

*Pour le multilinguisme : exploiter à l'école la
diversité des contextes européens*

Rapport final – facette 2

Janvier 2008

Equipes de recherche

Frédérique ARTUS et Marc DEMEUSE

Institut d'Administration scolaire (INAS)

Université de Mons-Hainaut (UMH)

Monique REICHERT et Romain MARTIN

Groupe de recherche EMACS, Université du Luxembourg



Projet soutenu par la Commission européenne

Direction générale de l'éducation et de la culture

Programme Socrates - Convention : N°2005-2397/001-001 SO2 61OBGE

Coordination

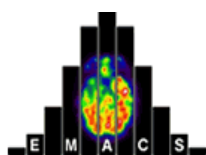


Université de Liège (Belgique)
Unité d'analyse des systèmes et des pratiques d'enseignement

Conception et réalisation



Université de Mons-Hainaut (Belgique)
Institut d'Administration Scolaire



Université du Luxembourg (Luxembourg)
Educational Measurement and Applied Cognitive Science

Partenaires



Ελληνικό Ανοικτό Πανεπιστήμιο (Hellenic Open University - Grèce)
Σχολή Ανθρωπιστικών Σπουδών (Ecole des Sciences Humaines)



Uniwersytet im. A. Mickiewicza (Pologne)
Instytut Filologii Romanskiej

SOMMAIRE

SOMMAIRE	3
Introduction	5
Chapitre 1 : Cadre théorique	7
1. Choix d'un outil permettant de mesurer les compétences linguistiques	8
2. Description de l'outil C-test	8
2.1. <i>Le C-test : caractéristiques générales</i>	8
2.2. <i>Le C-test : caractéristiques psychométriques</i>	10
2.2.1. La validité.....	10
2.2.2. La fidélité	11
2.2.3. L'objectivité	11
2.3. <i>Le C-test : caractéristiques pragmatiques</i>	11
3. La Construction des C-tests.....	12
4. Correction des résultats	14
5. Constats	14
Chapitre 2 : modalités d'administration	16
1. Utilisation du C-test dans le cadre du projet Socrates.....	16
1.1. <i>Choix des langues et des textes</i>	16
1.2. <i>Jeux de tests créés et planification du pré-test</i>	17
1.3. <i>Passation du pré-test dans les pays participants</i>	20
Chapitre 3 : constats concernant la mise à l'épreuve de l'outil	21
1. Défis rencontrés lors de la conception des textes C-test	21
2. Obstacles techniques	23
3. Défis posés lors de la passation	25
Chapitre 4 : résultats de la mise à l'épreuve des C-tests	26
1. Recueil de données	26
2. Analyse de la fidélité au niveau du C-test pour tout l'échantillon	28
2.1. <i>La fidélité des textes allemands</i>	29
2.1.1. Tous les pays (Communauté française de Belgique et Luxembourg).....	29
2.1.2. Par pays	29
2.2. <i>La fidélité des textes anglais</i>	30
2.2.1. Tous les pays (Communauté française de Belgique et Pologne)	30
2.2.2. Par pays	30
2.3. <i>La fidélité des textes français</i>	31
2.3.1. Tous les pays (Communauté française de Belgique, Luxembourg et Grèce)	31
2.3.2. Par pays	31
2.4. <i>La fidélité des textes néerlandais</i>	32
2.5. <i>La fidélité des textes polonais</i>	32
2.6. <i>La fidélité des textes grecs</i>	33
2.7. <i>Conclusions</i>	33
3. Analyse de la difficulté des textes	34
3.1. <i>Résultats en langue d'enseignement</i>	34
3.2. <i>Résultats en langue étrangère/seconde</i>	35
3.3. <i>La difficulté des textes allemands</i>	40
3.4. <i>La difficulté des textes français</i>	41

3.5. <i>La difficulté des textes anglais</i>	42
3.6. <i>La difficulté des textes néerlandais</i>	43
3.7. <i>Conclusions</i>	43
4. Analyse des inter-corrélations entre textes	44
5. Analyse des effets de séquence	46
6. Conclusions	47
Chapitre 5 : L'auto-évaluation	48
1. Présentation du questionnaire.....	48
2. Résultats	51
2.1. <i>Cohérence des items</i>	51
2.2. <i>Estimation du niveau de compétence linguistique par les élèves</i>	51
2.3. <i>Rapport entre les résultats à l'auto-évaluation et au C-test</i>	54
3. Conclusions	54
Conclusions et perspectives	56
Références	60
Annexes	63

Introduction

Le projet européen Socrates « Pour le multilinguisme : exploiter à l'école la diversité des contextes européens » vise « *une meilleure connaissance des attitudes des jeunes en obligation scolaire à l'égard d'une variété de langues de différents statuts et de leurs motivations au développement de compétences communicatives dans chacune d'elles, de leurs ressources en la matière, ainsi que la mise au point de pistes de nature à surmonter les obstacles au multilinguisme et la sensibilisation des différents acteurs à son intérêt* »¹. Le projet se divise en deux facettes. La première consiste en une prise de données relatives à la motivation à l'apprentissage des langues. La deuxième vise à compléter la première par une évaluation des compétences des élèves dans la première langue étrangère apprise afin d'étudier les liens entre attitudes, motivations et niveau de compétences. Le présent rapport expose les résultats du deuxième volet de l'étude - évaluation des compétences en langues - ; la première facette a fait l'objet d'un rapport distinct².

Deux visées complémentaires s'associent à l'objectif principal de la facette 2 de fournir des données sur les compétences linguistiques des élèves afin d'étudier les liens entre attitudes et motivations et niveau de compétences. D'une part, il s'agit de disposer d'une mesure objective des compétences de l'élève dans au moins une langue étrangère afin de mieux appréhender et discuter les limites éventuelles de l'auto-évaluation des compétences en tant qu'indicateur des compétences réelles dans un contexte international. D'autre part, on cherche à identifier les problèmes potentiels lors de la planification et la réalisation, à grande échelle, d'une évaluation des compétences linguistiques. Ce, dans l'objectif de proposer des solutions praticables.

Ce rapport est le fruit d'un travail mené en collaboration par quatre des six équipes de recherche impliquées dans l'étude, grâce au soutien de la Commission européenne et des institutions auxquelles ces équipes appartiennent : l'Université de Mons Hainaut en Belgique (Frédérique ARTUS et Marc DEMEUSE), l'Université du Luxembourg (Romain MARTIN, Véronique PELT et Monique REICHERT), l'Hellenic Open University en Grèce (Georges

¹ Rapport d'activité de la deuxième étape du Projet 42/2006 Action 6.1.2. – 6.2 Socrates, Août 2006

² Androulakis, G., Beckmann, C., Blondin, C., Fagnant, A., Goffin, C., Górecka, J., Martin, R., Mattar, C., Meißner, F.-J., Mosorka, E., Nowicka, A., Pelt, V., Schröder-Sura, A., Skrivánek, H., Wilczyńska, W., & Wojciechowska, B. (2007). *Pour le multilinguisme : Exploiter à l'école la diversité des contextes européens. Résultats d'une étude internationale*. Liège : Éditions de l'Université de Liège.

ANDROULAKIS) et l'Université A. Mickiewicz de Pologne (Joanna GÓRECKA, Edyta MOSORKA, Agnieszka NOWICKA, Hanna SKRIVÁNEK, Weronika WILCZYŃSKA, Bernadeta WOJCIECHOWSKA). L'Université de Giessen en Allemagne (Christine BECKMANN, Franz-Joseph MEIßNER et Anna SCHRÖDER-SURA) a par ailleurs participé aux réflexions et échanges qui ont permis d'aboutir à la mise en place de cette expérience, tout comme l'Université de Liège en Belgique (Christiane BLONDIN, Annick FAGNANT et Christelle GOFFIN) qui assurait en outre la coordination de l'ensemble du projet.

L'équipe de l'Université de Mons-Hainaut a assumé la responsabilité de cette facette 2 et a travaillé de concert avec les différents partenaires lors des diverses phases clés du projet. Signalons le rôle plus spécifique joué par l'Université de Luxembourg (Romain Martin et Monique Reichert) pour son apport d'expertise dans la création du test C-test. Par ailleurs, chaque pays partenaire a contribué aux différentes phases de l'étude et a organisé le recueil de données dans son propre pays. L'analyse des résultats des tests a été prise en charge par les équipes de l'Université de Mons Hainaut en Belgique (Frédérique Artus et Marc Demeuse) et de l'Université de Luxembourg (Romain Martin et Monique Reichert).

Ce rapport est structuré en trois parties. La première partie présente le cadre théorique et méthodologique de l'enquête. La deuxième partie s'intéresse à l'élaboration et à la mise à l'épreuve (pré-test) de l'outil d'évaluation des compétences en langues. Les modalités d'administration des épreuves et l'organisation de la phase de pré-test sont exposées et analysées, mettant en évidence les limites atteintes dans la mise au point de l'outil. La troisième partie du rapport présente les analyses des données recueillies lors du pré-test sous trois angles principaux dont les mesures psychométriques sont le point central. Pour autant, toutes les données n'ont pu être exploitées suite aux difficultés exposées dans la deuxième partie. Enfin, en conclusion, le rapport énonce des pistes d'action, destinées aux institutions habilitées à fixer les orientations au niveau national ainsi qu'aux instances européennes, en matière d'évaluations internationales de compétences linguistiques.

Chapitre 1 : Cadre théorique

L'apprentissage des langues au sein de l'Union Européenne est une préoccupation prioritaire des responsables politiques européens (Commission européenne, 2005, pp. 84-89) et pourtant il n'y a encore que très peu d'évaluations internationales qui se sont centrées sur la mesure des compétences en langues, acquises au cours ou en fin du cursus scolaire obligatoire. Les raisons en sont multiples et la lourdeur des dispositifs d'évaluation est certainement l'un des obstacles prépondérants.

En effet, dans une approche communicative qui tend actuellement vers une perspective actionnelle (Conseil de l'Europe, 2001), l'accent est clairement mis sur les compétences de réception, production et interaction dans la langue dont l'évaluation repose sur des mises en situation ou productions « spontanées » qui demandent un investissement en temps difficilement compatible avec des enquêtes à large échelle. Si l'on suit les critères proposés par la Commission européenne (2004, § 69, p. 16) en matière d'indicateur dont la portée *« devrait, par exemple, (1) mesurer les compétences des apprenants dans au moins deux langues autres que la langue maternelle, (2) mesurer leurs compétences en lecture, écoute, production orale et écrite et (3) enregistrer ces compétences sur les échelles du Cadre européen commun de référence pour les langues »*, nous sommes loin du compte au niveau des outils existants qui ne peuvent être utilisés tels quels dans le cadre d'une évaluation internationale (Blondin, 2005).

Jusqu'à présent, seules les auto-évaluations des personnes interrogées dans le cadre d'un Eurobaromètre (INRA, 2001) ont pu servir d'indicateur des compétences linguistiques en langue(s) étrangère(s) même si, à moyen terme, la Commission souhaite mettre en place un « indicateur européen des compétences linguistiques » basé sur un outil d'évaluation exploitable au niveau international.

Les équipes de recherches impliquées dans le présent projet ont donc décidé de mettre à l'épreuve un outil qui réponde à une partie des exigences complexes qu'implique l'évaluation des compétences en langues à l'échelle internationale.

1. Choix d'un outil permettant de mesurer les compétences linguistiques

Le premier défi à relever a donc été celui du choix d'un outil adéquat pour mesurer d'une manière valide et fidèle les compétences linguistiques d'une population internationale. En effet, mener, dans plusieurs pays en parallèle et pour des langues différentes, une évaluation du niveau de maîtrise d'une langue seconde (LS) et/ou étrangère (LE) au niveau de plusieurs compétences (lecture, écoute, productions écrite et orale), peut engager un dispositif très lourd, coûteux en temps et en argent.

La question se pose alors quant aux alternatives possibles à des évaluations différenciées, en particulier lorsque l'objectif n'est pas de réaliser une certification au niveau individuel, mais d'estimer la distribution des compétences linguistiques à l'intérieur d'une population. Pour autant, l'alternative retenue devrait prouver qu'elle est en capacité de mesurer le niveau de maîtrise globale d'une langue de manière comparable à celle d'autres tests d'évaluation des compétences linguistiques. Dans le cadre de la réalisation de la facette 2, nos équipes de recherche ont opté pour la méthode du C-test qui fait l'objet d'un nombre croissant de publications scientifiques. Les sections suivantes exposent, sur la base d'une argumentation scientifique, les avantages de cet outil.

2. Description de l'outil C-test

2.1. Le C-test : caractéristiques générales

Le C-test est un test écrit qui mesure la compétence linguistique générale aussi bien en langue maternelle qu'en langue étrangère (Grotjahn, 2002). Il consiste à mettre en évidence la capacité à retrouver l'information de textes partiellement mutilés. C'est donc un outil qui permet d'évincer la problématique des questions ouvertes, de leur type et de leur niveau qui peut s'avérer discriminant selon la langue et la population ciblée (Monseur et Demeuse, 2004). Le C-test est une forme alternative au test de closure (Klein-Braley, 1985 ; Raatz & Klein-Braley, 1982).

La caractéristique essentielle définissant le C-test est la réduction de la redondance d'un texte donné³ en effaçant, à partir de la deuxième phrase, la seconde moitié d'un mot sur deux. Seules la première et la dernière phrase restent intactes, présentant ainsi un minimum d'information concernant le contenu du texte original. Dans ce contexte, un locuteur natif adulte est capable de reconstruire le message verbal abîmé en se référant à ses connaissances des règles et des structures de sa langue, alors qu'un apprenant non natif de cette même langue ne sait pas (encore) aussi bien « tirer profit » des redondances.

Le texte suivant présente un exemple extrait d'un C-test en langue anglaise (d'après Bisping & Raatz, 2002, p.152) :

The English are great lovers of competitive sports. The ga_____ peculiarly assoc_____ with Eng_____ is cri_____. Many ot_____ games t_____ are Eng_____ in ori_____, but ha_____ been ado_____ with enthu_____ in ot_____ countries; b_____ cricket h_____ been seri_____ and exten_____ adopted on_____ in t_____ Commonwealth, partic_____ in Aust_____, India, Paki_____ and t_____ West Ind_____. So a lik_____ for cri_____ seems to go along with an English spirit and attitude and English institutions.

La solution:

The English are great lovers of competitive sports. The game peculiarly associated with England is cricket. Many other games too are English in origin, but have been adopted with enthusiasm in other countries; but cricket has been seriously and extensively adopted only in the Commonwealth, particularly in Australia, India, Pakistan and the West Indies. So a liking for cricket seems to go along with an English spirit and attitude and English institutions.

³ On dit qu'un message est redondant lorsqu'il contient plus d'informations que celles nécessaires pour la compréhension du message. Lorsque le message n'est pas communiqué entièrement, par exemple lorsqu'une partie est abîmée, les parties manquantes peuvent être reconstruites à partir des parties qui sont restées intactes. La redondance est considérée comme étant une caractéristique essentielle du langage (Aylett, 2000; Pinker 1994) parce qu'un message verbal peut être abîmé dans maintes situations quotidiennes (sur une mauvaise photocopie, dans un message sonore dans une gare, dans les informations échangées lors d'une fête, etc.).

2.2. Le C-test : caractéristiques psychométriques

Les C-tests étant développés et adaptés au(x) échantillon(s) en fonction du contexte de recherche et des langues évaluées, il en existe un nombre important (voir, p.ex. Grotjahn, 2002b, 2006), souvent basés sur différents textes, et composés d'un nombre variable de textes et de mots à compléter. Néanmoins, leurs caractéristiques psychométriques restent stables.

2.2.1. La validité

La validité du C-test est généralement élevée : des corrélations hautes entre le C-test et les compétences tant en compréhension qu'en expression écrites et orales sont observées (voir, p.ex., Arras, Eckes & Grotjahn, 2002). Ainsi, des résultats à des C-tests ont pu être corrélés avec des tests plus complets, comme notamment le TOEFL (Hastings, 2002) ou le « English Placement Test » (EPT) (Jafarpur, 2002). Les corrélations avec l'EPT, qui mesure la compréhension orale, la grammaire, le vocabulaire et la lecture en anglais, se situent entre 0,84 (grammaire) et 0,87 (compréhension orale) (Jafarpur, 2002)⁴. Il existe aussi des corrélations élevées avec les dimensions du Test DaF, mesurant les compétences linguistiques d'apprenants en allemand. Pour ce test, une corrélation de 0,64 avec la performance en expression orale a pu être mise en évidence. Des corrélations légèrement plus élevées ont été constatées en compréhension orale et écrite, ainsi qu'en expression écrite (Arras, Eckes & Grotjahn, 2002).

Par ailleurs, une étude de comparabilité entre le Test de Connaissance du Français (TCF) – test critérié, certificatif et référencé au Cadre Européen Commun de Référence pour les langues (Conseil de l'Europe, 2001) – et un C-test français réalisé au Luxembourg a mis en évidence des résultats très positifs. Ainsi, les résultats du TCF ont en très grande partie été prédits par les données du C-test et ce, surtout en ce qui concerne le score global du TCF (coefficient de corrélation de 0,78 entre le C-test et le score global du TCF). De même, les données du C-test ont permis de faire des prédictions valides quant au positionnement d'élèves sur un des niveaux du Cadre Européen Commun de Référence pour les langues (Reichert, Keller & Martin, 2006).

⁴ Sur une échelle allant de 1 (corrélation parfaite) à -1 (corrélation inversée) en passant par 0 (absence de toute corrélation).

2.2.2. La fidélité

Grotjahn souligne des niveaux de fidélité très élevés, avec des alphas de Cronbach se situant entre 0,80 et 0,90⁵. La fidélité test-retest varie entre 0,70 et 0,85 (Grotjahn, 2002). Les résultats montrent donc que ce type de test mesure de manière cohérente l'ensemble des compétences en question. Ils peuvent être d'ailleurs comparés à d'autres tests comme par exemple celui du français langue étrangère, le Test d'Evaluation de Français (TEF) de la Chambre de Commerce et de l'Industrie de Paris, dont la fidélité globale est de 0,96 (alpha de Cronbach) pour l'ensemble des 150 items des deux épreuves de compréhension (orale et écrite) et de lexique/structure. Cette fidélité, toujours calculée sur le même échantillon,⁶ s'élève à 0,91 (compréhension écrite), 0,92 (compréhension orale) et 0,87 (lexique/structure) si on considère séparément chaque épreuve (Demeuse et al., 2005).

2.2.3. L'objectivité

Lors de l'élaboration d'un C-test et avant même sa passation, il est recommandé de préparer des listes de solutions attendues, ainsi que des alternatives aux solutions principales. La réalisation de pré-tests, au sens de test de l'outil, permet d'identifier les solutions alternatives acceptables, dans le sens où elles maintiennent la cohérence et le sens du texte. Ces listes préalables augmentent l'objectivité des C-tests par rapport à une procédure qui consisterait à accepter ou refuser des solutions en cours de correction. En outre, les listes sont garanties d'objectivité lors d'une correction impliquant plusieurs correcteurs.

2.3. Le C-test : caractéristiques pragmatiques

D'autres paramètres caractérisent le C-test : le temps de passation limité est à priori compatible avec les périodes horaires scolaires. Le temps requis pour la correction peut également être circonscrit une fois les paramètres définis. Le recours au support informatique, en fonction des possibilités techniques (saisie directe de la réponse des élèves sur un ordinateur ou lecture optique des formulaires), permet d'automatiser la passation du test et de réduire encore le temps de ces deux étapes. Dans la mise au point de l'outil, l'équipe de

⁵ Sur une échelle allant de 0 à 1.

⁶ Echantillon de 16.526 candidats qui ont passé le TEF en situation réelle

recherche a opté pour une correction manuelle des pré-tests doublée d'une saisie automatisée avec reconnaissance scripturale.

Dans les modalités pratiques d'organisation de cette évaluation, des aspects pragmatiques non négligeables sont la possibilité de faire passer les tests par des non spécialistes (simples surveillants), les coûts modestes de reproduction et la possibilité de centraliser la correction.

3. La Construction des C-tests

Quatre étapes conduisent à la construction de C-tests :

- a) Choix des textes initiaux : un ensemble de textes est choisi en adéquation avec la population cible (Grotjahn, 2002), en veillant à éviter un vocabulaire trop spécifique ou des sujets trop particuliers (p.ex. scientifiques). Hastings (2002) recommande ainsi d'éviter les textes "dealing with fantastic, esoteric, arcane, or technical topics beyond the experience of the ordinary person"⁷ (Hastings, 2002, p. 16).

- b) Choix de la règle : les textes choisis doivent être « mutilés » en suivant une règle précise, systématique et d'application homogène. La règle suivante a été la référence de notre expérimentation. A partir de la deuxième phrase, la moitié de chaque second mot doit être effacée. Pour des mots ayant un nombre impair de lettres, la « moitié plus une lettre » est éliminée. La phrase finale est, comme la première, gardée intacte afin de pouvoir offrir aux sujets un minimum d'informations de base. Concernant la mutilation des mots, il y a néanmoins différentes particularités, en fonction de la langue en question, à prendre en considération. Par exemple, pour la construction de C-tests en français, les mots composés contenant une apostrophe ou un trait d'union sont à considérer comme un seul mot. Dans le cas où ceci mène à une impossibilité de reconstruction du mot ou si trop d'alternatives peuvent être jugées comme étant des solutions correctes, il faut avoir recours à d'autres techniques : soit le mot composé est pris comme étant deux mots, soit on propose aux sujets la première lettre de la deuxième partie du mot composé.

⁷Textes « d'un genre fantastique, ésotérique, obscurantiste ou traitant de sujets techniques qui dépassent l'expérience d'une personne lambda » (traduction libre par les auteurs).

Jafarpur (1999) montre qu'un non-respect du principe canonique d'élimination de lettres décrit plus haut peut avoir un effet négatif sur la fidélité du C-test. Grotjahn (2002) souligne, qu'en outre, la validité peut être affectée lorsque des principes d'élimination autres que le canonique sont utilisés.

Un texte C-test typique comprend 20 ou 25 lacunes. Afin d'aboutir à une fidélité élevée, et en considérant comme « item » un texte entier, 3 à 5 textes C-test doivent composer le test.

- c) Préparation de la mise en forme : la partie manquante d'un mot est en principe indiquée par une seule ligne continue (p.ex. « l_____ jeunes enf_____ »). L'indication par un trait court pour chaque lettre manquante est utilisée dans certains cas de simplification d'un texte C-test. En général, le C-test est présenté aux sujets sous la forme de test papier-crayon⁸ mais des études récentes montrent que la passation informatisée du C-test engendre des résultats équivalents à ceux d'une passation papier-crayon en termes de fidélité, de performances moyennes et de dispersion de celles-ci (Bisping & Raatz, 2002).

- d) Pré-test : les textes ainsi créés sont à pré-tester, au sens de mis à l'épreuve avant l'utilisation à plus grande échelle, auprès d'un nombre restreint de locuteurs natifs.

Le respect de ces étapes suffit en général pour aboutir à un C-test présentant de bons à très bons résultats quant à sa validité et un haut niveau concernant sa fidélité. La comparaison des étapes et modalités d'élaboration d'un C-test avec celles d'autres tests d'évaluation de compétences linguistiques (par exemple avec celles d'un test de lecture ou d'écoute) est à l'avantage du C-test, en ce compris la procédure de qualité afin de garantir sa fidélité et sa validité. Nous examinerons dans la section suivante si de tels constats sont valables pour la phase de correction du C-test.

⁸ Une passation papier-crayon implique néanmoins un engagement à ne pas sous-estimer en ce qui concerne la correction, le codage et la saisie des solutions données.

4. Correction des résultats

Nous l'avons dit, un test est composé de plusieurs textes. On calcule dans un premier temps, pour chaque élève, le score par texte. Celui-ci est obtenu en calculant le pourcentage de mots correctement complétés par rapport à tous les mots mutilés d'un texte. Quant au score total d'un élève, celui-ci est obtenu en rapportant le nombre de mots correctement complétés au nombre total de mots mutilés dans l'ensemble du test. Le score par texte, ainsi que le score total d'un élève sont donc exprimés en pourcentage (du nombre total de mots mutilés).

D'autres scores peuvent être calculés en correspondance avec la règle de correction retenue : par exemple, comprenant ou non des solutions dans lesquelles des fautes d'orthographe n'ont pas été prises en compte, ou des solutions alternatives jugées correctes (Eckes & Grotjahn, 2006).

La décision de prendre en considération ou non des fautes d'orthographe pourrait dépendre, selon Arras, Eckes et Grotjahn (2002), et Grotjahn (2002), du niveau linguistique des participants. Ainsi, avec des apprenants avancés, ces fautes sont plutôt rares et la différence entre les scores comptabilisant ces fautes et ceux les négligeant est minimale – ce qui n'est néanmoins pas le cas pour des débutants. De plus, ces auteurs soulignent que l'objectivité et l'économie dans le codage sont beaucoup plus élevées si tous les mots contenant une faute sont rejetés, ce qui, en outre, assure une concordance entre les correcteurs. Au contraire, si on accepte un relâchement des règles d'orthographe afin d'augmenter les critères discriminants⁹, c'est l'utilisation de listes de variantes pré-sélectionnées qui permet de garantir l'objectivité.

5. Constats

Avant de résumer les qualités du C-test, rappelons que celui-ci ne permet pas de faire un diagnostic des niveaux atteints dans des domaines spécifiques d'une langue tels que les compétences en compréhension de l'oral, en vocabulaire, en expression écrite, etc. Cependant,

⁹ En fonction du niveau de compétence des sujets, il est parfois difficile de distinguer la compréhension d'un texte de sa non compréhension sans accepter les mots mal orthographiés

le C-test présente une plus grande économie par rapport à d'autres tests en termes de temps de construction, de passation du test et d'analyse des résultats.

Par ailleurs, la validité élevée du C-test, montrant que la force de cet outil réside dans l'évaluation d'une compétence linguistique globale, nous amène à le considérer comme étant une alternative possible à d'autres tests dans le cadre d'une évaluation des compétences linguistiques à plus grande échelle, et une très bonne mesure de la distribution des compétences linguistiques globales d'une population. Il permet également la mise en relation avec d'autres données relatives aux sujets (p.ex. leurs attitudes et motivations).

Finalement, il reste à souligner que les différentes caractéristiques psychométriques du C-test, notamment sa validité et sa fidélité, ont fait preuve de leur niveau élevé dans différentes langues secondes/étrangères (par exemple l'anglais, le français, l'allemand, l'espagnol, l'hébreu – voir les nombreuses publications à propos du C-test¹⁰) et pour des sujets originaires de différents pays. Il s'agit donc aussi, dans une certaine mesure, d'un outil ayant fait ses preuves sur un plan international – ce qui rejoint les considérations contextuelles du cadre de réalisation de la facette 2 du présent projet Socrates.

¹⁰ voir <http://www.c-test.de>

Chapitre 2 : modalités d'administration

1. Utilisation du C-test dans le cadre du projet Socrates

Etant donné les bonnes caractéristiques psychométriques du C-test évoquées précédemment, notamment sa validité et fidélité élevées, ainsi que l'économie lors de sa construction, les responsables du projet Socrates ont donc décidé de mettre cet instrument à l'épreuve des objectifs spécifiques poursuivis. Ainsi, l'évaluation objective des compétences dans une langue moderne au moyen d'un C-test est complétée par une auto-évaluation portant sur la même langue.

Afin de s'assurer de la validité et du niveau de difficulté des textes choisis, les modalités de pré-test ont privilégié la composition de plusieurs jeux reprenant des segments d'un même texte reproduits dans les différentes langues à tester. L'ordre de présentation des segments est alterné afin de vérifier si celui-ci influence les performances des élèves.

Conformément à la méthodologie de la première facette de l'étude, les élèves concernés sont ceux du grade 9, soit la 3^{ème} année secondaire dans quatre pays européens participant au projet, à savoir la Grèce, le Luxembourg, la Pologne et la Communauté française de Belgique (seule l'Allemagne n'y a pas pris part).

1.1. Choix des langues et des textes

Les tests ont été mis en œuvre en Grèce, en Pologne, au Luxembourg et en Communauté française de Belgique, à la fois dans la langue d'enseignement (LENS) et dans au moins l'une des langues étrangères (LE) ou langues secondes (LS)¹¹ enseignées afin de pouvoir comparer le niveau des compétences en langue étrangère à celui de la langue d'enseignement principale, ainsi que de juxtaposer le niveau atteint par une population en langue d'enseignement à celui d'une population l'apprenant en tant que langue seconde ou étrangère. Le Tableau 1 montre pour chaque pays participant les langues qui y ont été testées.

¹¹ Le statut des langues dans chaque pays est explicité dans le rapport de la facette 1 (Androulakis et al., 2007)

Tableau 1 : Pays participants et langues testées

Communauté française de Belgique	français, allemand, anglais, néerlandais
Grèce	grec, français
Luxembourg	allemand, français
Pologne	polonais, anglais

Pour la mise à l'épreuve de l'outil d'évaluation (pré-test), 16 segments, extraits de 8 textes (4 provenant de l'enquête PISA 2000¹² et 4 de l'enquête IEA 1990¹³) ont été sélectionnés pour être transformés en textes C-test en suivant les recommandations de Grotjahn (2002) précédemment exposées. La décision d'opter pour des textes provenant de ces études a été prise afin de garantir la qualité des traductions et donc une similitude du contenu sémantique. Par ailleurs, ces deux études s'adressant respectivement aux élèves de l'enseignement secondaire et primaire, le niveau de difficulté des textes était a priori diversifié.

1.2. Jeux de tests créés et planification du pré-test

Pour pouvoir tester les éléments constitutifs du test comme les effets éventuels d'un ordre de passation, les 16 segments de textes sont utilisés dans 8 livrets (voir annexe 1). Les livrets comprennent également en préambule des questions caractérisant l'élève (sexe, date et lieu de naissance, usage des langues), une page de questions d'auto-évaluation sur la compréhension écrite ainsi qu'un texte mutilé dans la langue d'enseignement pour permettre à l'élève de s'approprier la méthode de ce test. Chaque livret présente 4 segments de textes dans la langue d'enseignement et 4 segments de textes dans la langue étrangère ou seconde. Afin d'évaluer un effet éventuel de l'ordre de passation des différents textes, dans la moitié des jeux, on présente les textes suivant la séquence « langue d'enseignement » - « langue étrangère/seconde ». Cet ordre est inversé dans l'autre moitié (tableau 2).

¹² PISA: Programme for International Student Assessment, OECD (2002). Extraits des textes suivants : « Graffiti » ; « Les armes scientifiques de la police » ; « Bien dans ses baskets » ; « Le cadeau ».

¹³ IEA: International Association for the Evaluation of Educational Achievement. Extraits des textes suivants : « Le morse » ; « Les marmottes » ; « Grand-père » ; « Comment lire l'âge d'un arbre ». Ces textes nous ont été fournis par le Secrétariat de l'IEA, Amsterdam, Pays-Bas et par les partenaires nationaux.

Tableau 2 : Composition des jeux 1 à 8 au niveau des segments de textes

Jeu 1	LE/LS Cadeau 1 Baskets 1 Morse 1 Grand-père 1	LENS Cadeau 2 Baskets 2 Morse 2 Grand-père 2	Jeu 2	LE/LS Police1 Graffiti 1 Marmottes 1 Arbre 1	LENS Police2 Graffiti 2 Marmottes 2 Arbre 2
Jeu 3	LENS Cadeau 2 Baskets 2 Morse 2 Grand-père 2	LE/LS Cadeau 1 Baskets 1 Morse 1 Grand-père 1	Jeu 4	LENS Police2 Graffiti 2 Marmottes 2 Arbre 2	LE/LS Police1 Graffiti 1 Marmottes 1 Arbre 1
Jeu 5	LE/LS Cadeau 2 Baskets 2 Morse 2 Grand-père 2	LENS Cadeau 1 Baskets 1 Morse 1 Grand-père 1	Jeu 6	LE/LS Police2 Graffiti 2 Marmottes 2 Arbre 2	LENS Police1 Graffiti 1 Marmottes 1 Arbre 1
Jeu 7	LENS Cadeau 1 Baskets 1 Morse 1 Grand-père 1	LE/LS Cadeau 2 Baskets 2 Morse 2 Grand-père 2	Jeu 8	LENS Police1 Graffiti 1 Marmottes 1 Arbre 1	LE/LS Police2 Graffiti 2 Marmottes 2 Arbre 2

L'extrait à la page suivante est issu du texte « *Arbre* ». Dans les jeux 2 et 4 (version français allemand), son premier segment est utilisé en allemand et le second en français. Le jeu 2 propose les textes en langue étrangère en premier alors que le jeu 4 utilise les mêmes textes mais en commençant par les versions en langue d'enseignement. L'extrait montre une différence dans le nombre de cases à compléter pour les mots mutilés selon la langue : en effet, un nombre constant de cases est présenté pour chaque texte dans une même langue (p.ex. 6 en français, 11 en allemand). Ce nombre constitue le nombre maximum de cases nécessaires pour compléter les mots des textes dans une langue donnée. Etant donné que la longueur maximale des mots à compléter varie significativement d'une langue à une autre, le nombre de cases varie en conséquence (voir aussi plus bas, page 24).

Le temps de passation est de 50 minutes, soit une période de cours. Il se décompose en 2-3 minutes pour répondre aux questions d’auto-évaluation et un maximum de 5 minutes par texte C-test. Afin de tester chacun des jeux dans chacune des classes, les 8 jeux différents sont distribués de manière séquentielle.

1.3. Passation du pré-test dans les pays participants

785 livrets ont été distribués dans les écoles sélectionnées (échantillon de convenance) par les partenaires. Le tableau 3 présente la répartition du nombre d’élèves par combinaison de langues et par pays.

Tableau 3 : Nombre de participants par combinaison de langues et par pays participant

Communauté française de Belgique	Livrets français-allemand: 80 Livrets français-néerlandais: 253 Livrets français-anglais: 222
Grèce	Livrets grec-français: 46
Luxembourg	Livrets allemand-français: 66
Pologne	Livrets polonais-anglais: 115

Chaque responsable d’établissement concerné a été convié à une présentation de l’expérimentation et a reçu des instructions précises quant aux modalités de passation des pré-tests. Dans le cadre de la mise au point d’un outil d’évaluation en langues, il était important de préciser à cette étape que l’objet principal du pré-test était l’outil lui-même et non les élèves et leurs performances. On trouvera en annexes 2 et 3 les versions françaises des documents mis à la disposition des directions et des professeurs des écoles concernées. Dans la plupart des classes, les chercheurs ont eux-mêmes procédé à la mise à l’épreuve des livrets ; dans les classes restantes, ce sont les professeurs de langues qui ont administré les questionnaires.

Chapitre 3 : constats concernant la mise à l'épreuve de l'outil

Au-delà des résultats des élèves, les partenaires ont dû faire face à des défis de plusieurs ordres. Ainsi, des contraintes méthodologiques, linguistiques, techniques et logistiques ont été mises en exergue lors de la conception et la mise à l'épreuve de l'outil.

1. Défis rencontrés lors de la conception des textes C-test

Le choix de textes identiques et disponibles dans toutes les langues impose une limitation dans la sélection. Les textes choisis pour la présente étude sont d'un niveau assez élevé. En effet, ceux-ci sont détournés de leur premier usage qui était la compréhension à la lecture dans la langue d'enseignement. Dans le contexte qui nous préoccupe, ils sont, d'une part, utilisés pour la mesure du niveau global dans la langue d'enseignement et, d'autre part, pour la mesure du niveau global dans une langue étrangère/seconde dont la maîtrise n'est pas équivalente pour la plupart des élèves.

Par ailleurs, pour certaines langues, le polonais et le grec en particulier, la construction de C-tests ne connaît pas de précédent dans la littérature. Le pré-test de textes C-test dans ces langues constitue donc une première. Les conséquences sur la sémantique du texte provenant de la mutilation de mots n'ont pu dès lors être anticipées avant que les textes ne soient mutilés.

En outre, l'application du principe « canonique » de mutilation des mots dans un texte – demandant de tronquer chaque deuxième mot à partir de la deuxième phrase d'un texte donné afin d'avoir un échantillon représentatif des termes utilisés dans des textes – mène à la difficulté que, pour un même texte disponible dans plusieurs langues, le « troncage » peut alors concerner différents types de mots et morphèmes selon la langue du texte.

Ainsi, un même texte transformé en C-test ne présente pas nécessairement la même difficulté selon la nature des mots qui sont mutilés dans une version linguistique par rapport à une autre. Par exemple, l'extrait du texte « Cadeau » de PISA ci-dessous montre, pour la comparaison de la version française avec l'anglaise, que pour les 20 mots mutilés par texte, il n'y en a que 7 qui sont sémantiquement équivalents pour les deux langues (mots grisés). Pour les autres mots

ou morphèmes, soit ceux-ci n'existent pas dans l'autre langue ou ne sont pas utilisés dans ce contexte (p.ex. « se » dans *se souvenait* par rapport à *could remember*), soit ils ne sont pas placés au même niveau chronologique dans la phrase et donc sont non concernés par la mutilation. En outre, il se peut que, pour les mots qui sont sémantiquement équivalents, des différences existent en ce qui concerne la flexion ou la conjugaison du mot (p.ex. « monter » (infinitif) par rapport à « rising » (gérondif)).

Français

Combien de jours, se demandait-elle, était-elle restée assise ainsi, à regarder l'eau brune et froide monter peu à peu et engloutir le promontoire ? Elle **se** souvenait à **peine** quand **la** pluie s'**était** mise à **tomber**, arrivant **par** le **sud** sur **le** marais **et** frappant **la** charpente **de** sa **maison**. Ensuite, c'**était** la **rivière** qui **avait** commencé à **monter**, d'abord **lentement**, jusqu'à **ce** qu'enfin elle s'arrête, **pour** repartir de plus belle. D'heure en heure, elle remplissait le lit des ruisseaux et les fossés et submergeait les basses terres.

Anglais

How many days, she wondered, had she sat like this, watching the cold brown water inch up the dissolving bluff. She **could** just **faintly** remember **the** beginning **of** the **rain**, driving **in** across **the** swamp **from** the **south** and **heating** against **the** shell **of** her **house**. Then **the** river **itself** started **rising**, slowly **at** first **until** at **last** it **paused** to turn back. From hour to hour it slithered up creeks and ditches and poured over low places.

Par ailleurs, le nombre de mots concernés peut varier en fonction de la langue du texte : le fait de laisser la première et la dernière phrase du texte intacte, et de tronquer chaque deuxième mot dans les phrases restantes, peut mener à un nombre de mots mutilés plus élevé dans une langue par rapport à une autre, en fonction du nombre total de mots utilisés dans le texte en question. Alternativement, appliquer la règle de tronquer exactement 20 mots (procédure appliquée dans la présente étude) dans chaque texte peut donner un pourcentage en unités sémantiques concernées plus ou moins élevé, de nouveau en fonction du nombre de mots employés dans le texte.

L'élaboration de textes C-test équivalents pour différentes langues pose donc question quant au choix du nombre de mots, au type de mots mutilés, et, en conséquence, quant à la compréhension générale du contexte qui peut être variable selon les langues.

2. Obstacles techniques

Plusieurs étapes sont nécessaires avant de pouvoir traiter les données au sens statistique du terme :

- **passation des livrets C-test** : diffusion dans les classes, renseignement du questionnaire et du C-test par les élèves ;
- **correction des C-tests** : mise au point des listes de réponses acceptées, correction manuelle et/ou informatisée des réponses données par les élèves ;
- **lecture des données** : lecture optique des livrets avec reconnaissance des données graphiques et scripturales ; vérification de la lecture faite ;
- **nettoyage des données** : levée des ambiguïtés et incohérences, élimination de livrets inexploitable ;
- **exportation des données** : sélection et mise en forme des données en vue de leur exploitation statistique à l'aide d'un logiciel spécifique.

Nous présentons dans cette section les obstacles rencontrés lors de ces étapes successives.

Une contrainte technique importante lors de l'organisation d'une évaluation à large échelle est le facteur temps. Une possibilité d'économiser du temps lors de la saisie des données peut être fournie par des passations sur ordinateur, permettant de sauvegarder toutes les entrées faites par les sujets lors d'une passation. Lors d'une évaluation internationale et à large échelle telle que celle envisagée dans le contexte du projet Socrates, il s'avère néanmoins très difficile, voire impossible, d'assurer une telle passation, étant donné que seule une minorité des écoles concernées disposent d'une infrastructure informatique adéquate. La procédure qui a dès lors été ici choisie pour la mise à l'épreuve de l'outil est celle de la mise en forme du C-test en format papier-crayon permettant ensuite la lecture optique des solutions proposées.

Néanmoins, des problèmes techniques pour la configuration du logiciel de lecture optique¹⁴ et la mise en page résultent des particularités du C-test : la reconnaissance des données

¹⁴ Le logiciel utilisé pour la lecture optique des livrets est le Cardiff TELEform, Cardiff Software Inc., Vista CA 92083.

scripturales (reconnaissance des lettres et des mots) demande une préparation spécifique plus conséquente par rapport aux données graphiques (noircissement d'une case) ; les langues comme le polonais et le grec présentent des caractères spéciaux qui doivent être paramétrés ; la mise en forme des documents repose sur des caractéristiques différentes selon les langues (longueur des mots par exemple).

Par ailleurs, la présence de données scripturales requiert l'élaboration de dictionnaires des mots à accepter lors de la lecture optique des livrets. Afin de disposer, au-delà des solutions originales proposées dans les extraits de textes, de variantes jugées également correctes, la correction des C-tests doit donc tout d'abord se faire sur la base d'une correction manuelle des livrets du pré-test afin de pouvoir lister ces variantes. Lors du pré-test, chaque partenaire a corrigé manuellement la partie des livrets rédigée dans la langue d'enseignement (le grec en Grèce, le polonais en Pologne, l'allemand au Luxembourg, le français en Communauté française de Belgique). Sur la base de ces guides de correction, les C-tests dans les langues étrangères enseignées ont alors été corrigés manuellement par le coordinateur de la facette 2.

Deux corrections manuelles ont donc été effectuées sur les livrets : une correction des livrets papier, qui permet de relever le nombre de bonnes réponses et de lister les variantes dans un tableau séparé, ainsi qu'une correction manuelle de chaque feuille de réponse « informatisée », telle que lue par le logiciel. Cette deuxième étape s'est avérée indispensable avant l'exportation des données. En effet, la lecture informatisée des données scripturales reconnaît automatiquement la forme des images. Les images sont alors interprétées en champs d'entrée sans l'assistance d'un correcteur, mais dans certaines limites seulement. Par exemple, la couleur de l'encre utilisée par l'élève, si elle est trop claire, peut nuire la qualité de la reconnaissance. D'autre part, l'écriture doit être homogène (toujours en majuscules, ou toujours en minuscules) sous peine de fausser sa lecture. De nombreux livrets ont présenté ces défauts et ont donc alourdi le travail de vérification.

Un module « vérificateur » intégré au logiciel permet également d'évaluer et de corriger les entrées qui n'ont pas été reconnues à la première lecture informatisée, compte tenu d'un seuil de tolérance de lecture prédéfini manuellement. Etant donné ces étapes de vérification intégrée et parfois manuelle après lecture par le logiciel des livrets, le temps de traitement effectif d'un livret de 13 pages varie de 10 secondes à 20 minutes.

La coordination a lu à l'aide du logiciel Téléform l'ensemble des questionnaires récoltés lors de la phase de pré-test¹⁵.

Une seconde phase de « nettoyage » des données a permis de relever certaines incohérences dues aux difficultés de prise de données. Les questionnaires illisibles ou défectueux (principalement dans l'échantillon de la Communauté française de Belgique) ont de ce fait dû être éliminés. L'annexe 4 détaille les dysfonctionnements rencontrés.

L'ensemble de ces éléments a constitué un véritable obstacle opérationnel lorsqu'il s'est agi d'exporter et traiter les données dans l'objectif de mettre au point les textes composant le C-test pour l'enquête finale. Dès lors, les équipes de recherches ont dû décider de se limiter au stade du pré-test de l'outil élaboré et d'utiliser les données dans leur état pour effectuer les analyses prévues initialement.

3. Défis posés lors de la passation

La mise à l'épreuve de l'outil a également permis de mesurer la durée de passation effective lors de l'administration des livrets par les partenaires. Certains ont constaté que le temps imparti était trop court pour plusieurs raisons dont la rapidité de lecture surestimée dans le protocole de départ, ou l'effet fatigue chez les élèves qui enchaînent les livrets des facettes 1 et 2 du projet (nécessitant deux périodes de cours). Le découragement de certains élèves s'explique par la forme inhabituelle et inconnue du test et par la difficulté des textes qui provoque chez certains un blocage. Par ailleurs, les chercheurs ont été confrontés à l'incompréhension de certains professeurs de langue par rapport à ce type de test, inconnu dans les pratiques de classe. Ainsi, les enseignants ont parfois remis en cause la validité du C-test au vu de la difficulté de certains textes le composant.

¹⁵ A l'exception du Luxembourg : comme ce partenaire dispose également du matériel nécessaire à la lecture optique, il a été jugé plus simple de procéder localement à la lecture des questionnaires complétés par les élèves luxembourgeois.

Chapitre 4 : résultats de la mise à l'épreuve des C-tests

1. Recueil de données

Sur la base des épreuves de C-test mises au point, les différents jeux de pré-test constitués ont été administrés dans 4 des 5 pays participant à l'étude. Le tableau 4 synthétise la participation au recueil des données ; il présente le nombre de questionnaires complétés, lus par lecture optique et disponibles pour les analyses.

Tableau 4 : Langues concernées par l'évaluation au moyen du C-test et nombre d'élèves.

	Langues évaluées (par combinaison de langues)	Nombre d'élèves	Données
Belgique	Français - Allemand	80	Bio - AE - C-test
	Français - Anglais	228	
	Français - Néerlandais	222	
Grèce	Grec - Français	46	AE - C-test
Luxembourg	Français - Allemand	66	Bio - AE - C-test
	Allemand - Français		
Pologne	Polonais - Anglais	115	AE - C-test
Total	6 langues concernées	757	

La colonne « données » indique les informations recueillies à propos de chaque élève et pour chaque langue évaluée par pays :

- bio : des informations de type biographique (soit au travers du questionnaire de la facette 1, pour le Luxembourg et la Grèce qui ont administré les instruments des 2 facettes aux mêmes élèves lors de l'administration du pré-test, soit au moyen de la partie du questionnaire de la facette 1 inséré dans le livret de la facette 2 pour la Communauté française de Belgique et la Pologne, chez qui les 2 facettes ont concerné des élèves différents).
- AE : Une épreuve d'auto-évaluation portant sur la compréhension écrite, construite à partir du Portfolio européen des langues.
- C-test : Un des 8 jeux du C-test.

Des enseignements intéressants peuvent être dégagés des analyses descriptives qui ont été effectuées.

Cependant, tous les échantillons recueillis lors du pré-test ne sont pas nécessairement représentatifs et limitent donc les analyses initialement prévues dans la phase de pré-test. En effet, le nombre de quinze sujets¹⁶ par jeu n'est pas atteint pour toutes les langues, notamment concernant l'échantillon grec et luxembourgeois et les allemands de l'échantillon belge. Par ailleurs, le nombre de sujets par item, pays et langue n'est pas très élevé (à partir de 9 sujets par item, langue et pays – à l'exception du français en Communauté française de Belgique qui rassemble le plus grand nombre de sujets – de 118 à 150 selon les items, voir aussi annexe **). Il faut donc rester prudent quant aux interprétations des résultats.

Ces échantillons ne sont pas non plus nécessairement représentatifs de la population visée pour l'enquête principale, étant donné qu'il s'agissait en premier lieu de voir dans quelle mesure des C-tests développés dans des langues différentes ne sont pas trop difficiles et en conséquence irréalisables par les élèves, d'avoir une notion de la fidélité des C-tests construits et de rassembler des variantes de mots jugées correctes.

Par ailleurs, des erreurs de paramétrage du programme de lecture optique pour l'élaboration des livrets du pré-test ont été relevées sans pouvoir être toutes corrigées dans les données belges. Par conséquent, des jeux, des questionnaires ou encore un texte mal configuré, ont dû être retirés des données. Ces éléments seraient à prendre en compte et à corriger lors de la configuration d'une version définitive du test, étape qui n'a pas été entreprise dans le cadre de cette étude de faisabilité.

Trois langues – l'allemand, l'anglais et le français - ont été testées dans deux ou trois pays, le polonais et le grec n'étant testés qu'en tant que langue d'enseignement dans leur propre pays. Le néerlandais, langue seconde en Belgique, est testé en tant que langue étrangère enseignée en Communauté française.

Tableau 5 : Comparaisons des textes d'une même langue entre pays (LENS : langue d'enseignement ; LE : langue étrangère ; LS : langue seconde)

	Français	Allemand	Anglais	Néerlandais	Polonais	Grec
Communauté française de Belgique,	LENS	LE (LS)	LE	LE (LS)		
Grèce	LE					LENS

¹⁶ Nombre minimum d'observations par item fixé initialement, jugé nécessaire, pour permettre de tirer des conclusions pertinentes sur la base du pré-test..

Luxembourg	LS	LENS				
Pologne			LE		LENS	

Huit jeux de textes ont été constitués. Comme nous l'avons précédemment présenté, ils correspondent à 4 paires de jeux où les textes sont identiques mais présentés dans un ordre différent. Dans les analyses suivantes, les calculs ont été réalisés en considérant ces 4 paires : jeux 1 & 3, 2 & 4, 5 & 7, 6 & 8.

2. Analyse de la fidélité au niveau du C-test pour tout l'échantillon

La fidélité d'un test concerne sa capacité à mesurer un construit de manière cohérente et exacte. Elle est exprimée par le coefficient **alpha de Cronbach**. Elle est calculée par langue pour tous les pays ayant fourni des données, et séparément par pays. Etant donné que les lacunes sont interdépendantes à l'intérieur d'un texte, la fidélité du C-Test est calculée sur la base des textes (*Grand-père1*, *Basket2*, etc.) considérés comme items. Les corrélations les plus élevées se retrouvent dans les échantillons à la fois les plus grands (français, anglais et allemand) et constitués respectivement de deux ou trois pays.

Les corrélations Item-Total : Idéalement, un seul item (un texte) devrait nous fournir les mêmes informations sur la compétence des sujets que le test complet. Il faut donc que les résultats d'un seul texte soient corrélés le plus haut possible avec le résultat du test entier. Les corrélations Item-Total nous montrent dans quelle mesure le résultat d'un seul texte est corrélé avec le résultat du test complet. Les résultats issus de nos pré-tests nous montrent une assez grande variabilité en fonction des langues et des textes concernés.

Ensuite, le coefficient **d'alpha de Cronbach après élimination de l'item** (nommé ci-après « *Alpha if Item Deleted* ») montre dans quelle mesure la fidélité du C-test en question augmente si l'on n'inclut pas l'item. Il a ainsi été observé que, pour toutes les langues, en fonction de l'élimination de l'un ou l'autre item, le coefficient a – à quelques exceptions près – tendance à diminuer.

Enfin, il faut noter que les textes en fin de jeu n'ont pas été complétés par un nombre variable de sujets dans un certain nombre de cas – par exemple parce qu'il ne restait plus assez de temps pour compléter le(s) dernier(s) texte(s), ou que les sujets ont tout simplement

abandonné après les premiers essais. Ces données manquantes seront appelées ici « missing à la fin du livret ». Les résultats ont été calculés avec ou sans « missing à la fin du livret ». Dans les paragraphes suivants, l'alpha de Cronbach constitue, sauf indication contraire, le coefficient calculé sur la base de toutes les données. Les missings n'ont été exclus que dans les cas où l'alpha de Cronbach différait significativement en fonction de l'inclusion ou non des missings.

2.1. La fidélité des textes allemands

2.1.1. Tous les pays (Communauté française de Belgique et Luxembourg)

La fidélité exprimée par le coefficient alpha de Cronbach, ainsi calculé pour les textes allemands, varie entre 0,94 et 0,95 en fonction des jeux considérés. Quant aux corrélations Item-Total, qui nous montrent dans quelle mesure le résultat d'un seul texte est corrélé avec le résultat du test complet, celles-ci se situent pour tous les items sur un niveau élevé voire très élevé entre 0,82 (*Grand-père1*) et 0,94 (*Police2*).

2.1.2. Par pays

Si l'on calcule la fidélité des textes par pays, on observe que le coefficient alpha est plus élevé pour le Luxembourg que pour la Communauté française de Belgique dans les quatre paires et en particulier pour les jeux 5 et 7 – où la fidélité pour l'échantillon belge est de 0,63, alors que le Luxembourg présente un alpha de 0,96. Ceci pourrait être dû au fait que la variance observée dans les résultats du Luxembourg est beaucoup plus élevée que celle des résultats de la Communauté française de Belgique. Il faut cependant prendre en compte dans notre interprétation le nombre peu élevé de sujets par pays (entre 13 et 22 élèves).

Par ailleurs, pour l'échantillon belge, on remarque une corrélation Item-Total négative avec l'item *Grand-père2* dans les jeux 5 et 7. L'alpha augmente de 0,13 lorsqu'on l'élimine. En éliminant les sujets n'ayant pas abordé ce texte (venant en dernière position parmi les textes allemands), cette corrélation négative ne disparaît pas, mais a même tendance à être plus exprimée encore. A nouveau, une explication pour ce résultat peut résider dans l'observation d'une variance très peu élevée dans les quatre items de ces jeux, et notamment dans les données valides pour le texte *Grand-père2*.

Enfin, les résultats des sujets de l'échantillon belge issus des jeux 1 et 3 donnent une corrélation Item-Total moyennement élevée (0,52) pour l'item *Cadeau1*, et l'alpha calculé pour ces jeux augmente (de 0,03) lorsqu'on élimine cet item du livret. Ce « comportement » n'est pas visible pour les données du Luxembourg.

2.2. La fidélité des textes anglais

2.2.1. Tous les pays (Communauté française de Belgique et Pologne)

La fidélité exprimée par le coefficient alpha de Cronbach varie respectivement entre 0,80 et 0,91 (pour les analyses incluant les données manquantes à la fin d'un livret) et 0,75 et 0,89 (pour les analyses ne considérant pas les données manquantes à la fin d'un livret) selon les jeux considérés.

Quant aux corrélations Item-Total, elles se situent pour tous les textes entre 0,62 (*Police1*, sans missings) / 0,63 (*Grand-père1* avec missings) et 0,83 (*Basket2* sans missings) / 0,85 (*Basket2* avec missings). Les corrélations Item-Total varient significativement en fonction de l'inclusion ou non des données manquantes à la fin d'un livret, comme le montre le tableau 6.

Tableau 6 : les cinq textes en anglais présentant des corrélations Item-Total peu élevées

Textes présentant des corrélations Item-Total peu élevées	avec missings	sans missings
Grand-père 1	0,63	0,75
Grand-père 2	0,78	0,68
Police 1	0,41	0,62
Basket 1	0,69	0,68
Marmottes 2	0,68	0,51

2.2.2. Par pays

Vu le grand nombre des sujets n'ayant pas du tout abordé certains textes, les calculs des coefficients alpha par pays ne sont faits que sur la base des données n'incluant pas les données manquantes à la fin des livrets respectifs.

Si l'on calcule la fidélité des textes par pays, on observe que le coefficient alpha est plus élevé pour la Pologne que pour la Communauté française de Belgique à l'exception des jeux 6 et 8.

En outre, on observe une corrélation Item-Total faible (0,22) dans l'échantillon belge pour l'item *Grand-père2* dans les jeux 5 et 7, et l'alpha calculé pour ces jeux augmente (de 0,03) lorsqu'on l'élimine du livret. Des corrélations Item-Total moyennement élevées s'observent dans ces mêmes jeux, pour les items *Cadeau2* (0,43) et *Basket2* (0,40). Ces « comportements » ne sont pas visibles pour la Pologne.

Dans les jeux 6 et 8, il s'agit de l'item *Marmottes2* qui montre une corrélation Item-Total moyennement exprimée (0,47) dans l'échantillon belge, et un alpha augmentant de 0,05 lorsqu'on l'élimine.

2.3. La fidélité des textes français

2.3.1. Tous les pays (Communauté française de Belgique, Luxembourg et Grèce)

La fidélité, exprimée par le coefficient alpha de Cronbach varie entre 0,82 et 0,90 en fonction des jeux considérés.

Analysant l'alpha de Cronbach en fonction de l'élimination de l'un ou l'autre item (« Alpha if Item Deleted »), on constate que le coefficient a tendance à diminuer, sauf si l'on élimine l'item *Cadeau2* (jeux 1 et 3 – augmentation de 0,01), *Cadeau1* (jeux 5 et 7 – augmentation de 0,02), ou *Police2* (jeu 2 et 4 - augmentation de 0,07) qui mènerait à un alpha légèrement plus élevé. La corrélation Item-Total que présentent ces items (entre 0,60 et 0,68 pour les deux premiers, et de 0,51 pour *Police2*), est réduite (mais toujours de niveau moyen) par rapport aux autres textes français utilisés (dont la corrélation s'élève jusqu'à 0,85).

2.3.2. Par pays

Si l'on calcule la fidélité des textes par pays, le coefficient alpha reste assez stable. Concernant les résultats grecs, la variance est très peu élevée et le nombre de sujets très restreint (N=9 à 16).

Pour la Communauté française de Belgique, il faut d'abord noter que les fidélités et les corrélations Item-Total ont été calculées sans considération des missings à la fin du livret,

étant donné le nombre important de sujets n'ayant pas complété les tests. Les résultats de l'échantillon du Luxembourg et de la Communauté française de Belgique montrent des corrélations Item-Total moyennement élevées pour certains textes. Ainsi, les items *Police2*, *Arbre2* ainsi que *Graffiti1* ne présentent des corrélations Item-Total qu'entre 0,55 et 0,63 dans les deux échantillons. C'est l'item *Marmottes2* qui présente une corrélation Item-Total réduite (0,44) dans l'échantillon luxembourgeois alors qu'elle est de 0,70 dans l'échantillon belge. Par contre, les items *Cadeau1*, *Grand-père1* et *Arbre1* ont des corrélations Item-Total entre 0,49 et 0,51 dans l'échantillon belge, alors qu'elles sont au-dessus de 0,65 dans l'échantillon luxembourgeois. Finalement, il reste encore à mentionner l'item *Cadeau2* pour lequel la corrélation Item-Total n'est que moyennement élevée dans l'échantillon belge (0,58), et l'alpha calculé pour ces jeux augmente (de 0,05) lorsqu'on l'élimine du livret.

2.4. La fidélité des textes néerlandais

Les textes néerlandais n'ont été soumis qu'en Communauté française de Belgique.

La fidélité exprimée par le coefficient alpha de Cronbach varie entre 0,69 et 0,83 (pour les analyses incluant les données manquantes à la fin d'un livret) et entre 0,65 et 0,76 (pour les analyses sans les données manquantes à la fin d'un livret). Ce sont notamment les items *Police1*, *Grand-père1* et *Grand-père2* qui montrent des corrélations Item-Total peu élevées, se situant entre 0,20 et 0,36 pour les données ne considérant pas les données manquantes à la fin du livret, et entre 0,52 et 0,59 pour les données incluant les données manquantes des fins de livret. Comme c'était déjà le cas pour les jeux anglais, les corrélations Item-Total varient significativement en fonction de l'inclusion ou non des données manquantes à la fin d'un livret.

2.5. La fidélité des textes polonais

La fidélité exprimée par le coefficient alpha de Cronbach varie entre 0,78 et 0,87 en fonction des jeux considérés. Les items présentant des corrélations Item-Total peu élevées sont *Police1* (corrélation Item-Total de 0,28) et *Arbre2* (corrélation Item-Total de 0,48). La conception des items n'étant pas identique pour tous, certains ayant moins de 20 lacunes (dont le texte *Police1*), les résultats obtenus sont donc aussi difficilement interprétables.

2.6. La fidélité des textes grecs

La fidélité exprimée par le coefficient alpha de Cronbach, calculée pour les textes grecs, varie entre 0,85 et 0,88 en fonction des jeux considérés. La corrélation Item-Total la moins élevée est observée pour l'item *Police1*, avec un coefficient de 0,34. Cependant, les données sont trop peu nombreuses, notamment pour les items (dont *Police1*) des jeux 1, 2 3 et 4., pour pouvoir en tirer des conclusions.

2.7. Conclusions

Le niveau de fidélité observé pour les différents jeux, notamment en considération du nombre restreint en items et sujets, peut être considéré satisfaisant jusqu'à élevé. Cependant, trois observations rendent difficile de tirer des conclusions définitives quant aux fidélités observées. Premièrement, le nombre important de données manquantes notamment à la fin des livrets mène à des divergences importantes dans les fidélités ainsi que dans les corrélations Item-Total observées en fonction de l'inclusion ou non de ces données. On observe en outre qu'il s'agit souvent d'un item présenté à la fin d'un set d'items dans une langue (p.ex. *Grand-père1* et *Grand-père2* pour l'échantillon belge, *Arbre2* pour l'échantillon polonais), qui présente des corrélations Item-Total faibles. Deuxièmement, on observe une difficulté générale de certains groupes d'élèves à compléter les textes en langue étrangère – comme l'anglais, l'allemand et le néerlandais par les sujets de la Communauté Française de Belgique – ce qui mène à des variances peu élevées, et qui affecte ainsi aussi les fidélités y relatives. Troisièmement, des convergences entre certains textes de différentes langues nous laissent présumer que des difficultés inhérentes à ces textes peuvent avoir mené à des résultats peu satisfaisants au niveau des fidélités et des corrélations Item-Total. Ainsi, le texte *Police1* montre des corrélations Item-Total peu élevées pour le français, le polonais et le grec. Un schéma identique est observé pour le texte *Grand-père2*, montrant des corrélations Item-Total peu élevées pour l'allemand, l'anglais et le néerlandais. Des analyses supplémentaires seront nécessaires pour analyser plus en détail le « disfonctionnement » de ces textes. En conclusion, les résultats au niveau de la fidélité des différents jeux de C-test sont satisfaisants à élevés, et, en général, les C-tests créés mesurent donc de manière cohérente la capacité linguistique globale des élèves. Seuls quelques textes, notamment *Police1*, *Grand-père1* et 2 ainsi que

Arbre2 ne pourraient à priori pas être intégrés dans un test final et nécessitent des analyses supplémentaires.

3. Analyse de la difficulté des textes

L'un des objectifs du pré-test était de pouvoir évaluer la difficulté des textes proposés et de s'assurer qu'ils sont d'un niveau adéquat pour mesurer les compétences linguistiques de la population visée. En préambule aux résultats, il est important de rappeler que chaque langue dispose d'un statut différent selon le contexte et le pays concerné. Le grec et le polonais sont testés en tant que langue d'enseignement uniquement et les pourcentages de réussite (taux de bonnes réponses par item) moyenne sont supérieurs à ceux des langues testées à la fois comme langue d'enseignement et langue étrangère (français et allemand) ou comme langue étrangère uniquement (anglais et néerlandais).

3.1. Résultats en langue d'enseignement

Les textes testés dans la langue d'enseignement (le français, le grec et le polonais) sont en grande majorité réussis avec une moyenne supérieure à 40% à l'exception des textes en grec. En effet, les moyennes de ces derniers se situent entre 33 et 78%. La moyenne maximale obtenue de plus de 75% pour deux textes démontre que l'exercice peut être compris et réalisable dans la langue grecque. Néanmoins, le nombre limité de sujets grecs nous invite une fois de plus à la plus grande réserve.

Quant au français testé en Communauté française de Belgique, les moyennes se situent entre 49% (*Marmottes2*) et 70% (*Grand-père2*). Pour le polonais, il s'agit de moyennes entre 40,5 (*Cadeau2*) et 81,3 (*Graffiti2*).

Pour le Luxembourg, il est un peu plus difficile de parler de « langue d'enseignement » ou de « langue seconde », étant donné que a) l'allemand ET le français sont utilisés en tant que langue d'enseignement au Luxembourg à partir de l'enseignement secondaire, et ce en fonction de la matière enseignée, et que b) des classes « francophones » ont participé au pré-test au Luxembourg, dont les élèves présentent en grande majorité des arrière-fonds langagiers romanophones, et dont les performances en allemand ne sont pas nécessairement comparables à celles d'autres classes scolaires du Luxembourg. En effet, on observe une moyenne relativement peu élevée en allemand (entre 24,2% et 66,4%). Une étude approfondie sur la corrélation entre le profil des élèves et leur taux de réussite pourrait sans doute expliquer ce phénomène et nous fournir de plus amples renseignements sur la possibilité à

compléter, en tant que locuteur natif, les textes français ou allemands. En effet, on observe que la moyenne de réussite pour les textes allemands augmente considérablement si l'on ne prend en compte que les élèves luxembourgeois ayant choisi la version allemande du livret (moyennes entre 50% et 87,5%).

Finalement, dans les quatre langues constituant des langues d'enseignement, pour une partie des élèves, on remarque qu'il leur est possible de compléter la majorité des textes à plus de 90 voire à 100%. Il s'agit de huit items en allemand, 10 items en grec et de 14 items en français et en polonais, dont plus de 90% des mots ont été complétés correctement par quelques élèves, montrant qu'il est tout à fait possible de trouver une solution à la plus grande partie des mots mutilés.

3.2. Résultats en langue étrangère/seconde

Quatre langues ont été testées en tant que langue étrangère/seconde : l'allemand (en Communauté française de Belgique et en partie au Luxembourg), l'anglais (en Communauté française de Belgique et en Pologne), le français (en Grèce et en partie au Luxembourg) et le néerlandais (en Communauté française de Belgique). Les résultats observés sont peu élevés, et se situent entre 4% (*Cadeau1* allemand en Communauté française de Belgique) et 36,7 % (*Basket1* français en Grèce).

Étant donné qu'on ne dispose pas d'un groupe de contrôle de locuteurs natifs anglais ni néerlandais, il est difficile de se prononcer sur la possibilité pour un locuteur natif de compléter les textes à 90-100%.

Intéressant à observer également est la cohérence qui existe pour le rang de difficulté des différents items, analysé pour chaque séquence de quatre items d'une même langue par deux jeux (p.ex. *Police1*, *Graffiti1*, *Marmottes1*, *Arbre1*, compris dans les jeux français 6 et 8). Cette cohérence est observée non seulement dans une même langue entre les pays, mais aussi d'une langue à une autre :

- L'item *Cadeau2* est, pour les six langues testées (le français, l'allemand, le néerlandais, l'anglais, le grec et le polonais) le texte le plus difficile, à l'exception du français et de l'allemand testé en Communauté française de Belgique ;
- L'item *Cadeau1* est aussi le texte le plus difficile à résoudre par les élèves, à l'exception du français, ainsi que de l'anglais en Pologne (avant-dernière position) ;

- L'item *Basket1* est soit le dernier, soit l'avant-dernier texte dans le rang de difficulté (en assumant une difficulté croissante) – à l'exception de l'allemand testé en Communauté française de Belgique ;
- L'item *Arbre1* est toujours le texte le plus facile (à l'exception du Grec et du Néerlandais);
- *Arbre2* est, pour toutes les langues et tous les pays, soit le texte le plus facile, soit celui qui se trouve en deuxième position quant à son degré de difficulté (en assumant toujours une difficulté croissante) ;
- *Morse1* aussi se trouve toujours en première ou en deuxième position quant à son degré de difficulté, indépendamment du pays ou de la langue testée ;
- Même constat pour *Morse2* : cet item se trouve toujours en première ou en deuxième position quant à son degré de difficulté (à l'exception de l'allemand testé en Communauté française de Belgique) ;
- *Police2* est en rang 3 de difficulté pour le néerlandais, le polonais, et le grec, ainsi que pour l'anglais testé en Communauté française de Belgique.
- *Marmottes2* est, pour trois des six langues étudiées (allemand, polonais et grec), ainsi que pour le français et l'anglais testé en Communauté française de Belgique, le texte le plus difficile. Pour le français testé au Luxembourg, et l'anglais testé en Pologne, ce texte se trouve à l'avant-dernier rang de difficulté. Seul pour le néerlandais, il s'agit du texte le plus facile.

Etant donné que cette recherche porte sur l'évaluation des langues étrangères et n'a utilisé la langue d'enseignement que comme point de comparaison, les analyses qui suivent se centrent sur la description des résultats concernant uniquement la difficulté des textes testés en tant que langue étrangère. Les tableaux 7, 8, 9, 10 et 11 quant à eux, synthétisent tous les résultats.

Tableau 7 : Synthèse des résultats par item pour le français

Français												
	Belgique	N Fr mean	r(it)	N Fr r(it)	Luxembourg	N Fr Mean	r(it)	N Fr r(it)	Grèce	N Fr Mean	r(it)	N Fr r(it)
	sans missings		sans missings		sans missings		avec missings		sans missings		avec missings	
Cadeau 1	67.6 (SD 20.5)	124	0.51	92	62.06 (SD 26.3)	17	0.8434	17	17.5 (SD 10.4)	4	0.9843	11
Basket 1	53.7 (SD 22.1)	115	0.63	92	57.5 (SD 20.5)	16	0.9255	17	36.67 (SD 5.8)	3	0.9649	11
Morse 1	66.04 (SD 17.6)	96	0.66	92	65.63 (SD 21.3)	16	0.8167	17	26.67 (SD 7.7)	3	0.9646	11
Grand-père 1	65.4 (SD 21.3)	92	0.49	92	70.7 (SD 21.4)	14	0.7022	17	12.5 (SD 3.5)	2	0.8558	11
Cadeau 2	55.21 (SD 22)	119	0.58	103	45 (SD 16.4)	15	0.7221	15	12.5 (SD 10.6)	2	0.8268	9
Basket 2	52.74 (SD 27.8)	113	0.83	103	58 (SD 20.2)	15	0.7816	15	10 (SD 14.1)	2	0.3806	9
Morse 2	65.19 (SD 37.1)	103	0.74	103	80.36 (SD 21.5)	14	0.7318	15	20 (SD 7.1)	2	0.8439	9
Grand-père 2	70 (SD 15.1)	33			70.36 (SD 13.2)	14	0.8413	15	25 (SD 21.2)	2	0.7004	9
Police 1	67.34 (SD 20.9)	109	0.58	80	62.5 (SD 25.6)	16	0.827	16	11.67 (SD 7.6)	3	0.8932	13
Graffiti 1	58.74 (SD 20.4)	99	0.59	80	64.69 (SD 13.9)	16	0.6148	16	20 (SD 8.7)	3	0.8394	13
Mammottes 1	53.99 (SD 17.9)	84	0.6	80	56.88 (SD 17.6)	16	0.7106	16	12.5 (SD 10.6)	2	0.871	13
Arbre 1	69.13 (SD 18.4)	80	0.5	80	67.5 (SD 26.1)	16	0.6549	16		0	0	13
Police 2	62.88 (SD 21.8)	132	0.55	93	64.17 (SD 20.4)	18	0.6289	18	27.5	2	0.7461	13
Graffiti 2					49.17 (SD 16.7)	18	0.5999	18	10	1	0.7461	13
Mammottes 2	49.26 (SD 21.5)	102	0.7	93	49.72 (14.9)	18	0.4428	18		0	0	13
Arbre 2	56.13 (SD 21.5)	93	0.58	93	62.5 (SD 13.3)	18	0.5988	18		0	0	13

Tableau 8 : synthèse des résultats par item pour l'anglais

Anglais								
	Belgique	N Ang mean	r(it)	N Ang r(it)	Pologne	N Ang mean	r(it)	N Ang r(it)
	<i>sans missings</i>		<i>sans missings</i>		<i>sans missings</i>		<i>avec missings</i>	
Cadeau 1	11.25 (SD 11.8)	48	0.5833	25	12.69 (SD 12.7)	26	0.861	22
Basket 1	15.48 (SD 13.2)	42	0.622	25	11.8 (SD 13.1)	25	0.888	22
Morse 1	19.7 (SD 14.8)	36	0.7011	25	22 (SD 16.6)	25	0.893	22
Grand-père 1	18 (SD 13.9)	25	0.6504	25	22.73 (SD 16)	22	0.868	22
Cadeau 2	9.89 (SD 9.3)	46	0.4276	23	14.13 (SD 17.5)	23	0.903	16
Basket 2	15.92 (SD 11.7)	38	0.4022	23	20.91 (SD 22.23)	22	0.967	16
Morse 2	16.9 (SD 14.9)	29	0.6328	23	26.58 (SD 27.1)	19	0.817	16
Grand-père 2	14.78 (SD 8.9)	23	0.2202	23	23.13 (SD 20.9)	16	0.778	16
Police 1	10.75 (SD 15.4)	40	0.6678	29	13.04 (SD 9.4)	23	0.556	22
Graffiti 1	16.57 (SD 11.3)	35	0.5403	29	27.17 (SD 20.3)	23	0.848	22
Marmottes 1	13.91 (SD 12.03)	32	0.6577	29	15.22 (SD 13.9)	23	0.724	22
Arbre 1	23.28 (SD 17.2)	29	0.7012	29	34.55 (SD 19.01)	22	0.787	22
Police 2	11.068 (SD 11.6)	52	0.6582	30	15.93 (SD 13)	27	0.738	21
Graffiti 2		0			24.26 (SD 17.5)	27		21
Marmottes 2	12.5 (SD 9.9)	32	0.4689	30	12.05 (SD 9.7)	22	0.584	21
Arbre 2	25.67 (SD 18.9)	30	0.7803	30	30.48 (SD 24.3)	21	0.749	21

Tableau 9 : Synthèse des résultats par item pour l'allemand

Allemand								
	Belgique	N All mean	r(it)	N All r(it)	Luxembourg	N All mean	r(it)	N All r(it)
	<i>sans missings</i>		<i>avec missings</i>		<i>sans missings</i>		<i>avec missings</i>	
Cadeau 1	4.29 (SD 4.6)	21	0.5129	22	30.38 (SD 26.3)	13	0.9	15
Basket 1	19.75 (SD 12.8)	20	0.7298	22	45.42 (SD 29.4)	12	0.94	15
Morse 1	21 (SD 10.2)	20	0.6453	22	57.08 (SD 23.6)	12	0.86	15
Grand-père 1	18.75 (SD 14.2)	16	0.6968	22	46.5 (SD 34.1)	10	0.83	15
Cadeau 2	13.44 (SD 11.4)	16	0.6332	16	24.23 (SD 22.9)	13	0.82	17
Basket 2	13.44 (SD 15)	16	0.5422	16	45.42 (SD 23.9)	12	0.95	17
Morse 2	11.33 (SD 8.8)	15	0.6079	16	57.73 (SD 21.4)	11	0.93	17
Grand-père 2	10.71 (SD 3.9)	14	-0.1206	16	45.91 (SD 27.3)	11	0.94	17
Police 1	13.64 (SD 13.3)	22	0.7018	22	41.79 (SD 28.8)	14	0.92	18
Graffiti 1	12.27 (SD 7.8)	22	0.6612	22	41.79 (SD 22.9)	14	0.85	18
Marmottes 1	17.11 (SD 16.8)	19	0.8285	22	41.79 (SD 24.5)	14	0.91	18
Arbre 1	21.18 (SD 13.5)	17	0.7835	22	49.62 (SD 22.8)	13	0.92	18
Police 2	28.08 (SD 16.9)	13	0.8481	13	52.73 (SD 21.3)	11	0.97	16
Graffiti 2		0		0	49.55 (SD 19.2)	11	0.95	16
Marmottes 2	16.92 (SD 20.1)	13	0.8589	13	40.91 (SD 27.8)	11	0.9	16
Arbre 2	37.31 (SD 23.1)	13	0.8479	13	66.36 (SD 18.9)	11	0.93	16

Tableau 10 : Synthèse des résultats par item pour le néerlandais et le polonais

	Néerlandais				Polonais			
	Belgique	N NL mean	r(it)	N NL r(it)	Pologne	N Pol mean	r(it)	N Pol r(it)
	<i>sans missings</i>		<i>sans missings</i>		<i>avec missings (peu élèves avec missings)</i>		<i>avec missings</i>	
Cadeau 1	6.96 (SD 6.9)	23	0.64	21	43.28 (SD 25.8)	29	0.5224	29
Basket 1	18.7 (SD 16.5)	23	0.62	21	54.83 (SD 27.9)	29	0.7141	29
Morse 1	15.87 (SD 14.9)	23	0.67	21	61.85 (SD 24.3)	29	0.8253	29
Grand-père 1	21.4 (SD 13.3)	21	0.36	21	68.1 (SD 32.7)	29	0.6849	29
Cadeau 2	14.86 (SD 10.8)	36	0.49	24	40.54 (SD 20.74)	28	0.8615	28
Basket 2	18.29 (SD 12.9)	35	0.79	24	74.82 (SD 21.5)	28	0.6943	28
Morse 2	20.36 (SD 11.5)	28	0.69	24	74.46 (SD 22.25)	28	0.7625	28
Grand-père 2	21.67 (SD 10.2)	24	0.3	24	68.75 (SD 25.3)	28	0.6169	28
Police 1	9.77 (SD 7.9)	43	0.2	30	78.33 (SD 16.8)	30	0.2827	30
Graffiti 1	21.13 (SD 15.2)	40	0.29	30	70.67 (SD 21.6)	30	0.7199	30
Marmottes 1	13.91 (SD 9.2)	32	0.74	30	74.5 (SD 21.1)	30	0.784	30
Arbre 1	19.83 (SD 11.9)	30	0.61	30	79.83 (SD 20.1)	30	0.5738	30
Police 2	15.51 (SD 16.9)	49	0.6	20	72.32 (SD 20.3)	28	0.6337	28
Graffiti 2		0			81.3 (SD 16.8)	28	0.6059	28
Marmottes 2	21.36 (SD 12.1)	22	0.47	20	69.17 (SD 18.8)	28	0.7571	28
Arbre 2	20.5 (SD 11.7)	20	0.5	20	81.07 (SD 21.5)	28	0.4761	28

Tableau 11 : Synthèse des résultats par item pour le grec

	Grec			
	Grèce	N Gr mean	r(it)	N Gr r(it)
	<i>sans missings</i>			
Cadeau 1	33.13 (SD 17.7)	8	0.93	3
Basket 1	35 (SD 28.3)	5	0.86	3
Morse 1	36.25 (SD 25.6)	4	0.88	3
Grand-père 1	70 (SD 22.9)	3	0.52	3
	<i>avec missings (seule 1 personne avec missings)</i>			
Cadeau 2	60 (SD 31.3)	11	0.77	11
Basket 2	69.55 (SD 22.4)	11	0.88	11
Morse 2	77.73 (SD 20)	11	0.86	11
Grand-père 2	75 (SD 30.2)	11	0.56	11
	<i>sans missings</i>			
Police 1	63.5 (SD 28.8)	10	0.34	4
Graffiti 1	47.86 (SD 27.5)	7	0.72	4
Marmottes 1	61 (SD 22.2)	5	0.88	4
Arbre 1	58.75 (SD 37)	4	0.7	4
	<i>avec missings (aucune personne avec missings)</i>			
Police 2	46.54 (SD 20.1)	13	0.74	13
Graffiti 2	48.85 (SD 22.7)	13	0.74	13
Marmottes 2	44.62 (SD 26)	13	0.74	13
Arbre 2	60.77 (SD 20.6)	13	0.67	13

3.3. La difficulté des textes allemands

La difficulté calculée pour les textes allemands (total des mots complétés correctement divisé par le nombre de mots mutilés par texte, multiplié par 100) se situe pour tous les textes entre 13,1% (*Cadeau1*) et 57,9% (*Morse2*) (la Communauté française de Belgique et le Luxembourg confondus). Les élèves ne réussissent donc à compléter en moyenne que la moitié des lacunes, ou moins. Cependant, dans la moitié des textes (N=8), quatre élèves complètent avec succès 90 à 100% des mots mutilés montrant qu'il est possible de trouver des solutions à (presque) tous les mots mutilés de ces items. Il existe également des scores maximum obtenus élevés dans les autres textes, avec des pourcentages entre 75 et 85%. Les données brutes (résultat par mot) nous révèlent également qu'il y a certains mots qui ne sont résolus par personne.

Dans les jeux 2 et 4 ainsi que 6 et 8, on constate une cohérence totale entre la Communauté française de Belgique et le Luxembourg en ce qui concerne la séquence de difficulté des quatre textes. Pour les jeux 1 et 3, il y a cohérence entre les deux pays pour le texte *Morse1* (texte le plus facile) et *Cadeau1* (texte le plus difficile pour les deux pays). Aucune cohérence ne se manifeste cependant pour les jeux 5 et 7. De manière générale, et pour les deux pays on peut donc retenir que les textes les plus faciles sont *Morse1* (jeux 1 et 3), *Arbre1* (jeux 2 et 4) et *Arbre2* (jeux 6 et 8), les plus difficiles étant *Cadeau1* (jeux 1 et 3), *Graffiti1* (jeux 2 et 4) et *Marmottes2* (jeux 6 et 8).

En analysant les résultats par pays, on remarque que les sujets luxembourgeois ont des taux de réussite légèrement plus élevés que les élèves de la Communauté française de Belgique. Cette différence est significative ($p < .05$) pour 9 des 16 textes utilisés (*Cadeau1*, *Morse1*, *Police1*, *Graffiti1*, *Marmottes1*, *Arbre1*, *Basket2*, *Morse2*, et *Grand-père2*). Les moyennes pour la Communauté française de Belgique se situent entre 4,3% (*Cadeau1*) et 37,3% (*Arbre2*), celles pour le Luxembourg entre 18,5% (*Cadeau2*) et 45,7% (*Morse1*).

Pour le Luxembourg on remarque cependant que les résultats diffèrent en fonction de la langue du livret que les élèves ont choisie pour la passation. Les élèves ayant choisi le livret allemand ont ainsi des résultats beaucoup plus élevés que les élèves ayant choisi le livret français (comme langue d'enseignement). Seule une analyse des habitudes de ces élèves quant

à l'usage des langues, mis en regard avec leurs résultats, pourrait confirmer si ce sont bien les élèves dont la langue usuelle est l'allemand qui tirent la moyenne des résultats vers le haut.

3.4. La difficulté des textes français

La difficulté moyenne calculée pour les textes français se situe (pour la Communauté française de Belgique, la Grèce, et le Luxembourg confondus) entre 28,9% (*Graffiti2*) et 61% (*Cadeau1*). Pour tous les items, à l'exception de *Graffiti2* et de *Marmottes2*, entre 4 et 55 élèves réussissent à compléter 95 voire 100% des lacunes. Pour les items *Graffiti2* et *Marmottes2*, certains élèves réussissent jusqu'à 85 % des lacunes, soit 17 complétées sur 20. Par ailleurs, chaque mot mutilé a été complété correctement par au moins 1 sujet. Il n'y a donc aucun mot qui n'a pas pu être résolu.

A nouveau, on remarque une certaine cohérence entre la difficulté relative des items entre la Communauté française de Belgique et le Luxembourg (les données de la Grèce ne sont pas reprises ici, étant donné le nombre restreint de sujets) : pour les jeux 2 et 4, il y a une cohérence totale entre la Communauté française de Belgique et le Luxembourg au niveau de la séquence de difficulté des quatre items. Pour les jeux 5 et 7, il y a cohérence entre les deux pays pour l'item *Morse1* (en deuxième position de difficulté) et *Basket1* (item le plus difficile pour les deux pays). Pour les jeux 6 et 8, il s'agit des items *Arbre1* (le plus facile) et *Marmottes1* (le plus difficile) qui montrent une cohérence entre les deux pays. Aucune cohérence ne se manifeste cependant pour les jeux 1 et 3. De manière générale, et pour les deux pays, on peut donc retenir que les textes les plus faciles sont *Police1* (jeux 2 et 4), *Arbre1* (jeux 6 et 8) et les plus difficiles *Marmottes1* (jeux 2 et 4), et *Basket1* (jeux 5 et 7).

En analysant les difficultés des textes par pays et en se limitant aux textes abordés par les élèves, on remarque que la Grèce a des résultats beaucoup moins élevés que les deux autres pays. Ainsi, les moyennes du Luxembourg et de la Communauté française de Belgique par item se situent entre 45% (*Cadeau2*, échantillon luxembourgeois) et 80,36% (*Morse2*, échantillon luxembourgeois), alors que le maximum obtenu par les élèves grecs est de 36,7% (*Basket1*). Il faut cependant souligner ici que la majorité des élèves grecs ne semble pas avoir traité l'ensemble des items français, probablement parce qu'ils étaient largement trop difficiles pour eux. Leurs non-réponses aux lacunes sont comptabilisées comme une absence de réponse correcte et affecte donc la moyenne générale des élèves grecs.

Pour le Luxembourg, on remarque à nouveau une certaine différence entre les élèves en fonction de la langue choisie pour le test. La différence n'est cependant pas toujours statistiquement significative.

3.5. La difficulté des textes anglais

La difficulté calculée pour les textes anglais se situe – pour la Communauté française de Belgique et la Pologne confondues – pour tous les items entre 7,9% (*Grand-père2*) et 18,4% (*Arbre1*) pour les données incluant les missings à la fin du livret, et entre 11,3% (*Cadeau2*) et 28,1% (*Arbre1*) pour les données n'incluant pas les missings à la fin du livret. Les élèves ne réussissent donc en moyenne que très peu d'items. Seuls 6 items sont réussis par quelques élèves avec des scores allant de 80 à 90%. Le maximum de points obtenus pour les autres textes se situe cependant entre 45 et 75%. Certains mots mutilés ne sont résolus par personne. Etant donné qu'on ne dispose pas d'un groupe de contrôle de locuteurs anglais natifs, aucune conclusion ne peut être avancée sur la possibilité de compléter ces textes à 90-100%, ni si ces mots peuvent être complétés par un locuteur natif.

En analysant les difficultés des textes (abordés par les élèves) par pays, on remarque que les élèves polonais présentent en grande partie des résultats légèrement plus élevés que les élèves de la Communauté française de Belgique. Cette différence est significative ($p < 0,05$) pour 2 des 16 textes utilisés (*Graffiti1* et *Arbre1*).

Les moyennes pour la Communauté française de Belgique se situent entre 9,9% (*Cadeau2*) et 25,7 (*Arbre2*), celles pour la Pologne entre 11,8 (*Basket1*) et 34,6% (*Arbre1*). Le classement des textes d'après leur difficulté dans les deux pays n'est pas identique en tous points (voir tableau 8) : seuls les jeux 2 et 4 présentent une cohérence totale entre les deux pays pour la séquence de difficulté des textes. Egalement, *Morse2* et *Arbre2* sont les items les plus faciles dans la séquence des quatre items utilisés respectivement, et ce pour les deux pays. Ceci est vrai aussi pour *Cadeau2* en tant qu'item le plus difficile.

3.6. La difficulté des textes néerlandais

Les textes C-tests néerlandais ont été administrés à un échantillon d'élèves de la Communauté française de Belgique qui apprennent cette langue en tant que langue étrangère dans un contexte politique où elle est statutairement langue seconde. La moyenne des items complétés correctement dans les textes est peu élevée, et se situe entre 5% (*Cadeau1*) et 17,2% (*Grand-père1*) pour les données incluant les missings à la fin du livret, respectivement entre 6,7% (*Cadeau1*) et 25% (*Grand-père1*) pour les données sans missings à la fin du livret. Le score maximum obtenu se situe au-dessus de 90% seulement dans *Grand-père1*. Pour les autres textes, le score maximum obtenu varie entre 20% (*Cadeau1*) et 70% (*Basket1*).

3.7. Conclusions

L'objectif du pré-test étant de vérifier la difficulté des textes proposés dans toutes les langues, il est intéressant de confronter les résultats obtenus par chaque item dans toutes les langues et par pays afin de confronter les items les mieux réussis avec ceux qui manifestement obtiennent des résultats très faibles. La prudence reste toujours de rigueur étant donné le nombre de sujets restreint ayant complété certains jeux (voir tableaux 7 à 11).

Aucun texte ne fait l'unanimité au niveau de son degré de difficulté mais certains items comme *Cadeau1* ont des résultats très bas en particulier dans les langues testées comme langue étrangère. En Communauté française de Belgique par exemple, il obtient les scores les plus bas dans les trois langues étrangères testées alors qu'en français (LENS), c'est l'item qui obtient le meilleur résultat (avec une faible corrélation Item-Total de 0,47). En Pologne et en Grèce, il s'agit de l'item présentant les moins bons résultats en LENS et obtenant des résultats assez faibles en langue étrangère. Quant au Luxembourg, le même segment *Cadeau1* représente l'item se situant sur le dernier (en allemand), respectivement l'avant-dernier (en français) rang de difficulté – les deux groupes LENS et LS étant ici regroupés pour chacune des langues.

L'item qui obtient la majorité des meilleurs résultats (7 sur 8 - la Grèce n'étant pas prise en compte) est *Arbre1*. Il fait l'unanimité en Pologne et au Luxembourg et obtient les meilleurs résultats dans trois des quatre langues (français, anglais et allemand) testées en Communauté française de Belgique et le deuxième meilleur résultat en néerlandais. Toutefois, les résultats

obtenus dans les langues testées en tant que langue étrangère restent assez faibles pour tous les items testés en LE (de 4,3% pour *Cadeau1* jusqu'à 37,3% pour *Arbre2* en allemand pour la Communauté française de Belgique, si l'on ne considère pas l'allemand comme LE au Luxembourg) et nous incitent à la prudence quant à la légitimité du niveau de difficulté de tous les items testés en LE : aucun des items testés en LE ne présente donc une moyenne au-delà de 37,3% parmi les groupes d'élèves ayant participé au pré-test, faisant qu'une bonne discrimination entre élèves moyennement performants ainsi qu'entre élèves faibles ne sera donc possible. Les items pré-testés ici ne sont donc pas adéquats pour être utilisés isolément dans un contexte d'évaluation des compétences linguistiques en LE auprès d'élèves se trouvant encore en début du processus d'apprentissage d'une LE, contexte de la présente étude. Les combiner avec d'autres items moins difficiles, permettrait d'obtenir un test adapté à différents niveaux de compétences, ...

Enfin, il reste à souligner que les bonnes moyennes auprès des items testés en LENS semble démontrer que ce genre de test est à la portée des sujets, mais que les résultats peu élevés en LE reflètent plutôt des difficultés dans la langue spécifique testée.

4. Analyse des inter-corrélations entre textes

Le calcul des inter-corrélations entre items¹⁷ d'une même langue dans un même jeu montre leur niveau généralement élevé. Par ailleurs, lorsque les résultats des textes présentent de faibles variances, les corrélations calculées sont également faibles (effet plancher).

Ainsi, les inter-corrélations entre les textes allemands se trouvent toutes à un niveau élevé si l'on considère l'ensemble des deux pays (avec des coefficients de corrélation au-dessus de 0,70). Cependant, elles sont peu élevées pour la Communauté française de Belgique lorsqu'on fait l'analyse par pays, en raison de la variance très faible au niveau des résultats de la Communauté française de Belgique. Notamment le texte *Grand-père2*, qui présentait déjà une corrélation Item-Total négative dans l'échantillon belge, montre, ici aussi, des corrélations négatives, et ce avec *Cadeau2* et *Basket2*. La corrélation avec le troisième texte dans le rang, *Morse2*, est peu élevée.

¹⁷ L'item dans les analyses qui suivent est le texte (*Arbre1*, *Marmottes2*, etc.).

Concernant les textes anglais, les inter-corrélations sont très significatives pour la majorité des textes et relativement élevées à l'exception de l'inter-corrélation *Police1- Graffiti1*, s'élevant à 0,27, ainsi que celle entre *Marmottes2* et *Police2*, s'élevant à 0,389. Il s'agit dans les deux cas d'une corrélation avec un des items présentant une corrélation Item-Total peu élevée (*Police1* 0,62 et *Marmottes2* 0,50). En outre, le texte *Grand-père2* est problématique pour sa version anglaise aussi, étant donné qu'il présente des corrélations négatives, proches de zéro, avec les textes *Basket2* et *Marmottes2*.

Les inter-corrélations de tous les textes en français sont également très significatives bien que peu élevées pour les jeux 5 et 7 et 6 et 8. Ces faibles corrélations se retrouvent au niveau des pays. En Communauté française de Belgique, l'inter-corrélation *Cadeau1 – Basket1* est très faible (0,29) ainsi que celle entre *Arbre1* et *Graffiti1* (0,36). Dans le premier cas, l'item *Cadeau1* présente une corrélation Item-Total peu élevée (0,47) mais ce n'est pas le cas pour les items du deuxième exemple qui présentent tous les deux des corrélations Item-Total satisfaisantes (*Arbre1* : 0,75 et *Graffiti1* : 0,70). Les inter-corrélations aux niveaux des items de ces jeux dans l'échantillon luxembourgeois sont plus élevées que dans l'échantillon belge mais présentent d'autres particularités avec des corrélations non-significatives entre certains items dont *Graffiti1* et *Arbre1*.

Les items en néerlandais présentent des inter-corrélations peu élevées liées de nouveau à la variance très faible au niveau des résultats. Les items *Grand-père1* et *Grand-père2* ont des inter-corrélations non significatives avec tous les autres items de leur jeu.

Enfin, concernant les items en polonais, les items ont tous des inter-corrélations significatives à très significatives excepté pour l'item *Police1* qui présente des corrélations non significatives avec les autres items de son jeu. Ce même item a une corrélation Item-Total peu élevée (0,28).

Les résultats des inter-corrélations entre les textes d'un même jeu par langue, ne sont pas suffisamment concluant que pour en retirer des enseignements significatifs dans une perspective d'amélioration des épreuves quant à leur contenu. Des analyses supplémentaires de nature qualitative seront nécessaires pour voir de plus près quelles pourraient être les raisons du mauvais fonctionnement de textes tels que *Grand-père1* et *Grand-père2*, ou *Police1*.

5. Analyse des effets de séquence

Une des questions posées initialement était celle d'un éventuel effet de séquence qui pourrait se manifester dans les performances, dû à une perte de motivation ou à un effet fatigue. Il se pourrait également que le fait d'avoir un premier segment de texte en LENS suivi d'un autre segment du même texte en LE, et inversement selon les jeux, les aident ou les démotivent. Etant donné qu'un certain nombre d'élèves ou de classes n'ont pu aller jusqu'au bout du test, les items de la fin du jeu qui n'ont pas du tout été entamés n'ont pas été pris en compte.

Dans les jeux allemands et polonais, les résultats ne diffèrent pas significativement en fonction de la position du texte dans le jeu. Les élèves n'ont donc pas de meilleurs résultats lorsqu'ils commencent le test par les textes allemands ou qu'ils terminent par ces mêmes textes.

Les résultats sont plus nuancés pour les autres langues.

En français, dans les jeux 1 et 3, un léger effet de séquence statistiquement significatif se manifeste : pour l'item *Basket2*, ainsi que pour celui de *Morse2*. Ici, les performances sont donc meilleures dans le jeu présentant les textes français au début du livret (jeu 3) que dans celui où les textes sont présentés dans la deuxième moitié du livret (jeu 1). Dans les jeux 2 et 4, la même différence significative est observée pour le l'item *Marmottes2* dont les performances sont meilleures dans le jeu présentant le texte français au début du livret (jeu 4). C'est également le cas pour les items *Basket1* (jeux 5 et 7) et *Graffiti1* (jeux 6 et 8). Ces items sont toujours en deuxième ou en troisième position dans la séquence des quatre items.

L'effet de séquence dans les items anglais se manifeste de façon significative uniquement pour l'item *Police1* dans les jeux 2 et 4, mais comme effet inverse : les résultats sont meilleurs lorsque cet item est présenté après les textes en langue d'enseignement.

Enfin, les analyses pour le néerlandais comparant les résultats en fonction de la position du texte dans le jeu ne révèlent pas d'effet de séquence, à l'exception du texte *Cadeau2* dans les jeux 5 et 7. Les moyennes dans le jeu 7 sont plus élevées que celles dans le jeu 5 alors que le texte vient dans la deuxième moitié du livret. Les élèves (N=16) ont des scores plus élevés que ceux (N=20) qui découvrent le texte au début du livret.

Le nombre important de livrets qui n'ont pas été complétés dans leur intégralité (et qui n'ont pas été pris en compte dans les analyses) influence sans aucun doute les résultats non significatifs concernant les effets des séquences. A l'exception d'un item en néerlandais et d'un item en anglais, les effets significatifs montrent tous la même tendance à traduire un effet de démotivation ou de fatigue.

6. Conclusions

Les analyses psychométriques n'ont pas pu faire ressortir toutes les caractéristiques attendues du fait que le contenu des textes semble effectivement être inadapté au niveau de compétence linguistique en langue étrangère au moins pour les échantillons provenant de la Communauté française de Belgique et de la Pologne. L'échantillon grec est trop restreint que pour pouvoir en tirer des résultats à analyser et à comparer avec les autres échantillons. Le cas du Luxembourg est intéressant à souligner car le résultat des analyses met en évidence de meilleurs résultats au C-test, une meilleure consistance interne et une inter-corrélation entre les textes d'un même jeu par langue plus élevée. De là à en conclure que le C-test serait un outil mieux adapté à l'évaluation de compétences acquises à partir d'un niveau B1 est sans doute trop rapide mais il est de fait interpellant à la lecture des résultats, d'observer qu'entre les élèves belges et luxembourgeois par exemple, les mêmes tests (en langue française et allemande) soient réalisés avec une telle divergence de résultats. Il est évident que seule la qualité psychométrique de l'outil ne peut donner une explication et qu'il faut aller la chercher dans des aspects liés au comportement ; à l'attitude et à la politique d'enseignement des langues mis en place dans les deux zones étudiées. Cette dimension renvoie directement au rapport de la facette 1 du projet (Androulakis et al., 2007).

Chapitre 5 : L'auto-évaluation

Le C-test était précédé d'un questionnaire d'auto-évaluation. L'objectif des chercheurs était notamment de confronter le niveau en langue mesuré grâce au C-test à l'auto-évaluation que pouvaient faire les élèves de leur propre niveau. Après avoir présenté le questionnaire utilisé, nous analyserons les résultats produits grâce à cet outil.

1. Présentation du questionnaire

Le questionnaire d'auto-évaluation est inspiré des portfolios européens des langues (Commission européenne, 2005)¹⁸. Ces modèles, tout en étant calibrés sur le CECR, présentent des descripteurs écrits en langage compréhensible par des adolescents et doivent être validés par le Conseil de l'Europe pour être homologués. Ceci constitue donc une garantie de la validité des descripteurs proposés.

Les items retenus dans le questionnaire concernent la compréhension à la lecture et portent sur les 6 niveaux de l'échelle du CECR (de A1 à C2 – 11 items, 2 par niveau, excepté pour le niveau C2 qui n'est représenté que par un item). Ce choix a été fait en fonction des contraintes temporelles et logistiques : étant donné que a) la facette 1 propose une auto-évaluation en compétences orales, et que b) la « face validity » (la première impression qu'un utilisateur quelconque a de ce que le test mesure) du C-test pointe plutôt vers une évaluation des compétences en compréhension de l'écrit, l'auto-évaluation de la facette 2 porte sur la compréhension écrite (CE).

Le questionnaire d'auto-évaluation des élèves se compose de deux ou trois parties:

- Une colonne relative à la langue d'enseignement en tant que langue étrangère a pour objectif de vérifier si l'élève considère que la langue d'enseignement n'est pas sa langue première. Cette situation pose néanmoins question si l'élève n'atteint pas un niveau B1 ou B2 dans la compétence de lecture car il risque de ne pas bien comprendre les items qui lui sont proposés dans l'auto-évaluation. Dans certains pays,

¹⁸ Les descripteurs d'auto-évaluation en version française, sont extraits du portfolio européen des langues – collège, modèle accrédité n° 44.2003 accordé à ENS/CIEP/DIDIER France. Chaque partenaire s'est référé au(x) portfolio nationaux pour y retrouver des traductions équivalentes.

comme la Pologne par exemple, où il existe très peu d'immigration, elle n'a pas beaucoup de sens. Cette colonne n'a donc été reprise que par la Grèce, le Luxembourg et la Communauté française de Belgique afin de refléter chaque contexte linguistique.

- Une colonne relative à la langue moderne apprise à l'école fait référence à la langue évaluée dans les épreuves du C-test incluses dans le questionnaire.
- Une colonne relative à une autre langue « apprise en plus ». Le cas échéant, il est demandé à l'élève de la définir afin de pouvoir déterminer quelles sont les langues apprises le plus fréquemment (excepté au Luxembourg où elle est prédéfinie).

Contrairement à l'échelle de Likert proposée comme possibilité de réponse aux questions d'auto-évaluation dans la facette 1 de l'étude, le choix de réponses dans le questionnaire d'auto-évaluation de la facette 2 est dichotomique et se limite à *oui* ou *non*, afin de se conformer au CECR et d'éviter toute interprétation dans l'auto-proclamation de compétences acquises (Alderson, 2005, pp.221-212). Les versions proposées dans le cadre du pré-test ont pour objectif de vérifier la pertinence de chacun des items et de sélectionner les langues renseignées par les élèves, ainsi que de s'assurer de la consistance interne de l'auto-évaluation.

Les extraits ci-dessous sont tirés des livrets polonais, luxembourgeois, grecs et belges. Ils montrent les différences de possibilités de réponses liées au contexte linguistique de chaque pays. Les items sont par contre identiques dans toutes les langues. Ils sont présentés dans la langue d'enseignement.



61666

5 0 [] [] [] [] [] []

W pierwszej kolumnie wpisz język, którego się uczysz w szkole i który umiesz najlepiej. Jeśli dodatkowo jeszcze uczysz się jakiegoś języka wpisz go do drugiej kolumny. Jeśli uczysz się dodatkowo więcej niż jednego języka, wpisz ten, który umiesz najlepiej.

Co potrafisz, kiedy czytasz w języku obcym?

Table with 3 columns: Co potrafisz, Język, którego uczę się w szkole, Język, którego uczę się dodatkowo. Rows include recognizing words, understanding simple texts, and identifying political/cultural topics.



51009

4 1 [] [] [] [] [] []

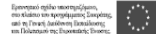
Was du schon kannst, wenn du einen Text liest: [Kreuze bei jeder Aussage für jede Sprache JA oder NEIN an.]

Table with 4 columns: Was ich schon kann, Französisch, Deutsch, Englisch. Rows include understanding simple words, written instructions, and recognizing news articles.



20903

[] [] [] [] [] []



Αυτο-αξιολόγηση - Επίπεδο CITE 2 (3η Γυμνασίου)

[An τα ελληνικά είναι για σένα "ξένη" γλώσσα (αν δηλαδή δεν είναι η πρώτη γλώσσα που έμαθες στο σπίτι σου), συμπλήρωσε την πρώτη στήλη. Αν τα ελληνικά είναι η πρώτη σου γλώσσα, μην ασχολείσαι καθόλου με την πρώτη στήλη. Στη δεύτερη στήλη, σημείωσε μόνο μια γλώσσα, από αυτές που μαθαίνεις στο σχολείο, και την οποία πιστεύεις ότι γνωρίζεις καλύτερα. Αν μαθαίνεις έξω από το σχολείο και κάποια πρόσθετη γλώσσα, εκτός από αυτή της δεύτερης στήλης, σημείωσε την στην τρίτη στήλη. Αν μαθαίνεις πολλές γλώσσες έξω από το σχολείο, σημείωσε εκείνη που πιστεύεις ότι γνωρίζεις καλύτερα.]

Τι μπορείς να κάνεις όταν διαβάζεις ένα κείμενο:

Table with 4 columns: Τι μπορεί να κ άνω, Ελληνικά ως «ξένη» γλώσσα, Ξένη γλώσσα που μαθαίνω στο σχολείο, Άλλη ξένη γλώσσα που μαθαίνω (και εκτός σχολείου). Rows include recognizing words, following instructions, and recognizing news topics.



2337

2 0 [] [] [] [] [] []

[Si le français est pour toi une langue « étrangère » (c'est-à-dire si ce n'est pas la première langue que tu as apprise), remplis la première colonne. Sinon, ne t'en préoccupe pas. Dans la deuxième colonne, indique la langue moderne que tu apprends à l'école et que tu estimes maîtriser le mieux. Si tu apprends une autre langue en dehors de l'horaire scolaire habituel, indique-la dans la troisième colonne. Si tu en apprends plusieurs, indique celle que tu estimes maîtriser le mieux.]

Qu'es-tu capable de faire quand tu lis un texte ?

Table with 4 columns: Ce dont je suis capable, Français en tant que langue « étrangère », Langue moderne apprise à l'école, Autre langue apprise en plus. Rows include recognizing words, understanding simple instructions, and recognizing news topics.

2. Résultats

Pour le présent rapport, nous nous intéresserons uniquement aux résultats relatifs à la langue moderne apprise à l'école et dont fait l'objet le C-test.

2.1. Cohérence des items

En vue d'analyser la consistance interne des items d'auto-évaluation (c'est-à-dire de vérifier si l'ensemble des items mesurent bien la même dimension), une analyse statistique (alpha de Cronbach) a été effectuée sur les résultats des échantillons belges, grecs, luxembourgeois et polonais par langue et par pays. Les alphas sont globalement satisfaisants, témoignant d'une bonne consistance d'ensemble. Par contre, la consistance interne de l'échantillon en langue néerlandaise est problématique (cf. tableau 10).

Tableau 10 : Consistance interne des items d'auto-évaluation (N=11) par langue et par pays

Langues	Alpha de Cronbach par langue	Pays	Alpha de Cronbach par langue et par pays
Allemand	0,88	Luxembourg	0,95
		Communauté française de Belgique	0,76
Anglais	0,80	Communauté française de Belgique	0,77
		Pologne	0,83
Français	0,89	Luxembourg	0,95
		Grèce	0,76
Néerlandais	0,54	Communauté française de Belgique	0,54

2.2. Estimation du niveau de compétence linguistique par les élèves

Dans la première analyse qui est présentée ici, le niveau de chaque candidat, à partir de son autoévaluation, a été calculé de manière simple, en totalisant le nombre de réponses affirmatives (Desroches et al., 2005).

La difficulté des items étant hiérarchisée et présentée dans cet ordre hiérarchique, les réponses positives devraient faire place aux réponses négatives à partir d'un certain niveau correspondant aux limites de compétences de chaque élève. Or, les résultats montrent que le

taux de réponses affirmatives n'est pas décroissant de façon constante. L'exemple de l'auto-évaluation réalisée par les élèves de la Communauté française de Belgique ($N= 228$) en langue anglaise est assez révélateur (cf. tableau 11).

Tableau 11 : Résultats de l'échantillon anglais de la Communauté française de Belgique dans le questionnaire d'auto-évaluation

Qu'es-tu capable de faire quand tu lis un texte ?

Ce dont je suis capable	Langue moderne apprise à l'école	
	OUI	NON
Je peux reconnaître des mots, des expressions et des phrases simples sur une affiche, un journal, un magazine.	<input type="checkbox"/> 91%	<input type="checkbox"/>
Je peux comprendre et suivre des indications écrites simples, par exemple pour aller d'un endroit à un autre.	<input type="checkbox"/> 84%	<input type="checkbox"/>
Je peux reconnaître si un article de presse traite d'un sujet politique, culturel ou économique, d'un fait divers, de la météo...	<input type="checkbox"/> 51%	<input type="checkbox"/>
Je peux trouver un renseignement spécifique dans des prospectus, des menus, des annonces, des horaires, un site Internet...	<input type="checkbox"/> 68%	<input type="checkbox"/>
Je peux localiser une information dans les différentes parties d'un texte long (lettres, prospectus, documents officiels, articles de journaux,...).	<input type="checkbox"/> 57%	<input type="checkbox"/>
Je peux identifier les principales conclusions d'un texte argumentatif bien structuré.	<input type="checkbox"/> 39%	<input type="checkbox"/>
Je peux comprendre des articles spécialisés hors de mon domaine à condition de pouvoir me référer de temps en temps à un dictionnaire.	<input type="checkbox"/> 58%	<input type="checkbox"/>
Je peux comprendre des instructions longues et complexes à conditions de pouvoir relire les passages difficiles.	<input type="checkbox"/> 37%	<input type="checkbox"/>
Je peux comprendre dans le détail beaucoup de textes complexes qui traitent de la vie sociale, professionnelle ou scolaire et identifier des points de détail implicites.	<input type="checkbox"/> 19%	<input type="checkbox"/>
Je peux comprendre presque toute forme d'écrits, y compris des textes abstraits (littéraires ou non) ou spécialisés.	<input type="checkbox"/> 15%	<input type="checkbox"/>
Je peux apprécier de subtiles distinctions de style et le sens implicite autant qu'explicite.	<input type="checkbox"/> 22%	<input type="checkbox"/>

Il apparaît que la gradation des items par ordre croissant de difficulté n'est pas totalement comprise par les élèves. Lors de la passation du pré-test, des observations ont été relevées quant à l'incompréhension de certains items par les élèves: implicite, explicite, abstrait sont

des termes trop compliqués. Dès lors, la révision de certains items en incluant des exemples concrets pourraient les aider à se référer à ce qu'ils font en général en classe.

Une analyse plus détaillée des réponses aux items pour toutes les langues et pays permettrait de dégager ceux à reformuler.

Les résultats de l'auto-évaluation permettent également de comparer l'estimation que les élèves ont de leur niveau en compétence linguistique selon les pays et les langues (tableau 12).

Tableau 12 : Pourcentage de réponses positives à l'auto-évaluation par pays et par langue

<i>Pays</i>	<i>Langues</i>	<i>% de oui</i>
Luxembourg	Allemand (N=49)	89%
	Français (N=66)	89%
Communauté française de Belgique	Allemand (N=80)	52%
	Anglais (N=228)	49%
	Néerlandais (N=222)	38%
Pologne	Anglais (N=115)	38%
Grèce	Français (N=46)	72%

Le nombre de réponses affirmatives est exprimé en % et ne peut être mis en équivalence avec la position occupée par chacun des six niveaux du CECR sur cette échelle (Desrosche, 2005) étant donné que le nombre d'items présentés aux élèves a été réduit pour les raisons évoquées plus haut. En conséquence, le nombre d'items correspondant à chaque niveau est trop réduit. Néanmoins, le niveau attendu des élèves en fin de parcours scolaire au terme de leur dernière année du secondaire étant d'un niveau maximal B1 sur l'échelle du CECR, on peut s'attendre à ce qu'une auto-évaluation reflétant le réel niveau de compétence d'un élève du grade 9 ne dépasse pas 55%, ce qui correspond à des réponses affirmatives jusqu'au 6ème item au maximum. Or les résultats ne reflètent pas entièrement cette estimation.

Le cas du Luxembourg qui pratique un enseignement dans les deux langues, est mis en exergue dans ces résultats. Les élèves estiment avoir un niveau de compétence linguistique assez élevé, équivalent dans les deux langues. En Communauté française de Belgique par contre, le pourcentage de réponses affirmatives est beaucoup moins élevé et inégal en fonction des langues étrangères apprises, l'allemand restant la langue dont les compétences estimées sont les plus élevées (les élèves provenant d'une région frontalière). Les élèves polonais semblent par contre s'auto-évaluer de façon raisonnable. Les élèves grecs ont un pourcentage de oui qui va bien au-delà de leur niveau de compétence exprimé dans les C-tests mais le nombre réduit de sujets ne permet pas d'interpréter ce pourcentage.

2.3. Rapport entre les résultats à l'auto-évaluation et au C-test

L'examen de la corrélation entre les résultats obtenus au C-test et à l'échelle d'auto-évaluation pour chaque élève, quelle que soit la langue considérée et le pays pris en compte, est faible et généralement non significative (comme le montre le tableau 13), ce qui conduit à une absence de lien entre les deux modes d'évaluation.

Tableau 13 : Corrélation entre les pourcentages de oui à l'auto-évaluation et de bonnes réponses au C-test pour l'échantillon

<i>Pays</i>	<i>Langues</i>	Corrélation entre auto-évaluation et résultat au C-test	Valeurs significatives (niveau de signification de 0,05)
Luxembourg	Allemand (N=49)	0,21	0,89
	Français (N=66)	0,20	0,87
Communauté française de Belgique	Allemand (N=57)	0,24	0,77
	Anglais (N= 175)	0,14	0,65
	Néerlandais (N=193)	0,19	0,08
Pologne	Anglais (N=115)	0,28	0,03
Grèce	Français (N=46)	0,17	0,25

Comme ce fut évoqué dans des travaux précédents (Desrosches et al., 2005), ces résultats doivent conduire à réexaminer la manière d'interpréter les échelles d'autoévaluation, en regard de l'évaluation par un test à réponse objective. En effet, si la cohérence de ces deux instruments est aussi faible, et si cela reflète bien une différence réelle entre la manière dont se perçoivent honnêtement les élèves et la manière dont ils sont évalués de manière externe, il convient de s'interroger sur la capacité qu'ont les élèves à s'auto-évaluer et à en comprendre le principe.

3. Conclusions

La mise au point d'un outil d'auto-évaluation calibré sur le référentiel européen à partir des portfolios des langues destinés à la population ciblée par l'étude semblait a priori pouvoir répondre aux objectifs des chercheurs. Les résultats conduisent au constat que cet outil n'est pas maîtrisé par les élèves qui ont dans leur grande majorité surestimé leur niveau par rapport

aux résultats de l'épreuve objective qui leur était proposée. Il faut néanmoins nuancer au vu de ces résultats qui ne reflètent peut-être pas le niveau réel de compétence linguistique de ces élèves confrontés à un type de test inconnu de la plupart et dont les textes sont apparemment inadaptés à leur niveau. Rappelons que l'auto-évaluation était préalable au C-test et que les élèves ne pouvaient donc nullement être influencés par leur performance au C-test.

Un autre questionnement est celui de l'efficacité du questionnaire d'auto-évaluation dans la forme réduite telle que proposée (11 items). Un nombre plus important d'assertions quant aux capacités à réaliser certaines tâches permettrait sans doute aux élèves de mieux cerner leur niveau de compétence.

Conclusions et perspectives

Les analyses des données recueillies lors du pré-test ont porté sur la **fidélité** – au sens de la capacité à mesurer un construit de manière cohérente et exacte - **des items** (segments de textes constituant le C-test) pour les pays et les jeux pour lesquels suffisamment de données étaient disponibles, sur les **inter-corrélations entre items**, sur les **différences dans les performances** (pour l'ensemble des pays ayant testé une même langue, mais aussi par pays) en fonction :

- de la langue testée (LENS, LS/LE),
- de la position du texte dans le jeu,
- des pays ayant testé la langue.

A part quelques exceptions, elles ont révélé une cohérence interne satisfaisante jusqu'à élevée au niveau des items sélectionnés pour le pré-test pour toutes les langues, confirmant ainsi les résultats des analyses d'autres études portant sur le C-test. Même pour le polonais, pour lequel on ne connaît pas encore d'étude C-test, il est intéressant de voir que la fidélité des tests créés est assez satisfaisante. Les difficultés quant à elles diffèrent en grande partie significativement en fonction du pays et du statut de la langue testée (LENS, LS/LE). Des analyses qualitatives, par exemple au niveau des structures morpho-syntaxiques, seront nécessaires afin d'identifier la nature des difficultés.

En outre, jusqu'à présent, on ne peut être certain que les textes C-test créés dans les différentes langues évaluent de manière équivalente le même construit. En effet, un premier constat est lié à la problématique linguistique et la comparaison de langues, en particulier, à la structure syntaxique et à la structure des mots. La création d'items équivalents en différentes langues présente des caractéristiques spécifiques liées à la particularité du C-test qui consiste en la mutilation de mots. Différentes unités sémantiques, grammaticales et morphologiques peuvent être concernées en fonction des langues, donnant pour résultat un texte à difficulté variable après mutilation. Des études de validité ainsi que des analyses qualitatives seraient nécessaires afin de pouvoir établir des points de comparaisons.

Même si l'on met donc de côté la problématique de l'équivalence des questions ouvertes dans différentes langues posée lors de l'élaboration d'enquêtes internationales telles que PISA,

d'autres problèmes surgissent au niveau de la comparabilité entre items dont l'équivalence dans différentes langues ne semble pouvoir être garantie a priori.

Un autre constat qu'on peut faire sur la base des expériences faites dans la présente étude est que, dans le cadre d'évaluations internationales (souvent d'une ampleur assez importante), l'économie de temps par l'utilisation du C-test n'est pas nécessairement assurée en termes de temps de correction, surtout si l'on décide de prendre en considération les variantes et les fautes d'orthographe. Cet élément peut être résolu d'une part, en se limitant éventuellement à l'acceptation de solutions sans examiner les variantes orthographiques mais aussi, en organisant la passation par ordinateur plutôt qu'à l'aide de feuilles à lecture optique, mais cela engendre d'autres contraintes logistiques (présence d'ordinateurs en nombre suffisant).

Par ailleurs, la plupart des textes C-test utilisés en langue seconde ou langue étrangère sont d'un niveau de difficulté très élevé pour la population pré-testée. Ce type de textes ne suffit pas à discriminer une population à large échelle se composant aussi bien d'élèves à des niveaux de compétences linguistiques faibles, que d'élèves se trouvant déjà à un stade de compétences élevé. Des textes C-test supplémentaires, de niveau de difficulté moyen et faible, devraient donc être intégrés afin de compléter le test dans une perspective de mise au point de l'outil. A cette fin, il serait utile de se référer aux facteurs qui déterminent la difficulté de textes C-test. Bien qu'il y ait déjà un nombre d'études à ce sujet, un commun accord semble encore loin d'être trouvé (voir, par exemple, Babaii & Moghaddam, 2006; Hastings, 2002 ; Klein-Braley, 1985 ; Sigott, 2002). Toutefois, les données issues du pré-test fournissent suffisamment d'informations permettant d'analyser de plus près les facteurs engendrant ces difficultés élevées, par exemple à partir de la nature et de la spécificité de certains mots. Un autre niveau d'analyse peut être envisagé sur la population cible. Les niveaux de compétences dans une langue seconde ou étrangère à un grade donné, peuvent varier fortement selon les systèmes éducatifs et les priorités en matière d'enseignement des langues, d'un pays à un autre. Dans le cadre d'une étude internationale, ciblant les élèves d'un même âge ou d'un même grade, on s'expose donc à une grande hétérogénéité concernant les niveaux de compétences visées. Cette composante importante doit être prise en compte dans les procédures de création du test qui devra être adapté au type de population ciblée (élèves du même âge ou grade, ou ayant plus ou moins le même niveau de compétence).

Enfin, le constat sans doute le plus sensible est celui de la reconnaissance du C-test ou sa « validité apparente ». En effet, enseignants et élèves éprouvent des difficultés à estimer la pertinence de ce type de test pour mesurer le niveau des compétences linguistiques générales d'une langue étrangère en cours d'apprentissage de manière comparable aux tests de langues « classiques ». Ce problème semble d'autant plus délicat que le niveau de compétences des élèves est faible.

Le défi, identifier une procédure de test compatible avec une évaluation internationale à large échelle, est donc loin d'être relevé, mais les données rassemblées dans le cadre de cette recherche, leurs analyses et les perspectives qui en découleront, contribueront certainement à améliorer l'outil afin de pouvoir l'envisager comme dispositif d'évaluation.

La simplicité apparente du principe du C-test ne dispense pas de prendre en compte les spécificités des langues du test. En effet, comme cela a bien été montré, certaines langues, non encore examinées à l'aide du C-test, comme le polonais et le grec, ne présentent pas les mêmes caractéristiques morphologiques et syntaxiques, ni les mêmes graphèmes, et ces particularités influent certainement sur la difficulté à compléter les lacunes. Cela place donc les évaluateurs face au problème de l'estimation d'un niveau de maîtrise de deux langues différentes.

Par contre, il semble bien possible d'utiliser le C-test pour estimer les compétences de sujets différents, issus de contextes semblables ou non, dans la même langue. Ce qui ouvre des perspectives non négligeables quant à la possibilité d'évaluer le niveau d'une langue étrangère choisie à travers plusieurs populations au sein de la Communauté européenne afin d'estimer, par exemple, le niveau de compétence linguistique terminal à l'issue du parcours scolaire obligatoire.

Mais afin d'arriver à transformer les résultats obtenus en un niveau correspondant à un des six niveaux de l'échelle du référentiel européen, le CECR, le travail d'étalonnage du C-test par rapport à d'autres épreuves serait particulièrement bien venu. Il existe encore très peu de travaux allant dans ce sens. Un travail a notamment été entrepris par l'Université du Luxembourg et du Ministère de l'Education nationale luxembourgeois (Reichert, 2006).

Une autre source d'embarras réside en l'absence de corrélation entre les auto-évaluations et les résultats aux C-tests, mais, nous l'avons vu, les échelles d'auto-évaluation proposées à partir du CECR ne sont sans doute pas encore totalement satisfaisantes. Même des outils

adaptés aux populations scolaires, tels les portfolios, ne parviennent pas à exprimer de manière suffisamment claire, des descripteurs qui permettent une bonne compréhension des tâches attendues en fonction des niveaux de compétence sur l'échelle du CECR.

Si la nécessité de développer pendant la scolarité obligatoire un profil de compétence élevé dans au moins deux langues étrangères est une réalité bien établie, faut-il encore s'en donner les moyens. Pour pouvoir se situer et progresser, tout apprentissage passe par l'évaluation et cette dimension doit donc être prise en compte dans sa complexité. Il nous semble important de mener de front des explorations pour améliorer les conditions d'apprentissage et motiver les jeunes Européens à maîtriser plusieurs langues, comme le développent largement les conclusions et perspectives du rapport de la facette 1 (Androulakis et al., 2007), et des travaux pour mettre au point des outils d'évaluation qui puissent répondre aux exigences de la politique européenne tout en respectant la rigueur scientifique afin de s'assurer de leur qualité objective en matière de mesure des compétences linguistiques.

Références

- Androulakis, G., Beckmann, C., Blondin, C., Fagnant, A., Goffin, C., Górecka, J., Martin, R., Mattar, C., Meißner, F.-J., Mosorka, E., Nowicka, A., Pelt, V., Schröder-Sura, A., Skrivánek, H., Wilczyńska, W., & Wojciechowska, B. (2007). *Pour le multilinguisme : Exploiter à l'école la diversité des contextes européens. Résultats d'une étude internationale*. Liège : Éditions de l'Université de Liège.
- Arras, U., Eckes, T. & Grotjahn, R. (2002). C-tests im Rahmen des "Test Deutsch als Fremdsprache" (TestDaF): Erste Forschungsergebnisse. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, pp.175-209). Bochum: AKS-Verlag.
- Aylett, M.P. (2000). *Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech*. In Proceedings of ICSLP-2000, Beijing.
- Babaii, E., & Moghaddam, M. J. (2006). On the interplay between test task difficulty and macro-level processing in the c-test. *System*, 34, 586-600.
- Bisping, M. & Raatz, U. (2002). *Sind computerisierte und Papier&Bleistift-Versionen des C-Tests äquivalent?* In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, pp.131-155). Bochum: AKS-Verlag.
- Blondin, C. (2005). *Enquêtes internationales et apprentissage des langues : un défi à relever*. Communication lors du Symposium « Langues modernes et évaluation ». Actes du 18e colloque international de l'Admée-Europe : Comment évaluer ? Outils, dispositifs, acteurs (Reims, du 24 au 26 octobre 2005).
- COMMISSION EUROPÉENNE (2004). *New Indicators on Education and Training*. Commission staff working paper, consulté sur l'Internet (01-02-2007) http://europa.eu.int/comm/education/policies/2010/doc/indicators_en.pdf .
- COMMISSION EUROPÉENNE (2005). *Progress towards the Lisbon objectives in education and training*, Commission staff working paper, consulté sur l'Internet (01-02-2007) <http://ec.europa.eu/education/policies/2010/doc/progressreport05.pdf>
- CONSEIL DE L'EUROPE (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Strasbourg : Division des Langues Vivantes.
- Demeuse M., Desrosches F., Crendal A., Renaud F., Casanova D., Artus F. (2005) : *La fiabilité de l'évaluation des compétences linguistiques pour des adultes non francophones :*

- présentation d'un protocole d'évaluation*, In Actes du 18e colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe), Reims, 24-26 octobre 2005.
- Desrosches, F., Casanova, D., Crendal, A., Renaud, F., Mahy, C., Demeuse, M. (2005). *Mise en équivalence des versions papier-crayon et électronique d'un test de français langue étrangère et mise en perspective des résultats par une procédure d'autoévaluation*, In Actes du 18e colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe), Reims, 24-26 octobre 2005
- Eckes, T. & Grotjahn, R. (2006). *C-Tests als Anker für TestDaF. Rasch-Analysen mit dem kontinuierlichen Ratingskalen-Modell*. In R. Grotjahn (Ed.), *Der C-Test. Theorie, Empirie, Anwendungen / The C-Test: Theory, Empirical Research, Applications*. Frankfurt/M.: Lang.
- Grotjahn, R. (2002). *Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis*. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4, pp. 211-225). Bochum: AKS-Verlag.
- Grotjahn, R. (2002b). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4). Bochum: AKS-Verlag.
- Hamez, M.P. (2006). *Le FLS en classe d'accueil, en France, Quelques repères*, APLV consulté sur l'internet (15-04-2007) http://www.aplv-languesmodernes.org/article.php3?id_article=398
- Hastings, A.J. (2002). *In defense of C-testing*. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4, pp.11-25). Bochum: AKS-Verlag.
- Jafarpur A. (1999). *Can the C-test be improved with classical item analysis?* *System*, 27, 79-89.
- Jafarpur, A. (2002). *A comparative study of a C-Test and a cloze test*. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, pp.31-51). Bochum: AKS-Verlag.
- Klein-Braley, C. (1985). *A cloze-up on the C-Test: a study in the construct validation of authentic tests*. *Language Testing*, 2, 76-104.
- INRA (2001), *Les Européens et les langues*, Eurobaromètre 54 spécial. Consulté sur l'Internet (01-02-2007) http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_147_fr.pdf .
- Monseur C. & Demeuse M. (2004). *Politiques d'éducatons et de formation, Quelques réflexions méthodologiques à propos des enquêtes internationales dans le domaine de*

- l'éducation*, (pp.37-47). Bruxelles : Institut Européen d'Education et de Politique Sociale & Institut européen pour la Promotion de l'Innovation de la Culture dans l'Education.
- OECD (2002). *Programme for International Student Assessment. Sample tasks from the PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: OECD.
- Pinker, S. (2000). *The Language Instinct: How the Mind Creates Language (1st edition)*. New York: Harper Perennial Modern Classics.
- Raatz, U. & Klein-Braley, C. (1982). *The C-Test - a modification of the cloze procedure*. In T. Culhane, C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and Problems in Language Testing* (pp.113-145). University of Essex : University of Essex Occasional Papers.
- Rapport d'activité de la deuxième étape du Projet 42/2006 Action 6.1.2. – 6.2 Socrates, Août 2006. Université de Liège.
- Reichert, M., Keller, U. & Martin, R. (2006). *Le Test de Connaissance du Français et le C-Test. Etude sur la comparabilité entre les deux instruments*. (Rapport établi pour le Ministère de l'Education nationale et de la Formation Professionnelle.) Luxembourg : Université du Luxembourg, FLSHASE, Groupe de Recherche EMACS.
- Sigott, G. (2002). High-level processes in C-Test taking? In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4, pp.67-82). Bochum: AKS-Verlag.
- Wockenfuß, V. & Raatz, U. (2006). *Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests*. In R. Grotjahn (Ed.), *Der C-Test. Theorie, Empirie, Anwendungen / The C-Test: Theory, Empirical Research, Applications* (pp. 211-242). Frankfurt/M.: Lang.

Annexes

** Nombre d'observations par jeu, par langue et par pays

1. Jeux par pays et par langues :

a. Communauté française de Belgique :

- i. français - néerlandais
- ii. français – anglais
- iii. français - allemand

b. Luxembourg

- i. allemand - français
- ii. français – allemand

c. Grèce

- i. grec – français

d. Pologne

- i. polonais – anglais

2. Instructions aux professeurs

3. Feuille de commentaires lors de la passation des pré-tests

4. Note technique concernant les anomalies des données de la Communauté française de Belgique