyserons la valeur du coefficient  $R^2$ , coefficient de détermination, qui nous donnera de l'information concernant la variance expliquée de la variable dépendante par la combinaison de toutes les valeurs introduites dans notre modèle. Ici, une valeur près de 1 est espérée. La prochaine étape est d'étudier chacun des coefficients de régression des trois régressions multiples. Ces coefficients seront ceux qui pourront être utilisés pour prédire les valeurs des variables dépendantes. Pour ce faire, nous analyserons aussi les coefficients bêta ( $\beta$ ), qui sont standardisés, et qui indiquent le poids relatif de chacune des variables dans le modèle. Aussi, le sens de la relation se voit à l'aide du signe positif ou négatif. Le seuil de signification statistique retenu sera de 0,05.

Afin d'appliquer nos régressions multiples, nous avons choisi comme variable dépendante le niveau de difficulté qui correspond à la proportion de réussite selon la théorie classique des tests. Pour chaque critère, soit les textes, les questions et les choix de réponses, plusieurs variables indépendantes sont retrouvées. En premier lieu, pour les textes, nous avons choisi comme variable: le nombre de caractères, ce qui inclue les mots, espace et signes de ponctuation dans le texte (`textecaractere`), le type du texte (`textetype`), le nombre de phrases dans le texte (`textephrase`), le nombre de mots qui font partie du premier corpus des élèves, (`motk1`) et du deuxième corpus (`motk2`), le nombre de mots qui font partie de la liste scolaire de vocabulaire (`motSAWL`) et les mots qui ne figurent sur aucune liste de vocabulaire des élèves (`motofflist`).

Ensuite, pour les questions, nous avons retenu comme variables : le nombre de mots dans la question (`nbmotquestion`), la difficulté des mots dans la question incluant les mots qui font partie du premier corpus des élèves (`diffmotK1`), les mots qui font partie du deuxième corpus des élèves (`diffmotK2`), les mots qui figurent sur la liste académique des élèves (`diffmotAWL`) et les mots qui ne figurent pas sur la liste scolaire des élèves (`diffmotofflist`), le nombre de mots (`qapmot`) et caractères (`qapcarac`) dans le texte qui précède le moment où la question est posée.

La dernière variable vise à vérifier le nombre de mots qui se retrouvent dans la question et à l'intérieur du texte (`qmottexte`). En dernier lieu, pour les choix de réponses, nous avons retenu comme variables le nombre de choix de réponses (`nbchoixrep`), les mots qui se retrouvent dans le texte et les choix de réponses (`motrépontexte`), le nombre de mots dans le texte et la question (`nbmotQ`) et la difficulté des choix de réponses (`diffchoixrep`). Toutes les analyses ont été effectuées à l'aide du logiciel SPSS 18 (Kinnear et Grey, 2010).

### 3.5 Considérations éthiques

Puisque notre projet d'intervention demande l'implication de sujets mineurs, certaines considérations éthiques ont dû être respectées. D'une part, une rencontre avec les directions des deux établissements scolaires a été effectuée afin de leur présenter l'intervention et l'implication des jeunes. Ensuite, chaque direction a rempli un formulaire de consentement pour que les élèves et les enseignantes participent au projet d'intervention.

Les élèves ont été rencontrés avant l'intervention afin de leur expliquer le projet et le formulaire de consentement. Il n'y avait aucune obligation de leur part et la participation était volontaire. Les formulaires pour les directions d'école et les élèves se retrouvent aux annexes A et B. Afin de garder l'anonymat des sujets, aucune donnée personnelle concernant les élèves n'a été recueillie.

Afin que les participants soient informés sur l'intervention, une présentation concernant celle-ci et les résultats obtenus sera offerte aux enseignants et directions d'écoles. De plus, une copie du rapport leur sera remise pour qu'ils puissent le consulter et pour que d'anciens élèves ou parents d'élèves puissent obtenir de l'information sur le projet.

## CHAPITRE IV

### RÉSULTATS

#### 4.1 Analyses descriptives

Pour débiter, il est primordial d'avoir une idée de la distribution de l'échantillon utilisé pour l'évaluation en lecture. Pour ce faire, on présentera au tableau 4 le nombre d'étudiants qui ont réussi chacun des 40 items ainsi que l'indice de difficulté. On se souvient que 113 élèves ont répondu au questionnaire en lecture.

Tableau 4. Fréquence des élèves qui ont réussi chacun des items et proportion des 40 items, indice de difficulté, pour l'évaluation en lecture

Item	Nombre d'élèves et (indice de difficulté)	Item	Nombre d'élèves et (indice de difficulté)	Item	Nombre d'élèves et (indice de difficulté)	Item	Nombre d'élèves et (indice de difficulté)
1	<b>107 (0,95)</b>	11	73 (0,65)	21	89 (0,79)	31	36 (0,32)
2	51 (0,45)	12	64 (0,57)	22	69 (0,61)	32	76 (0,67)
3	75 (0,66)	13	95 (0,84)	23	33 (0,29)	33	48 (0,42)
4	43 (0,38)	14	48 (0,42)	24	63 (0,56)	34	54 (0,48)
5	69 (0,61)	15	82 (0,73)	25	95 (0,84)	35	93 (0,82)
6	93 (0,82)	16	67 (0,59)	26	97 (0,86)	36	81 (0,72)
7	77 (0,68)	17	75 (0,66)	27	23 (0,20)	37	83 (0,73)
8	95 (0,84)	18	92 (0,81)	28	90 (0,80)	38	89 (0,79)
9	96 (0,85)	19	84 (0,74)	29	63 (0,56)	39	76 (0,67)
10	93 (0,82)	20	77 (0,68)	30	68 (0,60)	40	85 (0,75)

Note: Les caractères en gras indiquent les items avec un indice de difficulté faible (plus bas que 0,10) ou élevé (plus élevé que 0,90)

Si on regarde, dans un premier temps, le nombre d'élèves qui ont réussi l'évaluation de lecture, on peut déjà constater qu'un seul item pourrait poser problème puisqu'il est réussi par un très grand nombre d'élèves (95 %). Il s'agit de l'item 1. Cet item devrait être modifié ou retiré dans une version ultérieure de l'instrument. Toutefois, aucun item ne présente un indice de difficulté inférieur à 0,10 et peut être considéré comme très difficile.

Selon la théorie classique des tests, on obtient ainsi un indice de difficulté; moyen de 0,65 avec un minimum de 0,20 et un maximum de 0,95, ce qui signifie que notre test est considéré en général comme étant facile. Il aurait été possible de créer un examen un peu plus difficile, mais puisque c'est un examen qui a été conçu pour des débutants ainsi que pour une intervention pédagogique, ceci pourrait être acceptable.

#### 4.2 Analyse de la précision de mesure et des coefficients de détermination

Il est nécessaire de vérifier la fidélité de cet outil, soit la consistance interne de notre examen de lecture. Lorsqu'on vérifie la fidélité d'un test, on désire vérifier si les items qui sont à l'intérieur de cet outil sont bien reliés entre eux et nous permettent de mesurer ce qui est voulu. Ceci se fera à l'aide du coefficient alpha de Cronbach. On se rappelle que le seuil espéré pour la fidélité qu'on s'est fixé est de 0,80. Dans un premier temps, nous avons obtenu un alpha de Cronbach de 0,89 pour les 40 items inclus. Le résultat obtenu ici nous permet de dire que notre instrument de mesure est fidèle à lui-même puisque les items sont homogènes et permettent de mesurer ce que l'on veut. Dans un deuxième temps, il faut vérifier la corrélation entre chacun des items qui se retrouvent à l'intérieur de cet examen de lecture et le score total au test, aussi connu sous le nom d'indice de discrimination. Ce coefficient vérifie si chacun des items permet de faire la différence entre les élèves faibles et forts.



D'une part, aucun de ces coefficients n'est négatif. Donc, on ne possède aucune indication au sein de ces questions qui voudraient dire que plus un étudiant est fort, moins il réussit l'item : c'est ce qu'on recherchait. Seuls les chiffres qui étaient relativement en dessous du seuil espéré de 0,30 ont été analysés. Les résultats sont présentés au tableau 5. Les items 1, 4, 8, 14, 23, 27, 31, et 39 affichent un indice de discrimination en dessous du seuil fixé de 0,30, et devront donc être retirés ou modifiés dans le futur. Au total, il s'agit de 8 items, donc 20 % de l'examen, qui poseraient éventuellement problème. Deux de ces items présentent un indice de discrimination inférieur ou égal à 0,05 : ce sont les items 27 et 39.

Tableau 5. Valeur de la corrélation item-total, soit l'indice de discrimination pour chacun des items utilisé lors de l'évaluation en lecture

Item	r	Item	r	Item	r	Item	r
1	<b>0,20</b>	11	0,59	21	0,44	31	<b>0,20</b>
2	0,26	12	0,55	22	0,48	32	0,41
3	0,31	13	0,38	23	<b>0,21</b>	33	0,53
4	<b>0,28</b>	14	<b>0,23</b>	24	0,42	34	0,58
5	0,40	15	0,57	25	0,43	35	0,41
6	0,615	16	0,37	26	0,34	36	0,41
7	0,322	17	0,41	27	<b>0,04</b>	37	0,44
8	<b>0,29</b>	18	0,34	28	0,46	38	0,47
9	0,42	19	0,39	29	0,51	39	<b>0,05</b>
10	0,50	20	0,47	30	0,45	40	0,46

Maintenant que nous avons en main toutes les analyses concernant nos items, en regardant les indices de difficulté et de discrimination, un seul item pourrait poser problème au regard de ces deux indices simultanément : il s'agit de l'item 1.

#### 4.3 Analyses de régression linéaires

Dans un deuxième temps, nous avons aussi effectué trois analyses de régression linéaire avec les variables indépendantes des textes, des questions et des choix de réponses. Les trois sous sections suivantes en présentent les résultats.

#### 4.3.1 Analyse de régression linéaire multiple pour les textes

Notre première série d'analyses de régression multiple se fera en considérant les textes qui ont été utilisés pour rédiger les items, afin de savoir si le choix des textes peut influencer la réussite aux items. Par exemple, on aurait tendance à croire que si un texte est court et que le vocabulaire associé est simple, celui-ci serait plus facile à interpréter et donc que l'élève aura plus de facilité à répondre aux questions qui l'accompagnent. Le résultat de ces analyses se retrouve aux tableaux 6 et 7. Tout d'abord, nous regarderons le coefficient de détermination qui se retrouve au tableau 6. Ce dernier affiche une valeur de 0,14, ce qui signifie que 14 % de la variance en compréhension de lecture peut être expliquée par la combinaison de toutes les variables indépendantes concernant les textes. Si on utilise l'échelle de Cohen pour interpréter l'importance de ce coefficient, cette valeur se situerait dans la moyenne (Kinnear et Gray, 2010). On pourrait supposer que cette valeur n'est pas assez représentative pour dire que la combinaison de ces variables peut avoir une importance sur la prédiction de la proportion de réussite à chacun des items étudiés. Ceci voudrait dire que les variables indépendantes retenues pour les textes ne sont pas suffisantes pour prédire adéquatement l'indice de difficulté des items ou encore que l'échantillon d'items retenus n'est pas adéquat, car il est trop faible ou encore qu'il n'y a pas assez de variabilité dans le niveau de difficulté des textes retenus.

Tableau 6. Coefficients de corrélation<sup>a</sup> et de détermination pour la régression multiple des textes

Modèle	R	R <sup>2</sup>	Erreur-type	Sig.
1	0,37 <sup>a</sup>	0,14	0,18	0,64

a. Variables indépendantes: (constantes), *textecaractere*, *typedetexte*, *textephrase*, *motk1*, *motk2*, *motawl*, *motofflist*

b. Variable dépendante: *propreussite*

Maintenant que nous connaissons la qualité de l'ajustement du modèle de régression aux données, il est nécessaire d'analyser les coefficients de régression non standardisés  $b$  du modèle. Ces coefficients sont représentés par l'équation 1 suivante. Les symboles qui se retrouvent dans cette équation et les suivantes sont ceux qui ont été présentés lors de la méthodologie lors de l'explication des méthodes d'analyses choisies, section 3.3.3. Ce sont les symboles qui représentent nos variables indépendantes.

$$\begin{aligned} \text{Proportion de réussite} = & b_0 + b_1 \text{textecaractere} + b_2 \text{typedetext} + \\ & b_3 \text{textephrase} + b_4 \text{motk1} + b_5 \text{motk2} + b_6 \text{motAWL} + b_7 \text{motofflist} + e \end{aligned} \quad (1)$$

Dans un premier temps, nous allons étudier les coefficients standardisés  $\beta$  afin d'interpréter le poids relatif de nos variables et quelles variables peuvent être conservées pour prédire l'indice de difficulté. Nous les présentons au tableau 7. La première variable qui ressort est celle qui porte sur le nombre de phrases à l'intérieur du texte ( $\beta = 2,88$ ) : plus un texte contient des phrases, plus la proportion de réussite augmente. La deuxième variable concerne le nombre de mots qui font partie du premier corpus des élèves (de 0 à 1000). Il s'agit des mots que les élèves apprennent en premier et qui sont généralement les plus simples à apprendre ( $\beta = -3,92$ ). Ce coefficient de régression standardisé est négatif, ce qui signifie que moins j'ai de mots dans cette catégorie, plus ma proportion de réussite augmente. Ces deux variables pourraient donc avoir un impact pour prédire l'indice de difficulté : cependant ils ne sont pas statistiquement significatifs au seuil retenu ( $\text{textephrase} = 2,885$ ,  $p > 0,05$ ) et ( $\text{motk1} = -3,95$ ,  $p > 0,05$ ). Bref, même si elles ressortent, statistiquement parlant, on ne peut les considérer. Ceci peut être dû par la petite taille de notre échantillon de textes ou bien tout simplement par le choix même des textes. D'autres analyses seraient nécessaires afin de vérifier le potentiel de ces deux variables.

Tableau 7. Coefficients<sup>a</sup> de régression multiple pour prédire l'indice de difficulté des textes

Modèle	Coefficients non standardisés		Coefficients standardisés		
	<i>b</i>	Erreur type	$\beta$	<i>t</i>	<i>p</i>
1 (Constante)	0,668	0,153		4,373	0,00
Textecaractere	0,00	0,002	0,786	0,125	0,901
Texttype	-0,35	0,106	-0,098	-0,330	0,744
Textephrase	0,38	0,050	<b><u>2,885</u></b>	0,744	<b><u>0,462</u></b>
motK1	-0,006	0,12	<b><u>-3,926</u></b>	-0,489	<b><u>0,628</u></b>
motK2	-,005	0,005	0,308	0,441	0,662
motAWL	0,25	0,025	0,682	0,875	0,388
MotOffList	-0,002	-0,002	-0,233	-0,150	0,881

#### 4.3.2 Analyse de régression linéaire multiple pour les questions

Un autre facteur qui pourrait nous aider à prédire la réussite à une tâche de compréhension de textes est la formulation des questions qui s'y retrouvent. Une fois de plus, plusieurs variables indépendantes ont été déterminées afin de prédire l'indice de difficulté. L'équation qui a été retenue est la suivante :

$$\begin{aligned}
 \text{Proportion de réussite} = & b_0 + b_1 \text{nbmotQ} + b_2 \text{diffmotQ} + b_3 \text{diffmotK2} + \\
 & b_4 \text{diffmotAWL} + b_5 \text{diffmotOffList} + b_6 \text{Qapmot} + b_7 \text{Qapcarac} + \\
 & b_8 \text{Qmottexte} + e
 \end{aligned}$$

(2)

De même que pour l'analyse de régression concernant les textes, nous allons observer l'évaluation de la performance du modèle avec le coefficient de détermination. On retrouve les résultats au tableau 8.

Tableau 8. Coefficients de corrélation<sup>a</sup> et de détermination pour la régression multiple des questions

Modèle	R	R <sup>2</sup>	Erreur-type	p
Questions	0,57	0,32	0,16	0,11

a. Valeurs indépendantes : (constantes), nbmotq, diffmotq, diffmotk2, diffmotawl, diffmotofflist, qapmot, qapcarac, qmotttexte

b. Variable dépendante : propreussite

Le coefficient de détermination de cette analyse nous indique que 32 % de la variance du modèle peut être expliquée par la combinaison de nos variables indépendantes. Dans ce cas, ce coefficient présenterait une valeur élevée selon l'échelle de Cohen, ce qui serait acceptable (Kinnear et Gray, 2010). Ceci voudrait signifier que les variables indépendantes retenues pour les questions seraient en mesure de prédire l'indice de difficulté de nos items.

Afin de vérifier quelles variables ont le plus de poids, une analyse des coefficients de régression standardisés est nécessaire. Les résultats se retrouvent au tableau 9. Lorsqu'on regarde de plus près les coefficients que l'on obtient à cette analyse, deux variables indépendantes ressortent. Il s'agit des variables qui indiquent après combien de mots et après combien de caractères la question est posée. Ces deux variables ont été retenues dans le but de voir si l'ordre dans lequel les questions sont posées importe. D'une part, il y a la variable concernant le nombre de mots se retrouvant après que la question soit posée : le coefficient de régression standardisé associé est statistiquement significatif ( $Q_{apmot} = -2,682$ ,  $p \leq 0,05$ ). De plus, puisqu'on obtient un coefficient négatif, ceci indique que plus le nombre de mots qu'on retrouve suite à notre question augmente, plus la proportion de réussite diminue. Ensuite, il y a la variable concernant le nombre de caractères qu'on retrouve après que la question soit posée. Le coefficient est positif, ce qui indique que plus il y a de caractères à la suite des questions, plus la proportion de réussite augmente.

Cette variable est aussi statistiquement significative ( $Q_{apcarac} = 2,515$ ,  $p \leq 0,05$ ) et pourrait nous aider lorsque viendra le temps de proposer un modèle. On peut voir que ces deux variables contribuent au modèle et on a alors peu de chance de se tromper en rejetant l'hypothèse nulle qui dit que ce coefficient serait égal à 0.

Tableau 9. Coefficients<sup>a</sup> de corrélation multiple pour prédire l'indice de difficulté des questions

Modèle	Coefficients non standardisés		Coefficients standardisés		
	<i>b</i>	Erreur type	$\beta$	<i>t</i>	<i>p</i>
1 (Constante)	0,613	0,091		6,172	0,00
nbmotQ	0,005	0,017	0,1091	0,307	0,761
DiffmotQk1	-0,009	0,014	-0,184	-0,603	0,551
DiffmotQk2	0,059	0,043	0,239	1,355	0,185
DiffmotAWL	0,047	0,060	0,140	0,779	0,442
DiffmotOffList	0,019	0,014	0,229	1,332	0,193
Qapmot	-0,004	0,001	<b><u>-2,682</u></b>	-2,935	<b><u>0,006</u></b>
Qapcarac	0,001	0,000	<b><u>2,515</u></b>	2,715	<b><u>0,011</u></b>
Qmottexte	0,004	0,003	0,230	1,475	0,150

#### 4.3.3 Analyse de régression linéaire multiple pour les choix de réponses

La dernière analyse de régression multiple qui a été effectuée concerne les caractéristiques des choix de réponses. Pour débiter, les coefficients de détermination sont présentés au tableau 11. Avec les résultats obtenus, nous sommes capables d'expliquer 38 % de la variance de l'indice de difficulté, soit la variable dépendante. Ici, le coefficient de détermination est très élevé et permet de bien expliquer notre modèle (Kinnear et Gray, 2010).

Tableau 10. Coefficients de corrélation et de détermination pour la régression multiple des choix de réponses

Modèle	R	R <sup>2</sup>	Erreur-type	Sig.
Choix de réponses	0,67 <sup>a</sup>	0,38	0,15	0,02

a. Valeurs indépendantes : (constantes), nbchoixderéponse, mottxrep, nbmot, diffchoixrep  
 b. Variable dépendante : propreussite

Maintenant que nous possédons l'information concernant notre modèle vis-à-vis les caractéristiques des choix de réponses, le temps est venu d'analyser les différentes variables indépendantes et voir lesquelles sont plus déterminantes. L'équation qui a été choisie pour cette régression est la suivante :

$$\text{Proportion de réussite} = b_0 + b_1 \text{nbchoixderéponse} + b_2 \text{mottxrep} + b_3 \text{nbmotQ} + b_4 \text{diffchoixrep} + e \quad (3)$$

Tous nos coefficients de régression standardisés sont présentés au tableau 11. La variable qui ressort le plus afin de prédire la valeur de l'indice de difficulté selon les caractéristiques des choix de réponses est celle qui représente la complexité des choix de réponses (*diffchoixrep*) : si ces derniers se retrouvent dans le texte ou si les élèves doivent faire appel à leurs propres connaissances pour y répondre.

Pour cette variable indépendante, on obtient un coefficient de régression standardisé  $\beta$  égal à -0,631 qui est statistiquement significatif au seuil prédéterminé. Ceci voudrait indiquer que plus le choix de réponses est complexe et que les élèves doivent effectuer certaines inférences, plus la valeur de l'indice de difficulté diminue et, ainsi, la proportion de réussite à l'item.

Tableau 11. Coefficients<sup>a</sup> de régression multiple pour prédire l'indice de difficulté des choix de réponses

Modèle	Coefficients non standardisés		Coefficients standardisés		
	<i>b</i>	Erreur type	$\beta$	<i>t</i>	<i>p</i>
1 (Constante)	1,340	0,243		5,524	0,00
nbchoixrep	-0,07	0,047	-0,254	-1,412	0,167
Motrepontexte	-0,01	0,004	-0,243	-1,565	0,127
nbmotQ	-0,01	0,007	-0,169	-1,198	0,239
diffchoixrep	-0,23	0,061	<b>-0,631</b>	-3,746	<b>0,001</b>

Bref, de toutes nos variables choisies, celle concernant la complexité des choix de réponses serait celle qui nous aiderait à mieux prédire l'indice de difficulté des items.

Il est à souligner que des analyses avec le modèle de Rasch ont été effectuées afin de comparer les résultats obtenus avec la théorie classique des tests. Concernant les textes, on obtient un coefficient de détermination de 0,13. Pour les questions il est égal à 0,31 et pour les choix de réponses celui-ci est de 0,34. De plus, on se retrouve à observer les mêmes variables qui ressortent lors des analyses des coefficients  $\beta$ . Puisque les résultats se ressemblent, nous avons décidé de ne présenter que ceux obtenus avec la théorie classique des tests.

Au chapitre suivant, soit la discussion, des indications concernant l'interprétation de ces résultats sont présentés ainsi les enjeux que ceux-ci peuvent avoir dans le domaine de l'enseignement de l'anglais, langue seconde.



## **Chapitre V**

### **Discussion**

Dans le chapitre précédent, nous avons vu quelles variables pouvaient avoir une influence sur la proportion de réussite (indice de difficulté) selon la théorie classique des tests en compréhension de textes. Dans ce chapitre, nous expliquerons la signification de ces résultats et ce qui pourrait être fait en salle de classe pour évaluer les étudiants lorsqu'ils seront en situation d'évaluation de leur compétence en compréhension de textes. Finalement, à l'aide des coefficients de régression non standardisés obtenus, un outil informatisé sera proposé afin que les enseignants puissent prédire le niveau de difficulté de leurs propres items

#### 5.1 Explication concernant les textes

Les analyses concernant les textes nous ont démontré qu'il y avait 14 % de variance au sein du modèle et deux variables indépendantes qui pouvaient prédire les résultats. Malgré le fait que ces variables ne sont pas statistiquement significatives, nous pouvons tenter de donner des interprétations quant à la signification de ces résultats. Pour débiter, nous avons vu que les premiers mots du vocabulaire des élèves ont une certaine importance pour accéder à la bonne compréhension des textes ( $\beta = -3,92$   $p > 0,05$ ). Ces analyses ont démontré que même si notre texte contient peu de mots de cette catégorie, ils réussissent quand même à le comprendre. Ceci fournirait des pistes aux enseignants afin qu'ils puissent utiliser des textes beaucoup plus enrichissants qui développeraient la culture anglophone des élèves.

De plus, la deuxième compétence stipule que les élèves devraient être en mesure de lire et prendre connaissance de textes variés, ce qui est en lien avec cette affirmation. En ce qui concerne la longueur des textes, plusieurs enseignants croient que si un texte est trop long, les élèves vont se décourager et ne pourront pas le comprendre. Cependant, nos résultats semblent démontrer le contraire. D'autres analyses, avec un plus grand échantillon d'items, seraient cependant appropriées pour vérifier si la relation entre la valeur de l'indice de difficulté et ces deux variables indépendantes est statistiquement significative.

Une relation positive a été observée entre la proportion de réussite et le nombre de phrases que contient notre texte ( $\beta = 2,88$ ,  $p > 0,05$ ). Ceci pourrait encourager les professeurs à utiliser des textes qui sont un peu plus longs que la norme. Une explication derrière cette relation serait que lorsqu'un texte contient plus de phrases, plus d'informations sont présentées et qu'ainsi, la compréhension du texte est facilitée. Les élèves sont alors mieux outillés pour répondre aux questions accompagnant le texte. Il serait donc primordial d'effectuer d'autres recherches concernant cet aspect avec un échantillon d'items plus grand afin de valider cette hypothèse.

À l'intérieur du contexte théorique, on a pu voir que le choix des textes est un aspect crucial pour la compréhension. Stenner (1996) a soulevé le fait que si un élève se retrouve avec un texte qui n'est pas de son calibre, la réussite d'un item qui lui est associé se fera plus difficilement. Les analyses qui ont été effectuées nous portent à croire que les textes qui ont été choisis pour ce projet d'intervention ne sont peut-être pas appropriés pour les élèves qui se trouvaient dans l'échantillon. Ceci pourrait expliquer pourquoi les résultats obtenus n'étaient pas statistiquement significatifs : en fait, il ne semblait pas y avoir une grande variabilité entre les textes, ce qui ne facilite pas l'observation de liens statistiquement significatifs.

Par ailleurs, l'étude d'Enright et Sheehan (2002) a conclu que le contexte qui est choisi est tout aussi important que les items afin que les élèves aient le goût de lire le texte et s'investissent davantage dans la tâche.

Cobb (1997) a présenté des recherches où l'apprentissage soutenu par des listes de vocabulaire aide les apprenants à acquérir le vocabulaire nécessaire pour bien comprendre ce qui sera véhiculé dans le texte et ainsi répondre convenablement aux questions. Les enseignants pourront appliquer ce principe lorsque viendra le temps d'effectuer l'évaluation de la compréhension de textes. Une fois que les textes appropriés au niveau scolaire des élèves seront choisis, les enseignants pourront créer des listes de vocabulaire à présenter aux élèves avant la réalisation de la tâche, de façon à favoriser la bonne compréhension du texte. Par ailleurs, les recherches de Cobb (1997) ont démontré que plus un apprenant s'expose à un grand nombre de mots de vocabulaire, plus ses chances de rétention augmentent.

## 5.2 Explications concernant les questions

D'un autre côté, les questions que comportent nos compréhensions de textes sont tout aussi importantes que les textes pour les élèves. Nos résultats ont montré que notre modèle pouvait expliquer 32 % de la variance de la proportion de réussite des items, pourcentage qui est plus élevé qu'avec les analyses effectuées avec les textes : les variables indépendantes qui en découlent sont d'ailleurs statistiquement significatives.

D'une part, on a pu constater que le nombre de mots que l'on retrouve suite à notre question influence négativement la proportion de réussite des items ( $\beta = -2,68$ ,  $p \leq 0,05$ ). Plus le nombre de mots que l'on retrouve suite à nos questions augmente, plus la proportion de réussite diminue.

Donc, lorsque les questions se retrouvent déjà les unes à la suite des autres dans le texte et que les réponses sont faciles à retracer dans le texte, le travail des élèves est facilité. Aussi, les élèves ne font face à aucun piège et ils se sentent plus en confiance : ceci augmente la probabilité qu'ils réussissent les items.

Par extension, avec un ordre de présentation des questions qui est simple à suivre, les élèves auraient plus de facilité à se retrouver et à répondre correctement aux items. Par ailleurs, une deuxième variable concernant le nombre de caractères qui se retrouvent après la question, qui représente la longueur du texte, s'est montrée statistiquement significative ( $\beta = 2,51$ ,  $p \leq 0,05$ ). Les résultats nous indiquent que lorsqu'un grand nombre de caractères se retrouve après que la question soit posée, la proportion de réussite augmente. Une explication possible pour ce cas serait que les élèves doivent posséder assez d'informations pour pouvoir répondre correctement aux questions accompagnant le texte. Donc, si les élèves ont l'information nécessaire, ils auront plus de facilité à répondre aux questions. Ces deux résultats vont dans le même sens que ceux concernant la longueur des textes qui démontreraient que la longueur des textes n'est pas un signe automatique de difficulté. Les élèves doivent avoir les bons outils, donc des textes suffisamment informatifs, pour répondre aux questions.

Avec ces résultats, une importance devrait être accordée aux différents aspects qui entourent l'évaluation de la compréhension de textes. Comme Bormuth (1970) le suggère, il serait important que les enseignants présentent la matière en fonction des items auxquels les élèves devront répondre. De plus, Bejar (2002) suggère que le construit doit être bien cerné par l'enseignant afin de le présenter clairement aux élèves. Puisque nos résultats suggèrent que le contenu du texte soit bien compris afin que les élèves puissent répondre aux questions, il serait intéressant d'expliquer aux élèves comment déduire le sens d'un texte et identifier les éléments clefs de celui-ci.

En lecture, il existe différentes stratégies d'apprentissage qui font partie du programme (MELS, 2001) qui peuvent être utilisées en classe; ceci représente une excellente opportunité pour les utiliser plus régulièrement. Afin que les élèves comprennent leur utilité, les stratégies doivent être présentées individuellement dans des contextes appropriés. Comme l'ont soulevé Irvine et Kyllonen (2002), tout ce qui entoure le texte doit être bien saisi par les élèves.

Le principe de hiérarchie qui ressort de ces analyses est comparable à ce que Bormuth (1970) suggère dans son modèle. On se souviendra que les enseignants devront s'assurer de présenter les concepts en classe dans un ordre logique afin que les élèves puissent acquérir toutes les connaissances nécessaires et les appliquer par la suite. Bormuth (1970), comme d'autres auteurs, a soulevé le fait que les items qui nécessitent un effort cognitif spécifique doivent être regroupés ensemble afin de faciliter la tâche aux élèves : ce qui fera en sorte que le lien entre les items sera fait plus rapidement. Les enseignants peuvent présenter différentes situations aux élèves de façon à favoriser leur apprentissage quand ils seront en situation d'examens : l'information sera plus simple à retrouver.

### 5.3 Explications concernant les choix de réponses

Pour les choix de réponses, les résultats démontrent que la réalisation de ces derniers est plus simple si la réponse se retrouve mot pour mot ou est semblable à ce qui est présenté dans le texte, ( $\beta = -0,631$ ,  $p \leq 0,05$ ). De plus, notre modèle permet d'expliquer 38 % de la variance de la difficulté des items avec la combinaison de nos variables indépendantes, ce qui correspond au pourcentage le plus élevé en comparaison à la valeur observée aux analyses de régression associées aux textes, 14 % et aux questions, 32 %.

Si les élèves doivent faire part de leurs propres connaissances et faire preuve d'inférence, les questions sont moins bien réussies. Par contre, un questionnement survient suite à la lecture de ces résultats sur le lien qui existe entre les questions et les choix de réponses à savoir si les résultats obtenus lors des analyses sont plutôt expliqués par les questions et non par les choix de réponses. Même si on a pu constater lors de la lecture du contexte théorique que les choix de réponses ne doivent pas être trop difficiles et comporter trop de pièges, ceux-ci doivent tout de même représenter un défi auprès des élèves.

La recherche présentée par Enright et Sheehan (2002) nous a fait comprendre que plus il existait une variabilité entre les items, plus il était facile d'en cerner la difficulté. Donc, les items présentés doivent être structurés, avoir un niveau de complexité qui est adéquat pour les élèves et posséder un nombre de caractéristiques qui ne sera pas trop intimidant pour les élèves. Puisque plusieurs items affichaient un niveau de difficulté assez similaire, ceci pourrait nous amener à nous questionner sur la sélection des choix de réponses utilisés lors de cette étude ainsi que sur la variabilité entre les items afin que l'on puisse déterminer adéquatement leur niveau de difficulté.

Si nous regardons les aspects qui ont été soulevés dans le contexte théorique en comparaison aux résultats obtenus lors de ce projet d'intervention, on pourrait conclure que l'instrument de mesure choisi est acceptable, mais pourrait bénéficier de certaines modifications. Nous pouvons constater qu'il est vrai que les caractéristiques qui sont choisies pour déterminer le niveau de difficulté peuvent être utilisées lors de la production d'items.

Par ailleurs, le *Lexile Framework for Reading*, s'est démontré pertinent lorsqu'il a été comparé à d'autres instruments de mesure avec un coefficient de 0,91. Cela démontre que les principes concernant la sélection des textes par ce logiciel devraient être pris en considération. Enfin, on pourrait effectuer d'autres recherches concernant les caractéristiques des textes et leur influence sur la proportion de réussite des compréhensions de textes en anglais, langue seconde.

Au sein de ce projet d'intervention, les caractéristiques des items ont été pour leur part statistiquement significatives et pourront être utilisées dans d'autres contextes. Nous avons pu, grâce aux résultats, mettre sur pied une équation qui pourra prédire le niveau de difficulté d'autres items afin de permettre aux enseignants de présenter un outil d'évaluation adéquat pour leurs élèves. Bref, ceci nous permet d'explorer plusieurs pistes afin de continuer à trouver des solutions qui pourront améliorer les tâches présentées dans nos classes.

## CONCLUSION

Notre rapport s'inscrit dans le courant de recherche portant sur la production de tâches d'évaluation en enseignement. L'objectif de ce projet d'intervention était de proposer un modèle qui permet de prédire le niveau de difficulté des tâches en compréhension de lecture. Nous nous sommes intéressés aux recherches portant sur la production d'items et sur les outils pédagogiques qui permettent d'informer les enseignants sur le niveau de difficulté des items et des tâches.

Afin de réaliser notre projet d'intervention, deux périodes de collecte de données ont été effectuées. Ces collectes se sont faites auprès des élèves du 3<sup>e</sup> cycle du primaire qui étaient inscrits aux programmes réguliers et d'anglais intensif. On a présenté à ces élèves différents textes ainsi que des questions de compréhension qui s'y rattachaient. Par la suite, trois analyses de régression multiples, pour les textes, les questions et les choix de réponses, ont été effectuées afin de prédire le niveau de difficulté des items selon la théorie classique des tests. Pour expliquer en détail le résultat de ce projet d'intervention, ce sont les coefficients de détermination  $R^2$  et les coefficients de régression standardisés ( $\beta$ ) ou non ( $b$ ) qui ont été utilisés.

Les résultats à ces analyses démontrent que les différentes variables indépendantes étudiées avec les textes ne sont pas statistiquement significatives et ne nous permettent pas d'apporter des conclusions définitives sur cette dimension. En ce qui concerne les questions, on a remarqué que l'ordre dans lequel celles-ci sont présentées influence la proportion de réussite à l'examen. Ainsi, lorsque celles-ci sont présentées dans un ordre logique, les élèves réussissent mieux. Aussi, on a pu conclure que lorsque les élèves possèdent plus d'information pour les aider à répondre aux questions, ils performant mieux.



Pour les choix de réponses, lorsque ceux-ci sont moins complexes et se complètent plus facilement, la proportion de réussite des élèves augmente. De plus, lorsque le choix de réponses n'est pas présenté dans le texte et que les élèves doivent utiliser leurs propres connaissances afin d'y répondre, ceci fait en sorte que la proportion de réussites diminue.

Ce projet a permis de saisir la réalité qui est vécue dans les classes d'anglais, langue seconde, et a contribué à l'application d'un modèle qui pourra être utilisé par des enseignants dans leur milieu. De plus, ce modèle pourrait être généralisé à différentes disciplines. L'originalité de ce projet d'intervention se retrouve dans son sujet et du fait qu'elle permet d'examiner un aspect qui est souvent laissé de côté en évaluation, soit la mise en œuvre de procédures pour viser l'équivalence des interprétations en évaluation des apprentissages.

Comme tout projet de recherche, celui-ci possède certaines limites. D'une part, le petit échantillon d'items que nous avons nous a permis d'obtenir un nombre de données restreintes qui peut être une cause des résultats non statistiquement significatifs obtenus à certaines analyses. Aussi, puisque les variables indépendantes associées aux textes étaient non significatives, on pourrait se questionner sur le choix de ces dernières et aussi sur le fait que celles-ci n'aient pas une aussi grande variabilité. De plus, lors de l'analyse des résultats, on s'est aperçu que certains items avaient un niveau de difficulté trop similaire, ce qui limitait la variabilité des items. Dans ce contexte, certains items devraient être modifiés.

Selon les résultats obtenus, nous pouvons proposer des pistes qui pourront nous être utiles lors de recherches futures.

D'une part, il serait intéressant de faire des recherches concernant aussi les réponses construites courtes ou élaborées. On pourrait utiliser ces réponses à développement afin de voir si on obtient des résultats qui vont dans le même sens que ceux obtenus ici avec des items à réponse choisie. Une autre recherche qui porterait sur les questions et l'ordre de leur présentation nous permettrait de valider les résultats obtenus dans ce projet d'intervention.

De plus, à la suite de ces analyses certains questionnements surviennent concernant la variabilité des textes. Pour les élèves, il est primordial que durant leur apprentissage de la langue, une variété de textes leur soit présentée. Avec l'instrument qui a été utilisé ici, plusieurs formats ont été utilisés, mais on pourrait se questionner sur la variabilité entre ceux-ci. La sélection de texte doit être faite avec précision afin que les élèves puissent bien en comprendre le contenu. Sinon, les questions qui s'y rattacheront n'auront aucun poids et les résultats obtenus à l'évaluation n'auront aucune pertinence. Une recherche concernant la variabilité des textes serait pertinente dans le cas où on voudrait voir l'importance de cet aspect en compréhension de textes.

À la lumière de ce qui a été présenté plus haut, certaines étapes peuvent être entreprises afin de créer une application informatisée pour créer une échelle pour les enseignants en anglais, langue seconde. Cela nous permettrait de faire une application directe dans le milieu. Cette échelle aurait comme objectif de prédire l'indice de difficulté de leurs items et ensuite celui de l'instrument de mesure en considérant simultanément les trois dimensions étudiées : textes, questions et choix de réponses. À l'aide de ce modèle, les enseignants pourraient évaluer les questions qu'ils utilisent afin de créer un instrument de mesure adapté à leur clientèle.

Cette application a été réalisée à l'aide des résultats obtenus aux différentes analyses. Ces résultats nous permettront de bâtir des équations qui pourront alors être généralisées à différentes situations d'évaluation en compréhension de lecture en anglais, langue seconde. Cette application est disponible sous la forme d'un fichier Excel qui comporte une section pour les textes, pour les questions ainsi que pour les choix de réponses. Elle est déposée sur le site internet du Collectif sur le développement des applications en mesure et évaluation (<http://www.cdame.uqam.ca/activites/projets.html>). De plus, il y a une section où il est possible de faire la moyenne des trois aspects afin d'avoir un aperçu de l'indice de difficulté global pour tous les items, avec les trois dimensions confondues. Pour chaque aspect nommé ci-dessus, on y retrouve les variables indépendantes retenues lors de nos analyses. Un enseignant doit tout simplement introduire ses propres données afin d'obtenir une prédiction de l'indice de difficulté de chacun des items et le score moyen attendu à l'examen. Les informations concernant les calculs possibles à l'intérieur de cette application Excel sont disponibles à l'annexe C ainsi qu'un exemple de sa mise en œuvre.

## RÉFÉRENCES

- Bejar, I. (2002). Generative testing : from conception to implementation. Dans S. H. Irvine et P. C. Kyllonen (Dir) : *Item generation for test development*. Mahwah, New Jersey : Lawrence Erlbaum Associates.
- Bertrand, R. et Blais, J-G. (2004). *Modèles de mesure : l'apport de la théorie de la réponse aux items*, Sainte-Foy, Québec : Presses de l'Université du Québec.
- Black, W., Marsha et McAuliffe M., Michelle. (2001). *Arts and artists : through the centuries*. Westminster, California : Teacher created materials, Inc.
- Bormuth, J. R. (1970). *On the theory of achievement test items*. Chicago, Illinois : The University of Chicago Press, Ltd.
- Cobb, T. (2004). *Compleat lexical tutor V.6.2., for data-driven language on the Web*, [<http://www.lex tutor.ca/>]. Montréal, Québec.
- Cobb, T.(1997). *Concord to lexicon*. Thèse de doctorat inédite, Université Concordia, Montréal, Québec.
- Enright, M. K. et Sheehan, K. M. (2002). Modeling the difficulty of quantitative reasoning items : implications for item generation. Dans S. H. Irvine et P. C. Kyllonen (Dir) : *Item generation for test development*. Mahwah, New Jersey : Lawrence Erlbaum Associates.
- Flores, D., Moran, A. et Orzo, T. (2004). *Reading comprehension : using graphic organizers to teach literal, inferential and critical comprehension*. Huntington Beach, Californie : Creative teaching press.
- Folse, K. (2004). *Vocabulary myths: applying second language research to classroom teaching*. Ann Arbor, Michigan : University of Michigan press.
- Gunning, P., Lalonde, R., Schinck, M. et Watts, W. (2001). *A new TWIST to English : a competency-based approach*. Montréal, Québec : LIDEC.
- Irvine, S. H. et Kyllonen P. C. (2002). *Item generation for test development*. Mahwah, New Jersey : Lawrence Erlbaum Associates.
- Kinnear, P. R. et Gray, C. D. (2010). *IBM SPSS 18 statistics made simple*. New York, New Jersey : Psychology Press.

- Kucera, H. et Francis, W.N. (1967). *A computational analysis of present-day American English*. Providence, RI : Brown University Press.
- Newstead, S., Brandon, P., Handley, S., Evans, J. et Dennis, I. (2002). Approaches to modeling item-generative testing. Dans S. H. Irvine et P. C. Kyllonen (Dir) : *Item generation for test development*. Mahwah, New Jersey : Lawrence Erlbaum Associates.
- Nichols, S. (2007). Children as citizens: literacies as social participation. *Early years : an international journal of research and development*, 27(2), 119-130.
- Ministère de l'Éducation, des Loisirs et du Sport (2001). *Programme de formation de l'école québécoise. Enseignement primaire - Domaine des langues*. Québec, Québec : Gouvernement du Québec, 99-105.
- Snowling, M., Stainthorp, R. et Stuart, M. (2008). Literacy as a complex activity : deconstructing the simple view of reading. *Literacy*, 42(2), 59-66.
- Stenner, J, A. (1996). *Measuring reading comprehension with lexile framework*. Paper presented at the Fourth North American Conference on adolescent literacy. Durham, Caroline du Nord.
- West, M. (1953). *A general service list of English words*. London : Longman, Green and Company.

**ANNEXE A**  
**FORMULAIRE DE CONSENTEMENT POUR LES DIRECTIONS**  
**D'ÉCOLE**

## **Formulaire de consentement**

(direction d'école)

### **Projet de recherche d'intervention sur la production automatisée de tâches d'évaluation en anglais, langue seconde**

Votre école a été choisie pour participer à un projet d'intervention dans le cadre d'un projet de maîtrise de la Faculté des Sciences de l'éducation de l'Université du Québec à Montréal.

Ce projet vise, dans un premier temps, à élaborer différentes tâches d'évaluation en lecture qui pourront être utilisées par les enseignants. Dans un deuxième temps, il vise à soutenir les enseignants pour la préparation de leurs élèves de sixième année du primaire à l'épreuve ministérielle en anglais, langue seconde.

#### **Buts généraux du projet**

Ce projet poursuit les objectifs spécifiques suivants :

- 1) Déterminer différents niveaux de difficulté des questions associés aux textes.
- 2) Créer une banque de textes et de questions qui pourront être utilisées par les enseignants.
- 3) Aider les élèves à réussir la partie lecture de leur examen de synthèse.

#### **Procédure**

Ce seront les élèves de la sixième et cinquième année suivant le programme régulier et le programme intensif qui seront sollicités. Les élèves devront lire une série de textes différents et devront ensuite répondre à des questions en lien avec ces derniers. Il y aura deux sessions et chacune des sessions aura une durée de 45 minutes.

#### **Avantages et risques**

En participant à cette recherche, vous contribuez à l'avancement des connaissances et à l'amélioration de la réussite dans le domaine de l'anglais, langue seconde au primaire. De plus, vous nous permettez de développer du matériel pédagogique qui pourra être utilisé ultérieurement.

Soyez assuré qu'aucune de ces tâches ne sera utilisée comme source d'évaluation pour le bulletin des enfants.

### **Confidentialité**

Afin d'assurer la confidentialité des réponses des élèves, aucune information normative relative à l'identification des élèves participants ne sera conservée.

### **Participation volontaire**

La participation de votre école à ce projet d'intervention est volontaire. Ceci signifie que même si vous consentez aujourd'hui à ce que vos élèves y participent, vous demeurez entièrement libre à ce que votre école ne participe pas ou mettre fin à sa participation en tout temps sans justification ni pénalité. Vous pouvez ainsi retirer votre école du projet en tout temps.

En ce qui concerne les élèves qui ne participent pas au projet, des exercices leurs seront proposés pendant que les autres enfants participeront au projet.

### **Compensation financière**

Votre école ne recevra aucune forme de compensation pour votre participation au projet puisque celle-ci ne génère aucuns frais de votre part.

### **Questions sur le projet ou sur vos droits**

Si vous avez besoin d'information supplémentaire concernant ce projet avant de donner votre autorisation ou encore par la suite, n'hésitez pas à me contacter par courriel ([cristina\\_carlomusto@hotmail.com](mailto:cristina_carlomusto@hotmail.com)) ou contacter mon superviseur, M. Gilles Raiche, ([raiche.gilles@uqam.ca](mailto:raiche.gilles@uqam.ca)). Cela nous fera un plaisir de répondre à vos questions.

Sachez aussi que vous pouvez vous adresser au Comité institutionnel d'éthique de la recherche avec des êtres humains de l'UQAM (CIÉR). Toute inquiétude ou tout commentaire relativement aux questions d'éthique de la recherche peut être adressé au président du CIÉR, M. Joseph Lévy. Il peut être rejoint au numéro (514) 987-3000 poste 4483 ou au poste 7553.

### **Remerciements**

Votre collaboration est essentielle pour la réalisation de mon projet et je tiens à vous en remercier.



## **Consentement**

Je reconnais avoir lu le formulaire de consentement et consens volontairement à ce que mon école participe à ce projet. Je reconnais aussi que j'ai obtenu des réponses à mes questions de manière satisfaisante et que j'ai disposé suffisamment de temps pour réfléchir à ma décision de participer. Je comprends que ma participation à ce projet est totalement volontaire et que je peux y mettre fin en tout temps sans pénalité d'aucunes forme, ni justification à donner.

**Je consens à ce que mon école participe à ce projet de recherche sur l'élaboration des tâches d'évaluation en anglais, langue seconde** (veuillez cocher cette case)

Nom de votre école (en caractères d'imprimerie) :

\_\_\_\_\_

Votre signature : \_\_\_\_\_ Date : \_\_\_\_\_

**Veuillez conserver ce formulaire pour vos dossiers ou pour communiquer avec la personne responsable du projet d'intervention**

\_\_\_\_\_  
Cristina Carlomusto, étudiante à la maîtrise

**ANNEXE B**  
**FORMULAIRE DE CONSENTEMENT POUR LES ÉLÈVES**

**Formulaire de consentement  
(sujet mineur)**

Projet de recherche d'intervention sur la production automatisée de tâches  
d'évaluation en anglais, langue seconde

Chers parents,

Je sollicite la participation de votre enfant à un projet d'intervention pédagogique dans le cadre de mon projet de maîtrise à la Faculté des sciences de l'éducation de l'Université du Québec à Montréal. La direction de l'école de votre enfant a donné son accord à ce projet.

Ce projet vise, dans un premier temps, à élaborer différentes tâches d'évaluation en lecture qui pourront être utilisées par les enseignants. Dans un deuxième lieu, celui-ci vise à soutenir les enseignants dans la préparation de leurs élèves de la sixième année du primaire à l'épreuve ministérielle en anglais, langue seconde.

**Procédure**

Avec votre autorisation ainsi qu'avec son accord, votre enfant sera invité à lire en classe une série de textes et à répondre à des questions en lien avec les textes. Il y aura deux sessions et chacune des sessions aura une durée de 45 minutes. Aucun autre engagement et aucune autre sollicitation ne suivront.

**Avantages et risques**

En participant à ce projet d'intervention, votre enfant contribuera à l'avancement des connaissances et à l'amélioration de la réussite dans le domaine de l'anglais, langue seconde au primaire.

Soyez assuré qu'aucune de ces tâches ne sera utilisée comme source d'évaluation pour le bulletin de votre enfant.

**Confidentialité**

Afin d'assurer la confidentialité des réponses de votre enfant, chacun des questionnaires sera codifié anonymement. Aucune information normative relative à l'identification des élèves participants ne sera conservée.

### **Participation volontaire**

La participation de votre enfant à ce projet d'intervention est volontaire. Cela signifie que même si vous consentez aujourd'hui, votre enfant demeure entièrement libre de ne pas participer ou de mettre fin à sa participation sans justification à n'importe quel moment du projet. En ce qui concerne les enfants qui ne participeront pas au projet, des exercices leur seront proposés pendant que les autres enfants participeront au projet.

### **Questions sur le projet ou sur vos droits**

Si vous avez besoin d'informations supplémentaires concernant ce projet avant de donner votre autorisation, ou encore par la suite, n'hésitez pas à me contacter par courriel ([carlomusto.c@csgm.qc.ca](mailto:carlomusto.c@csgm.qc.ca)) ou contacter mon superviseur, M. Gilles Raïche ([raiche.gilles@uqam.ca](mailto:raiche.gilles@uqam.ca)). Cela nous fera un plaisir de répondre à vos questions.

Sachez que vous pouvez vous adresser au Comité institutionnel d'éthique de la recherche avec des êtres humains de l'UQAM (CIÉR). Toute inquiétude ou tout commentaire relativement aux questions d'éthique de la recherche peut être adressé au président du CIÉR, M. Joseph Lévy. Il peut être rejoint au numéro (514) 987-3000, poste 4483 ou au poste 7553.

### **Remerciements**

Votre collaboration est essentielle pour la réalisation de mon projet d'intervention et je tiens à vous en remercier.

## Consentement

Je reconnais avoir lu le présent formulaire de consentement et consens que mon enfant participe à ce projet. Je reconnais aussi que j'ai obtenu les réponses à mes questions et que j'ai eu assez de temps pour réfléchir à ma décision de participer. Je comprends que la participation de mon enfant à ce projet est volontaire et que je peux y mettre fin en tout temps, sans pénalité d'aucune forme, ni justification.

**Je consens à ce que mon enfant participe à ce projet d'intervention** (veuillez cocher cette case)

**J'autorise mon enfant à lire et à faire les différentes tâches en lecture en anglais mentionnées précédemment** (veuillez cocher cette case)

Nom de votre enfant (en caractères d'imprimerie) :

\_\_\_\_\_

Votre signature : \_\_\_\_\_ Date : \_\_\_\_\_

**Veuillez remettre un exemplaire de ce formulaire à l'enseignant de votre enfant**  
et vous pouvez conserver l'autre pour vos dossiers

**Merci de votre collaboration**

\_\_\_\_\_  
Cristina Carlomusto, étudiante à la maîtrise

**ANNEXE C**  
**INFORMATIONS SUR LES CALCULS UTILISÉS DANS**  
**L'APPLICATION EXCEL ET EXEMPLE DE MISE EN OEUVRE**

Les équations qui ont été produites à l'intérieur de l'application Excel (<http://www.cdame.uqam.ca/activites/projets.html>) ont été élaborées à partir de données issues d'un projet d'intervention pédagogique en anglais langue seconde. Il s'agit des équations 1, 2 et 3, disponibles à l'intérieur du projet d'intervention et reproduites plus bas. Pour chaque aspect étudié, soit les textes, les questions, et les choix de réponses, la moyenne pondérée par le coefficient de détermination respectif à chacun des trois aspects et la proportion de bonnes réponses prédites à chacun des items sont calculés. Le calcul de la proportion de bonnes réponses prédites à chacune des dimensions selon les caractéristiques des items est réalisé à l'aide des coefficients de régression non standardisés obtenus dans la présente recherche.

Pour utiliser l'application Excel, les trois étapes suivantes doivent être effectuées :

1. Il faut, en premier lieu, déterminer le nombre d'items qui constituent notre examen. Le calcul du score total et du pourcentage de bonnes réponses attendues au test ne sera valide que si nous avons ces données.
  2. Ensuite, il faut indiquer les valeurs de chacune des caractéristiques associées à chacun des trois aspects des items : une feuille du chiffrier Excel est disponible pour chacun de ces trois aspects. À chacune de ces feuilles, des indications sont aussi disponibles pour définir les caractéristiques ainsi qu'informer quant aux valeurs admissibles à celles-ci.
  3. À partir des valeurs inscrites par l'enseignant aux trois feuilles, le niveau de difficulté de chacun des items, le score total au test ainsi que le pourcentage de bonnes réponses au test sont calculés et reportés à une feuille nommée EXAMEN.
- Voici les équations qui ont été retenues pour chaque aspect. Il est à noter que les valeurs des coefficients de régression standardisés ont été arrondies à trois décimales. Dans l'application EXCEL, les valeurs utilisées vont au-delà de ces trois décimales.

$$\begin{aligned}
\text{Difficulté des textes} = & 0,668 + 0,000\text{textecaractere} - 0,035\text{typedetext} + \\
& 0,38\text{textephrase} - 0,006\text{motk1} + 0,005\text{motk2} + 0,025\text{motAWL} - \\
& 0,002\text{motofflist}
\end{aligned}
\tag{1}$$

$$\begin{aligned}
\text{Difficulté des questions} = & 0,613 + 0,005\text{nbmotQ} - 0,009\text{diffmotQ} + \\
& 0,059\text{diffmotK2} + 0,47\text{diffmotAWL} + 0,19\text{diffmotOffList} - \\
& 0,004\text{Qapmot} + 0,001\text{Qapcarac} + 0,004\text{Qmotttexte}
\end{aligned}
\tag{2}$$

$$\begin{aligned}
\text{Difficulté des choix de réponse} = & 1,34 - 0,07\text{nbchoixderéponse} - \\
& 0,01\text{mottxrep} - 0,03\text{nbmotQ} - 0,23\text{diffchoixrep}
\end{aligned}
\tag{3}$$

Afin de pouvoir utiliser l'échelle de prédiction, voici un exemple d'application pour trois items : facile, moyen et difficile. Vous trouverez dans le tableau qui suit, les items accompagnés de leurs choix de réponses.

Pour chaque item, les caractéristiques de chaque aspect ont été identifiées et ensuite introduites dans l'échelle. Le tableau 12 qui suit donne un exemple d'items facile, moyen et difficile qui ont été utilisés lors du projet d'intervention. Ensuite au tableau 13, on retrouve les différentes valeurs pour les aspects qui sont retrouvés dans l'outil. Ces chiffres sont ensuite introduits dans l'échelle dans leur catégorie correspondante. Une fois que ces chiffres sont introduits, les calculs sont faits automatiquement.



Tableau 12. Exemples de trois items pour la mise en œuvre du modèle de prédiction

---

Item facile : Groundhog Day is celebrated in Canada and in \_\_\_\_\_.

a) Africa  
b) America  
c) United States  
d) Toronto

Item moyen : Preparing a digital multimedia show is easy?

a) True  
b) False

Item difficile : What was Raven`s plan?

a) To get a moon  
b) To travel to the rich man`s house  
c) To buy a sphere  
d) To get a sun

---

Tableau 13. Valeurs à chacun des trois items pour les différents aspects du modèle de prédiction

<b>Items</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Niveau de difficulté prédit à partir des questions</b>	<b>(0,91)</b>	<b>(0,50)</b>	<b>(0,41)</b>
(Q01) Mots dans la question	8	7	4
(Q02) Difficulté avec les mots du premier corpus	5	5	3
(Q03) Difficulté avec les mots du deuxième corpus	0	0	0
(Q04) Difficulté avec les mots de la liste scolaire	0	0	0
(Q05) Difficulté avec les mots hors la liste de vocabulaire	3	2	1
(Q06) Nombre de mots après que la question est posée	14	133	133
(Q07) Nombre de caractères se retrouvant après que la question soit posée	78	743	684
(Q08) Nombre mots dans le texte et question	2	5	2
<b>Niveau de difficulté prédit à partir des choix de réponses</b>	<b>(0,73)</b>	<b>(0,53)</b>	<b>(0,40)</b>
(R01) Nombre de choix de réponses	4	4	4
(R02) Mots dans le texte et les choix de réponses	2	0	16
(R03) Mots dans la question et choix de réponses	8	7	4
(R04) Difficulté des choix de réponses	1	2	2
<b>Niveau de difficulté prédit à partir des textes</b>	<b>(0,52)</b>	<b>(0,45)</b>	<b>(0,15)</b>
(T01) Nombre de caractères dans le texte	2	1356	2298
(T02) Type de texte	1	1	2
(T03) Nombre de phrases dans le texte	12	17	39
(T04) Mots du premier corpus	108	199	380
(T05) Mots du deuxième corpus	5	16	18
(T06) Mots sur la liste scolaire des élèves	2	13	13
(T07) Mots hors la liste	0	20	34
<b>Niveau de difficulté pondéré prédit à partir des trois aspects</b>	<b>(0,77)</b>	<b>(0,51)</b>	<b>(0,36)</b>

Une fois que ces valeurs sont introduites dans les trois feuilles du chiffrier spécifiques aux trois aspects, on obtient les indices de difficulté prédits pour chacun des aspects sélectionnés qui sont arrondis à deux décimales. Selon les résultats obtenus avec l'application Excel, pour l'item 1, à l'aspect *questions*, on a une valeur de 0,91, de 0,50 pour l'item 2 et de 0,41 pour l'item 3.

Ensuite, à l'aspect réponse, le premier item affiche une valeur de l'indice de difficulté de 0,73, l'item 2 de 0,53 et l'item 3 de 0,40. Finalement, pour l'aspect des textes, l'item 1 affiche une valeur de 0,52, l'item 2, de 0,45 et l'item 3, de 0,15.

Les moyennes obtenues pour les trois aspects combinés sont ainsi de 0,77 pour l'item 1, de 0,51 pour l'item 2 et de 0,36 pour l'item 3. Ensuite, la moyenne pour les questions est de 0,60, pour les choix de réponses, de 0,83, et pour les textes, de 0,56. Finalement, le tout nous permet d'obtenir la moyenne de l'indice de difficulté pour les trois questions qui est de 0,66.

Enfin, en pondérant les trois niveaux de difficulté par le coefficient de fidélité associé à chacun des aspects, le score total moyen attendu au test constitué de trois items serait égal à 1,63, tandis que ce score transformé en pourcentage serait égal à 54%.

Bref, comme on peut le constater, on peut obtenir différents résultats selon nos besoins. Par la suite, nous pourrions effectuer les modifications nécessaires afin que nos items correspondent au niveau de nos élèves. Afin de connaître les caractéristiques de ces aspects, les enseignants peuvent se référer à [lexutor.ca](http://lexutor.ca) ou tout simplement se fier à leurs propres connaissances concernant le vocabulaire acquis de leurs élèves.