



Commission
européenne



Etude sur la comparabilité des tests linguistiques en Europe

Résumé

COMMISSION EUROPEENNE

Direction générale Education et Culture

Direction A — Modernisation de l'enseignement I: Europe 2020, Analyse par pays, Coordination d'Erasmus+

Unité A.4 — Etudes, études d'impact, analyses et statistiques

Contact: eac-unite-a4@ec.europa.eu

Commission européenne
B-1049 Bruxelles

Etude sur la comparabilité des tests linguistiques en Europe

Résumé
Septembre 2015

Ce document a été préparé pour la Commission européenne; néanmoins, il reflète uniquement l'avis des auteurs et la Commission ne saurait être tenue responsable de toute utilisation qui serait faite des informations qu'il contient.

***Europe Direct est un service destiné à vous aider à trouver des réponses
aux questions que vous vous posez sur l'Union européenne.***

Un numéro unique gratuit (*):

00 800 6 7 8 9 10 11

(* Les informations sont fournies à titre gracieux et les appels sont généralement gratuits (sauf certains opérateurs, hôtels ou cabines téléphoniques).

De nombreuses autres informations sur l'Union européenne sont disponibles sur l'internet via le serveur Europa (<http://europa.eu>).

Luxembourg: Office des publications de l'Union européenne, 2015

ISBN 978-92-79-50998-8
doi: 10.2766/20788

© Union européenne, 2015

Reproduction autorisée, moyennant mention de la source

Image de couverture: © Shutterstock.com

Printed in Belgium

1 Résumé

1.1 Objectif et portée de l'étude

Suite à l'adoption des "Conclusions sur le Multilinguisme et le Développement des Compétences Linguistiques" par le Conseil de l'Union Européenne en mai 2014, une nouvelle approche pour mesurer les compétences linguistiques au niveau européen a été suggérée. Plutôt que de développer un système de références linguistiques commun à tous les États Membres, il a été convenu qu'il serait préférable de mettre en œuvre des mesures visant à promouvoir le multilinguisme et améliorer la qualité et l'efficacité de l'apprentissage et l'enseignement des langues, et à développer des dispositifs permettant d'évaluer les compétences linguistiques, de préférence au sein du système éducatif de chaque pays.

Afin d'obtenir une meilleure connaissance des compétences linguistiques en Europe, le Conseil a invité la Commission Européenne à explorer la possibilité d'évaluer les dites compétences dans les différents États Membres, en s'appuyant sur les tests linguistiques nationaux existants. Le but de cette étude est d'évaluer la comparabilité des tests linguistiques nationaux existants pour les élèves des niveaux CITE 2 et CITE 3 (enseignement secondaire premier et deuxième cycle). L'étude utilise dans ce but des données rassemblées par le Réseau Eurydice relatives aux tests linguistiques nationaux existants dans les États Membres de l'Union Européenne.

1.2 Langues et examens pris en compte dans l'étude

Les langues incluses dans l'étude sont les langues qui ne sont pas les principales langues d'instruction. Seules les langues officielles de l'UE qui sont utilisées dans au moins un autre État Membre de l'Union Européenne ont été prises en compte dans cette étude. Pour chaque territoire considéré, seules les langues étudiées par plus de 10% des élèves de l'enseignement secondaire (selon Eurostat: 2013, 2014) ont fait l'objet d'une analyse (voire Table 1 du rapport, section 2.2.1).

Sur la base des données rassemblées par le Réseau Eurydice, 133 examens nationaux de langues provenant de 33 territoires juridiques (territoires avec indépendance dans le domaine de l'éducation ressortissants des 28 États Membres de l'UE) ont été identifiés comme pertinents pour cette étude de comparabilité. De ces 133 examens, 77 sont au niveau CITE 2 et 56 au niveau CITE 3. L'Annexe 1 offre une liste détaillée des examens nationaux inclus dans cette étude, ainsi que les raisons pour lesquelles certains examens ont dû en être exclus.

1.3 Participation des États Membres

Afin de garantir que les résultats de cette étude soient les plus fiables et transparents possibles, la Commission Européenne a facilité la collaboration entre les membres du Groupe Expert pour les Indicateurs sur le Multilinguisme (Indicator Expert Group on Multilingualism) (IEG). Ces membres sont tous des experts en enseignement des langues et/ou travaillent sur les questions d'évaluation linguistique pour les Ministères de l'Éducation Nationale ou Offices des Statistiques Nationales, sur leur territoire respectif.

Après une réunion initiale avec la Commission Européenne et l'IEG, l'Équipe du Projet s'est mise directement en contact avec chaque membre du groupe pour discuter en détail les examens nationaux de langues existants dans chaque territoire juridique. La contribution de ces membres experts a été essentielle pour confirmer les langues et les examens inclus pour chaque territoire, ainsi que d'autres informations concernant ces examens, telles que des exemples de leurs examens ; des exemples de performance de leurs étudiants ; et la documentation d'appui concernant leurs examens (procédures pour la création et l'administration des examens ; les matériaux pour la formation des créateurs des tâches et des évaluateurs ; les résultats nationaux, etc.).

1.4 Structure de l'étude

Les 5 principales tâches examinées dans ce rapport sont :

- 1^{ère} tâche: Évaluation de la comparabilité des tests linguistiques nationaux existants pour les étudiants des écoles secondaires.
- 2^{ème} tâche: Propositions pour un ajustement *a posteriori* qui pourrait faire augmenter la comparabilité des résultats existants.
- 3^{ème} tâche: Propositions d'évolution en vue d'augmenter la comparabilité des tests linguistiques existants.
- 4^{ème} tâche: Propositions pour les États Membres qui n'ont pas de système de tests linguistiques et qui seraient intéressés pour en développer un
- 5^{ème} tâche: Synthèse comparative des données existantes selon les pays au sujet des tests linguistiques.

Le Cadre Européen Commun de Référence (CECR) a été utilisé comme cadre comparatif dans cette étude. Le CECR est très largement utilisé partout en Europe et sert de point de référence, de schéma pertinent pour l'apprentissage et de construction pour un système de mesure.

1.5 Résultats

La 1^{ère} tâche a été accomplie grâce à une méthode de recherche mixte comprenant l'analyse des données qualitatives aussi que quantitatives. Celle-ci est décrite en détail dans la section 5. Les données qualitatives ont été rassemblées par une analyse experte du contenu des examens de langues et a été accomplie par une équipe de spécialistes de haut niveau en évaluation linguistique provenant de différents pays Européens. Ces experts ont utilisé un instrument en ligne pour l'analyse du contenu des examens et ont été spécifiquement formés dans l'utilisation de cet instrument pour assurer l'uniformité et la fiabilité de leur travail. Les données quantitatives ont été rassemblées par un exercice de jugement comparatif qui a été accompli par 49 experts en éducation et en évaluation linguistique sur une plateforme en ligne désignée spécialement pour ce but (www.nomoremarking.com).

L'analyse qualitative du contenu et des caractéristiques des examens a été faite sur les 133 examens de langues (provenant de 33 territoires juridiques, 28 États Membres de l'UE). Cette évaluation a trouvé qu'il y avait une diversité considérable entre les tests, diminuant le potentiel pour une comparabilité directe des résultats des

tests. Quatre principaux domaines ont été analysés: les constructions de capacité (ce qui est mesuré par le test), les interprétations données aux résultats des tests, les populations testées, et les caractéristiques du système de mesure (particularités contextuelles qui peuvent affecter la comparabilité.) Sur la base d'un certain nombre d'éléments, les indicateurs suggèrent un manque de comparabilité.

En ce qui concerne les **constructions de capacité**, les examens de langue observés montrent une diversité importante, même si les parties composantes utilisent souvent les mêmes termes (par exemple, « compréhension écrite »). Par conséquent, il serait probablement incorrect de comparer les résultats des différents tests et en conclure qu'ils sont interchangeables. En réalité, les tests examinés semblent avoir pour but d'évaluer des compétences différentes même s'ils utilisent les mêmes termes pour faire référence aux différentes parties de ces examens.

Ainsi, à partir des **interprétations des résultats**, la principale conclusion concerne les tests qui ne garantissent pas de suivre le CECR. Dans de nombreux de cas, il n'était pas possible d'établir comment les résultats de ces tests devaient être interprétés. A défaut, certains résultats ont été normalisés selon les références (pour être interprétés en comparant le classement d'un candidat par rapport aux autres). Une telle approche n'est pas propice à une véritable comparaison des résultats entre différents tests, car les populations dans chaque cas sont différentes.

Les **populations** ayant passé les tests de langues aux niveaux CITE 2 et CITE 3 ont présenté une homogénéité satisfaisante en termes d'âge, la seule caractéristique de la population ayant été examinée.

En termes de **caractéristiques du système de mesure**, comme dans le cas des constructions, beaucoup de résultats suggèrent une comparabilité réduite. Par exemple, une proportion significative des tests n'étaient pas garantie d'être équivalents entre les sessions. Dans ce cas, la comparabilité de ces tests avec d'autres tests était impossible étant donné que les résultats d'une session ne pouvaient même pas être comparés à ceux d'une autre session pour la même épreuve. Bien que la comparabilité des résultats entre les sessions soit souhaitable pour un grand nombre de raisons, et devrait être ainsi traitée, les tests étaient différents pour des raisons légitimes. Par exemple, le type de questions choisi a un effet sur les résultats des tests qui est lié à la nature même de la construction mentale évaluée, et certains types de questions peuvent avoir un nombre d'effets particuliers, comme l'augmentation ou diminution des distinctions entre les candidats.

Une approche quantitative a aussi été présentée pour comparer les résultats existants grâce à un **exercice de jugement comparatif**. Celle-ci a été illustrée par un nombre limité de tâches de compréhension et expression écrite qui ont été extraites des examens de langue inclus dans cette étude. Cette méthode montre la possibilité d'aligner au CECR les résultats nationaux de tous les territoires en fonction de la difficulté des tâches dans leurs différents examens linguistiques nationaux. Cet exercice a aussi démontré des disparités dans la difficulté relative des tâches des examens nationaux, mais la portée limitée de cette étude n'a pas permis de fournir une comparaison complète des résultats des tests individuels faute de données plus précises à propos des distributions des points pour chaque examen. Compte tenu du manque de comparabilité actuel entre les résultats nationaux des examens

linguistiques, cette méthode suggère une nouvelle approche qui permettrait de comparer ces résultats nationaux dans l'avenir, notamment en appliquant la technique du jugement comparatif aux exemples de performance des étudiants dans les tâches de compréhension et de production écrite.

1.6 Propositions d'évolution

Prenant en compte les résultats de la 1^{ère} tâche, un nombre de propositions ont été présentées pour accomplir les 2^{ème}, 3^{ème} et 4^{ème} tâches. Les propositions suivantes sont expliquées en détail dans les sections 6, 7 et 8 du rapport.

1.6.1 Propositions pour un ajustement a posteriori visant à augmenter la comparabilité des résultats nationaux existants

Cette étude suggère l'utilisation de la technique de jugements comparatifs comme la méthode la plus utile pour l'ajustement a posteriori des résultats nationaux existants. Cette méthode a pour but de développer une échelle commune des compétences linguistiques sur laquelle les examens et résultats nationaux pourraient être placés. Cependant, afin de mettre cette méthodologie en place, certaines conditions doivent être remplies :

- Un format commun pour la communication des résultats nationaux ;
- Un engagement de la part des différents territoires pour fournir les données nécessaires ;
- Un plan annuel établi et surveillé par un organisme responsable.

1.6.2 Propositions pour le développement des tests linguistiques existants dans le but d'augmenter la comparabilité des résultats

La comparabilité des résultats des tests dépend de la qualité et de la diversité des tests. Ces différences, légitimes, proviennent du contexte et de la finalité des examens. La qualité du test affecte la comparabilité parce qu'un système d'évaluation plus faible ou moins fiable conduit à des résultats aléatoires. Les conclusions de cette étude montrent qu'il existe un certain nombre de problèmes de qualité affectant les tests qui devraient être traités par les concepteurs de ces tests. Nous devons néanmoins reconnaître que la possibilité d'améliorer la qualité de certains tests est parfois limitée en fonction des moyens financiers disponibles. Le manque de comparabilité du fait de différences légitimes entre les tests est plus difficile à réduire, et une comparabilité transversale entre les territoires devrait être la finalité. Les principales recommandations sont donc :

Construction

- élargir l'éventail des types de lecture et de compréhension orale qui font l'objet de tests au niveau B2 et au-dessus.
- concevoir des tâches qui suscitent des processus cognitifs appropriés pour chaque niveau de compétences selon le CECR.

Interprétations

- développer des interprétations de résultats sur la base de critères pouvant être détaillés au CECR si l'harmonisation avec le CECR n'a pas été sollicitée.

Population

- rassembler des informations concernant les caractéristiques de la population testée.

Caractéristiques du système d'évaluation

- s'assurer que le recrutement de toute l'équipe (concepteurs des tests, auteurs des questions, rédacteurs, évaluateurs, analystes, etc.) est basé sur une liste de compétences requises pour ce travail.
- garantir une formation pour les membres de l'équipe qui n'ont pas toutes les compétences requises.
- garantir que le jugement des évaluateurs est standardisé de façon à ce que les analyses soient cohérentes.
- garantir que les procédures d'évaluation impliquent une surveillance et des actions correctives dans l'éventualité où la surveillance révélerait des problèmes.
- développer des procédures afin de corriger les différences (particulièrement concernant le niveau de difficulté) entre les plusieurs versions de la même épreuve.
- établir un programme transversal dont le but est d'aligner le test avec le CECR.
- rassembler périodiquement les données sur les notes et les réponses, et les analyser afin de permettre une amélioration des procédures de développement et de gestion.
- améliorer la rédaction des questions et la correction de celles-ci.
- vérifier les caractéristiques du test et déterminer s'ils peuvent être rendus plus comparables avec les tests des autres territoires.
- envisager l'utilisation d'un test unique pour comparer les compétences des candidats avec ceux des autres territoires.

1.6.3 Propositions pour mettre en place de nouveaux examens linguistiques nationaux

Il existe beaucoup d'ouvrages contenant des recommandations théoriques et pratiques pour la création et l'administration d'examens linguistiques, et ceux-ci sont référenciés dans la section 8. Au-delà de leurs indications générales, ce rapport suggère quelques recommandations plus spécifiques qui ont le potentiel d'augmenter dans l'avenir la comparabilité des résultats d'examens nationaux.

- Intégrer le CECR dans le test: la conception de tests fondée sur le CECR sera plus facile si le CECR est utilisé dès le départ.
- Développer des procédures visant à une amélioration constante du test: la mise à disposition du test doit être vue comme un cycle au sein duquel l'information est en permanence rassemblée dans une tentative de détecter les problèmes et les résoudre pour les futurs tests.
- Développer une procédure visant à maintenir les niveaux: déterminer où sont les limites entre les niveaux du CECR, et ensuite s'assurer que ces limites sont respectées dans toutes les sessions, si possible utilisant une banque d'items.

1.7 Synthèse comparative des données existantes dans les pays sur les tests linguistiques

La 5^{ème} tâche consistait d'une synthèse des données existantes actuellement dans chaque territoire concernant les résultats des examens de langues. Des 133 examens des langues inclus dans cette étude, nous avons essayé de trouver des données pour 62 tests de premières langues étrangères d'un total de 33 territoires, mais il n'a été possible de trouver des données que pour 45 de ces tests provenant de 26 territoires. Les raisons par lesquelles ces résultats ne sont pas disponibles sont décrites dans la section 8.2.

Les données disponibles pour chaque territoire varient énormément, ainsi que le format dans lequel ces données sont fournies. La section 9.2 offre une synthèse des observations faites à propos du format dans lequel les résultats nationaux des tests de langues sont présentés.

Dans le but de produire pour l'avenir un tableau européen avec une synthèse des résultats nationaux ajustés de façon qu'il puisse être utilisé pour contrôler régulièrement les compétences linguistiques des étudiants dans une ou plusieurs langues, il faut considérer d'abord très attentivement un nombre d'éléments pour assurer que ce tableau soit compilé et interprété de la manière la plus significative et représentative possible. Ces éléments sont expliqués en détail dans la section 9.3, et comprennent la sélection des données qui doivent être fournies, la signification de « réussite » à un test, et les populations qui passent les examens.

1.8 Conclusion

La mesure dans laquelle des résultats des examens linguistiques nationaux peuvent être comparés dépend sur un nombre important de facteurs. Premièrement, une comparaison des résultats nationaux n'est possible que lorsque les données comparées ont assez d'éléments en commun. Actuellement, il existe trop de variabilité dans la nature des données fournies par les autorités éducatives des différents territoires ainsi que le format dans lequel ces données sont présentées. De plus, cette étude a montré que les examens linguistiques dans les différents territoires présentent une très grande variabilité dans les constructions testées, les populations des étudiants qui passent les examens, les interprétations des résultats et les caractéristiques des systèmes de mesure sur lesquels ces examens sont basés. Ces caractéristiques déterminent en grande partie la qualité des tests, et à leur tour impactent sur la validité et la fiabilité des résultats obtenus. Par conséquent, **une comparaison sérieuse des résultats nationaux des examens de langues des États Membres de l'UE ne dépendra pas seulement du format uniforme dans lequel ils pourraient être exprimés, mais aussi de la mise en pratique des mesures à niveau national et européen visées à une augmentation de la qualité des examens linguistiques actuels et, par la suite, à assurer que les résultats soient similairement valables et fiables dans tous les territoires.**

COMMENT VOUS PROCURER LES PUBLICATIONS DE L'UNION EUROPÉENNE?

Publications gratuites:

- un seul exemplaire: sur le site EU Bookshop (<http://bookshop.europa.eu>);
- exemplaires multiples/posters/cartes:
auprès des représentations de l'Union européenne (http://ec.europa.eu/represent_fr.htm),
des délégations dans les pays hors UE (http://eeas.europa.eu/delegations/index_fr.htm),
en contactant le réseau Europe Direct (http://europa.eu/europedirect/index_fr.htm)
ou le numéro 00 800 6 7 8 9 10 11 (gratuit dans toute l'UE) (*).
(* Les informations sont fournies à titre gracieux et les appels sont généralement gratuits (sauf certains opérateurs, hôtels ou cabines téléphoniques).

Publications payantes:

- sur le site EU Bookshop (<http://bookshop.europa.eu>).



Publications Office

ISBN: 978-92-79-50998-8