

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 19, Nummer 2 (Oktober 2014)

Évaluation de la production écrite dans le cadre de l'Enquête européenne sur les compétences linguistiques (ESLC)

Sylvie Lepage

Centre international d'études pédagogiques
Département évaluation et certifications
1, avenue Léon Journault
92318 – Sèvres
Tél : 01 45 07 63 85
Courriel : lepage@ciep.fr

Roselyne Marty

Centre international d'études pédagogiques
Département évaluation et certifications
1, avenue Léon Journault
92318 – Sèvres
Tél : 01 45 07 63 92
Courriel : marty@ciep.fr

Dr. Michaela Perlmann-Balme

Goethe-Institut e.V.
Bereich 41 Sprachkurse und Prüfungen
Dachauer Straße 122
80637 München
Telefon: +49 89 15921-382
E-Mail: perlmann-balme@goethe.de

Claudia Stelter

Goethe-Institut e.V.
Bereich 41 Sprachkurse und Prüfungen
Dachauer Straße 122
80637 München
Telefon: +49 89 15921-122
E-Mail: stelter@goethe.de

Abstract: Der vorliegende Aufsatz ist besonders für Fremdsprachenlehrkräfte in den Fächern Französisch und Deutsch von Interesse, aber auch Lehrende anderer Fremdsprachen können davon profitieren. Er bietet eine Einführung in die Methodik, die für die Bewertung schriftlicher Leistungen junger Lernender entwickelt wurde. Diese hatten an der *Europäischen Erhebung zur Fremdsprachenkompetenz* teilgenommen. Sie basiert auf dem Vergleich schriftlicher Leistungsbeispiele mit Modellarbeiten, welche jeweils ein mittleres Leistungsniveau der Stufen A2 und B1 des *Gemeinsamen europäischen Referenzrahmens* abbilden. Die angewendete Methode sowie die Materialien sollen zur Standardisierung von Bewertungen beitragen und können auch bei der Aus- und Fortbildung von Bewertenden verwendet werden. Im Beitrag wird eine sprachübergreifende Bewertungsübung

anhand von französischen und deutschen Leistungsbeispielen illustriert. Die Frage der Vergleichbarkeit von sprachübergreifenden Bewertungsverfahren wird diskutiert.

This article will be of interest to foreign language teachers mainly of French and German but also of other foreign languages. It will introduce the methodology developed specifically for the assessment of written productions from young learners who took part in the *European Survey on Language Competences*. It is based on the comparison of written performances with model scripts which illustrate medium performances for the A2 or B1 levels of the Common European Framework of Reference. It will introduce the methodology and show materials used for rater training. It will then demonstrate how samples in French and German were marked by trained raters. The article will also discuss the issue of comparability across languages.

Cet article est principalement destiné aux professeurs qui enseignent l'allemand et le français mais peut aussi intéresser ceux qui enseignent d'autres langues étrangères. Son objectif est de présenter la méthodologie spécifiquement élaborée pour évaluer la production écrite de jeunes apprenants dans le cadre de *l'Enquête européenne sur les compétences linguistiques*. Cette méthodologie est basée sur la comparaison de productions écrites avec des copies-étalon sélectionnées en tant que performances moyennes pour les niveaux A2 et B1. L'article présentera également la démarche utilisée pour la formation des correcteurs ainsi que le matériel de formation, notamment des exemples de production écrite en allemand et en français évalués et commentés. La question de la comparabilité de l'évaluation de la production écrite en différentes langues sera également soulevée.

Key words : European survey, assessment of writing, methodology, training of raters

Mots-clés : Enquête européenne, évaluation de la production écrite, méthodologie, formation des correcteurs

1. Introduction

L'objectif de cet article est de faire part de la méthodologie employée pour l'évaluation de la production écrite dans le cadre de *l'Enquête européenne sur les compétences linguistiques (ESLC)*, de fournir des exemples de productions en allemand et en français aux niveaux A2 et B1 (exemples ayant servi à la formation des correcteurs) et d'expliquer la démarche adoptée lors de cette formation à l'évaluation. La question de la comparabilité de l'évaluation de la production écrite en différentes langues sera également soulevée. Mais avant d'en venir à la production écrite, il est nécessaire de rappeler l'origine et les objectifs de l'enquête ESLC.

1.1. Origine et objectifs de l'enquête

En mars 2002, le Conseil européen réuni à Barcelone a demandé la mise au point d'un indicateur des compétences linguistiques concernant les élèves de dernière année du secondaire, avec l'objectif de permettre aux gouvernements européens de comparer le niveau de leurs élèves en langues étrangères et de s'inspirer des meilleures pratiques pour le développement de leurs politiques linguistiques.

La Commission européenne a confié l'organisation de cette enquête au consortium *SurveyLang*, constitué d'organismes spécialistes de l'évaluation linguistique (*Goethe-Institut, Instituto Cervantes, University of Cambridge ESOL Examinations, Universidad de Salamanca, Università per Stranieri di Perugia, Centre international d'études pédagogiques*), ainsi que de spécialistes de l'échantillonnage (*Gallup*) et de psychométrie (*Institut national d'évaluation en éducation - CITO*).

L'enquête ESLC, initiée par la Commission européenne en février 2008 et publiée en juin 2012, a permis d'obtenir des informations concrètes sur le niveau de langue des élèves, en référence aux niveaux du *Cadre européen commun de référence pour les langues* (CECR, Conseil de l'Europe 2001).

1.2. Organisation et mise en place

L'enquête ESLC a été menée dans 14 États membres/16 systèmes éducatifs : Angleterre, les trois communautés linguistiques de Belgique, Bulgarie, Croatie, Espagne, Estonie, France, Grèce, Malte, Pays-Bas, Pologne, Portugal, Slovénie, Suède (cf. *First European Survey on Language Competences – Technical Report 2012*).

Dans chaque pays, les deux langues étrangères les plus enseignées (parmi l'allemand, l'anglais, l'espagnol, le français et l'italien) ont été testées en compréhension orale et écrite et en production écrite, du niveau A1 au niveau B2

Sylvie Lepage, Roselyne Marty, Michaela Perlmann-Balme & Claudia Stelter (2014), Évaluation de la production écrite dans le cadre de l'Enquête européenne sur les compétences linguistiques (ESLC). *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 58-77. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Lepage_et_al.pdf.

du CECR. L'évaluation a porté sur un échantillon représentatif d'apprenants de 13 à 16 ans (au moins 1 500 élèves par langue et par pays), l'apprentissage d'une langue étrangère débutant plus ou moins tôt selon les pays. Les élèves ont été recrutés entre la dernière année du premier cycle du secondaire et la deuxième année du deuxième cycle du secondaire (ISCED 2 ou ISCED 3, *International Standard Classification of Education*). Au total, 53 000 élèves des pays cités ci-dessus ont été évalués. Par ailleurs, des questionnaires ont été transmis aux ministères de l'éducation, aux chefs d'établissements, aux enseignants et aux élèves. Ces questionnaires avaient pour objectif de faciliter la comparaison entre les États membres en matière de politiques linguistiques et de méthodes pédagogiques, en cherchant à identifier les bonnes pratiques.

2. Méthodologie de l'enquête pour l'évaluation de la production écrite

L'un des défis majeurs a été d'assurer la comparabilité des tests dans les cinq langues. Les cinq institutions partenaires ont travaillé étroitement sur la définition des spécifications des tests et sur la conception des tâches. De plus, les tâches de production écrite ont été adaptées d'une langue à l'autre et ont été communes aux cinq langues. Il s'agissait de la première expérience de cette envergure en Europe. Outre la comparabilité des épreuves de production écrite, il fallait aussi assurer la comparabilité de l'évaluation grâce à une méthodologie commune. Un séminaire de définition des scores de césure (*standard setting*) a été la dernière étape. Ces différents points seront exposés aux chapitres 4 et 5.

2.1. La conception des tâches d'évaluation de production écrite

Le construit des tests de langue a été basé sur les niveaux et les activités langagières du CECR. L'approche communicative et actionnelle telle qu'elle est décrite dans le CECR était l'approche partagée par tous les partenaires linguistiques. Cela signifie que les tâches ont été conçues en privilégiant des situations authentiques, c'est-à-dire des situations susceptibles d'être rencontrées par un adolescent de 13 à 16 ans apprenant une langue étrangère. Les 3 domaines suivants ont été utilisés dans l'élaboration des tâches : le domaine personnel (ma chambre, par exemple), le domaine public (par exemple : utiliser les transports en commun) et le domaine éducationnel (voyages scolaires, par exemple).

Les spécifications des tâches ont été déterminées après une étude comparative des tâches d'évaluation des différents organismes de certification du consortium *SurveyLang* et avec l'aide de la grille pour la production écrite (*ALTE CEFR Writing Grid*) élaborée par les membres d'ALTE (*Association of Language Testers in Europe*), disponible sur le site du Conseil de l'Europe.

Suite au prétest et à l'expérimentation sur le terrain, les spécifications retenues ont été les suivantes (First European Survey on Language Competences, Technical Report 2012 : 34) :

Tableau 1: Spécifications des tâches de production écrite

Task	Test focus	Text type	Task type	Levels
W1	Expressing general or topic-specific notions describing pictures or graphically-displayed information	Short personal text (email)	Candidates write a short personal text making reference to the picture / graphically-displayed information.	A1
W2	Expressing general or topic-specific notions in response to input text and content points	Short personal text (email, postcard)	Candidates write a short personal text explaining, describing etc.	A1 A2 B1

W5	Writing a referential text (intended to inform)	Personal text (email)	Candidates write a personal text explaining, describing etc.	A2 B1
		At B2 an article, essay, letter, report, review	At B2 candidates write article etc. explaining, describing, comparing etc.	B2
W6	Writing a conative text (intended to persuade or convince)	An essay, letter	Candidates write an essay / letter describing, explaining, comparing, justifying, giving opinion etc.	B2

2.2. L'évaluation de la production écrite

Une nouvelle méthodologie d'évaluation de la production écrite a été mise au point par Neil Jones, directeur du consortium *SurveyLang* et par les représentants des cinq institutions chargées de développer le test en cinq langues. Le sentiment général était que l'interprétation des niveaux du CECR pouvait varier en fonction des pays et qu'il serait difficile d'arriver à un consensus sur la manière de noter sans exemples précis auxquels se référer. D'autre part, comme l'évaluation était assurée par les correcteurs de chaque pays impliqués dans l'ESLC, il fallait s'assurer que ces correcteurs appliquent les critères de la même manière.

Plutôt que de demander aux correcteurs d'évaluer en fonction de leur connaissance des niveaux du CECR, il leur a été demandé d'évaluer par rapport à des copies-étalon. Pour les niveaux A, les copies-étalon étaient des copies représentatives d'une performance moyenne par rapport à une tâche donnée, c'est-à-dire que si toutes les productions pour cette même tâche avaient été classées de la moins bonne à la meilleure, la copie-étalon aurait occupé une place centrale. La copie-étalon représente une performance moyenne pour une tâche spécifique pour une population donnée. Pour les niveaux B, la première copie-étalon était représentative d'un petit niveau de performance et la deuxième copie d'une bonne performance, toujours par rapport à la tâche donnée. La raison pour laquelle deux copies-étalon ont été sélectionnées pour les niveaux B est que l'écart pour passer d'un niveau à l'autre devient plus important au fur et à mesure qu'on avance dans l'apprentissage. Il peut donc y avoir beaucoup de variation dans la représentation des niveaux B en partant de la production écrite la plus faible jusqu'à la production écrite la plus forte alors que pour les niveaux A, la variation est moins grande. Chaque niveau représente un continuum plutôt qu'un point précis sur une échelle. Le continuum étant plus grand pour les niveaux B, il était nécessaire de sélectionner deux copies-étalon représentatives d'un petit niveau et d'un bon niveau.

Ces copies-étalon ont été sélectionnées en fonction des scores obtenus lors de l'expérimentation sur le terrain et à l'aide des analyses de CITO, l'Institut national d'évaluation en éducation des Pays-Bas. Lors de cette phase d'expérimentation, les niveaux A avaient été évalués avec 2 critères : réalisation de la tâche et correction linguistique/organisation, et les niveaux B avec 4 critères : réalisation de la tâche, compétence communicative, compétence discursive, correction linguistique. Pour les niveaux B, les critères étaient subdivisés ainsi :

- ✓ **A. Réalisation de la tâche**
 - ✓ Combien de points sont réalisés ? (clarté)
 - ✓ De quelle manière les points sont développés ?
- ✓ **B. Compétence communicative**
 - ✓ Lexique
 - ✓ Aspect sociolinguistique
 - ✓ Effet sur le lecteur
- ✓ **C. Compétence discursive**
 - ✓ Degré d'élaboration des phrases
 - ✓ Cohésion
 - ✓ Cohérence
- ✓ **D. Correction linguistique**
 - ✓ Orthographe
 - ✓ Grammaire



Schéma 1 : Critères d'évaluation, niveau B, première version

Comme il s'était avéré difficile lors de l'expérimentation sur le terrain de noter les niveaux B avec autant de critères et de subdivisions, il a été décidé de ne retenir que deux critères pour l'enquête ESLC : communication et langue. La communication portait dorénavant sur le contenu et la correction sociolinguistique. Le correcteur devait se poser les questions suivantes : dans quelle mesure la tâche est-elle accomplie ? Autrement dit : combien de points de contenu sont traités ? Sont-ils traités de manière claire ? De manière développée ? Le style et le registre sont-ils appropriés ?

Le deuxième critère, plus linguistique, permettait d'évaluer le lexique, la grammaire, l'organisation du texte et l'orthographe.

Tableau 2 : Critères d'évaluation, niveau B, deuxième version

Communication (Réalisation de la tâche) (contenu, correction sociolinguistique)	Langue (organisation, orthographe, grammaire)
Dans quelle mesure la tâche est-elle accomplie ?	La correction lexicale/ linguistique et l'organisation sont-elles adéquates ?
Combien de points de contenu sont traités de manière claire ? Est-ce que ces points sont bien développés ? Style - registre	Aspects linguistiques à prendre en compte : cohérence lexique cohésion étendue

Pour les niveaux A et pour chaque critère, trois notes étaient possibles pour évaluer les productions des élèves testés (schéma 2).

- 1 : moins bien que la copie-étalon
- 2 : à peu près aussi bien que la copie-étalon
- 3 : meilleur que la copie-étalon

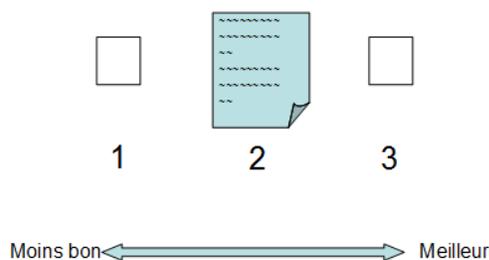


Schéma 2 : Echelle à 3 notes niveaux A1/A2 (évaluer en comparant avec une copie-étalon)

Aux niveaux B1 et B2, les notes ont été attribuées en classant les productions écrites observées par rapport à **deux copies-étalon**. Pour chaque critère, cinq notes étaient possibles (schéma 3)

- 1 : pas aussi bien que la copie-étalon de niveau inférieur
- 2 : à peu près aussi bien que la copie-étalon de niveau inférieur
- 3 : meilleur que la copie-étalon de niveau inférieur mais pas aussi bon que la copie étalon de niveau supérieur
- 4 : à peu près aussi bien que la copie-étalon de niveau supérieur
- 5 : meilleur que la copie-étalon de niveau supérieur

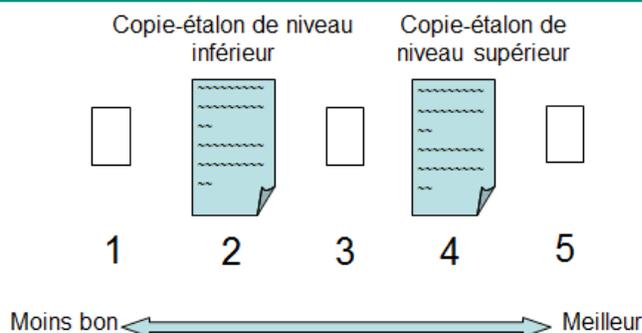


Schéma 3 : Échelle à 5 notes niveaux B1/B2 (évaluer en comparant avec deux copies-étalon)

2.3. La formation à l'évaluation de la production écrite

Pour garantir la fiabilité de l'évaluation, les cinq partenaires ont élaboré en commun leur matériel de formation. Ils se sont d'abord assurés de partager la même interprétation des critères en évaluant des productions rédigées en an-

glais, langue commune à tous. L'important était de privilégier l'approche actionnelle du CECR et de ne pas se focaliser sur les erreurs si celles-ci ne bloquaient pas la communication. Par exemple, en allemand, les erreurs d'orthographe telles que « Son » à la place de « Sonne », les erreurs de genre et de déclinaison n'ont pas été pénalisées car elles n'entravent pas la communication. Le matériel de formation, minutieusement élaboré, devait permettre de garantir une très grande comparabilité dans l'évaluation faite en différentes langues durant la formation des correcteurs, étape-clé de l'enquête ESLC.

La correction de la production écrite a été à la fois centralisée et décentralisée, c'est-à-dire qu'une partie des productions écrites ont été évaluées à la fois dans les pays des candidats et par les organismes certificateurs de SurveyLang. Pour diminuer les écarts de notation, deux séminaires d'évaluation ont été organisés au Centre international d'études pédagogiques (CIEP) en 2010 et 2011 pour former les coordinateurs nationaux à la méthodologie décrite ci-dessus, ceux-ci devant à leur tour former leurs équipes. Les différences d'interprétation que les correcteurs pouvaient avoir des niveaux du CECR ont pu être réduites grâce à la comparaison des écrits des élèves avec des copies-étalon.

Pour ces séminaires, les participants disposaient d'un guide du correcteur expliquant la méthodologie, du livret des tâches avec des copies-étalon commentées pour chaque niveau et de feuilles de notation. De leur côté, les formateurs disposaient du livret du formateur qui indiquait précisément les différentes étapes, le minutage et le matériel utilisé : un Powerpoint de présentation, un fichier électronique Excel pour la saisie des notes et la projection sur écran, et des jeux de copies pour le classement des productions écrites et des notes de référence et commentaires sur ces productions. Les différentes étapes de la formation seront présentées ci-dessous.

2.3.1. Première étape de la formation : exercice de familiarisation

L'objectif de cet exercice était de familiariser les participants avec la méthode de classement des productions écrites les unes par rapport aux autres.

Les participants répartis en petits groupes ont reçu 4 jeux de productions écrites, chacun comportant 4 copies d'élèves pour une même tâche (4 tâches x 4 productions écrites). Ils ont dû classer ces productions écrites de la meilleure à la plus faible en prenant en compte les critères de la « communication » et de la « langue ». Chaque production écrite apparaissait sur une page séparée de telle sorte que les participants puissent les classer sur la table. Cette phase de classement a duré 15 minutes.

Des exemples de productions en allemand et en français, sélectionnés pour le classement, peuvent être consultés en annexe.

2.3.2. Deuxième étape de la formation : familiarisation avec la méthodologie

Cette étape a permis d'introduire la méthodologie proposée par *SurveyLang*. Les participants ont pris connaissance d'une tâche de niveau A1 projetée à l'écran. Ils ont été invités à répondre aux questions suivantes :

- Combien de points de contenu la tâche inclut-elle ?
- Que pourrait écrire un candidat de niveau A1 pour répondre à cette tâche ?

Puis un exemple de production écrite moyenne d'un candidat a été projeté à l'écran. Cet exemple avait été sélectionné parmi les productions de l'expérimentation sur le terrain, après avoir croisé les résultats de la correction centralisée et décentralisée. Les participants ont dû dire si cet exemple répondait à leurs attentes pour les deux critères d'évaluation :

- Communication (nombre de points abordés ? L'information est-elle suffisante ?)
- Langue (le vocabulaire est-il approprié à la tâche ? La structure de la phrase simple est-elle correcte ? Pourrait-il y avoir de meilleures réponses que celle-ci ? (Oui, par exemple avec plus de vocabulaire) Pourrait-il

y avoir des réponses plus faibles que celle-ci ? (Oui, par exemple, des réponses moins claires avec des erreurs qui entravent la compréhension.)

Lors de la mise en commun les formateurs ont guidé la discussion de telle sorte que tous les participants concluent que la production était d'un niveau moyen.

La même procédure a été adoptée avec deux productions écrites de niveau B1, l'une illustrant un petit niveau de performance et l'autre un bon niveau de performance, de telle sorte à avoir 2 points d'ancrage sur une échelle de 1 à 5. Autrement dit, les productions à évaluer pouvaient ensuite être classées par rapport à la copie-étalon de niveau inférieur et à la copie-étalon de niveau supérieur pour le critère de la communication et celui de la langue.

2.3.3. Troisième étape de la formation : phase d'entraînement

Dans un premier temps, les participants ont évalué, en petits groupes, 4 productions écrites correspondant à chacun des niveaux A1, A2, B1 et B2 par comparaison avec les copies-étalon correspondantes. La discussion a été menée de telle sorte à obtenir les notes de référence pour chacune des productions écrites.

Dans un deuxième temps, les participants ont évalué individuellement 12 copies et inscrit leurs notes sur une feuille prévue à cet effet. Lors de la pause qui a suivi, les notes ont été reportées sur un fichier électronique avec un numéro d'identifiant correspondant à chaque correcteur, mais permettant de préserver l'anonymat. Les notes ont ensuite été montrées à l'écran pour que tous puissent les comparer et les commenter.

Ce fichier donnait des indications sur :

- la constance : dans quelle mesure les correcteurs sont capables de différencier les meilleures productions écrites des moins bonnes.
- le niveau des notes : dans quelle mesure les notes des correcteurs sont proches des notes de référence (plus élevées ou plus basses).

Les formateurs ont de nouveau guidé la discussion afin d'obtenir un consensus sur les notes de référence pour les critères de la communication et de la langue pour chacune des 12 productions écrites.

Correcteur		1	2	3	4	5	6	Note de référence		
Session n°1	H° d'identification du correcteur									
	Constance	-	-	-	-	-	-			
	Niveau des notes	-	-	-	-	-	-			
Session n°2	Constance	-	-	-	-	-	-			
	Niveau des notes	-	-	-	-	-	-			
Session n°1	Echantillon	Note de référence								
	1	Com							3	Com
		Lang							3	Lang
	2	Com							3	Com
		Lang							3	Lang
	3	Com							1	Com
		Lang							1	Lang
	4	Com							4	Com
		Lang							5	Lang
	5	Com							1	Com
		Lang							1	Lang
	6	Com							2	Com
		Lang							3	Lang
	7	Com							2	Com
		Lang							3	Lang
	8	Com							1	Com
Lang								1	Lang	

Schéma 4: Feuille de notes pour la standardisation de l'évaluation de la production écrite

2.3.4. Dernière étape de la formation : phase de correction

Pendant cette phase, les participants ont évalué 16 productions de manière individuelle. Ils se sont familiarisés avec les copies-étalon correspondantes pendant 15 minutes, puis ont eu 50 minutes pour évaluer les productions de manière individuelle et reporter leurs notes sur les feuilles prévues à cet effet. Leurs notes ont ensuite été reportées ensuite sur le fichier électronique donné en illustration (schéma 4).

La mise en commun a de nouveau permis que chacun comprenne les notes de référence pour la communication et la langue pour chacune des 16 productions. Toutes ces productions avaient déjà été évaluées auparavant lors de l'expérimentation sur le terrain dans le cadre de la correction décentralisée et centralisée, et sélectionnées après les analyses montrant un haut degré de corrélation dans leur évaluation. Les scores attribués à ces productions et les commentaires justifiant ces scores ont donc servi de référence pour les correcteurs. Grâce au fichier Excel qui produisait immédiatement un rapport sur chaque participant, chaque correcteur a pu voir si ses notes se rapprochaient des notes de référence et s'il était constant dans sa notation.

Cette session s'est terminée sur une note positive en faisant repérer aux participants les stratégies qui, selon eux, étaient les plus efficaces pour arriver à des jugements standardisés.

3. Exemple de tâches et de productions écrites en allemand et en français

On trouvera ci-dessous des exemples de tâches A2 et B1 et des productions commentées qui ont servi lors de la formation des correcteurs.

Tableau 3: Tâche de niveau A2 (First European Survey on Language Competences - Final report 2012:109-110).

FR - Nouveau passe-temps préféré	DE - Neues Hobby
<p>Tu as commencé une nouvelle activité. Tu écris un email à un ami français et tu lui dis :</p> <ul style="list-style-type: none"> • quelle est ta nouvelle activité • quand tu as commencé cette activité • pourquoi tu aimes cette activité <p>Tu écris 25–35 mots.</p>	<p>Du hast ein neues Hobby. Schreib einer deutschen Freundin eine E-Mail. Schreib:</p> <ul style="list-style-type: none"> • Was ist dein neues Hobby? • Wann hast du damit angefangen? • Was gefällt dir an dem Hobby? <p>Schreib 25–35 Wörter.</p>

Tâche A2 : Nouveau passe-temps préféré (Production commentée)

<p>Chèr Peter Bonjour Peter ! Ca va ? J'ai commencé une nouvelle activité il y a quelques semaines il y a très magnificint. J'ai joué à foot et rugby. Il es très difficult et entertaining. Salut xxx</p>

Communication	2	Langue	2
---------------	---	--------	---

Pour rappel, l'échelle de notes va de 1 à 3 (cf.chapitre 2.2)

Communication

2 points sont traités. Pour le point 3, il faut interpréter. Rituel de la lettre respecté.

Langue

Les points positifs :

- Bonne construction du passé-composé : « j'ai commencé », « j'ai joué »
- Expression du temps « il y a quelques semaines »
- Tentative pour construire une phrase complexe : « J'ai commencé une nouvelle activité il y a quelques semaines ».

Les points négatifs :

Des erreurs qui peuvent entraver la compréhension : interférences avec l'anglais pour le lexique (« magnificent », « difficult », « entertaining »).

Tableau 4 : Tâche de niveau B1

FR - Membre de la famille	DE - Familienmitglied
<p>Voici un extrait d'un message que tu as reçu de ta correspondante française.</p> <p><i>Dans ton prochain mail, parle-moi d'un membre de ta famille que tu aimes vraiment beaucoup. Qu'est-ce que vous faites ensemble ? Pourquoi est-ce que vous vous entendez bien tous les deux ?</i></p> <p>Tu écris un email à ta correspondante française et tu réponds à ses questions.</p> <p>Tu écris 80–100 mots.</p>	<p>Von einem deutschen Brieffreund bekommst du eine E-Mail. Darin schreibt er:</p> <p><i>... Bitte schreibe mir in deiner nächsten E-Mail, wen du in deiner Familie besonders gern magst. Was macht ihr gemeinsam? Warum versteht ihr euch gut? ...</i></p> <p>Schreib eine E-Mail an deinen Freund und antworte auf seine Fragen.</p> <p>Schreib 80–100 Wörter</p>

Tâche B1 : Membre de la famille (Production commentée)

<p>Cher</p> <p>Comment ça va ?</p> <p>J'aime mon père. Il s'appelle Cork et il a 50 ans. Souvent je regarde a television avec mon père. Nous regardons des films ou des sport. Il regarde aussi quand je joue du foot. Quand j'ai fini nous mangons quelque chose. Nous nous entendons bien parce quel nous adorons le meme chose. Mon père fait beaucoup pour moi. Quand je veux visiter une université, il aime aller avec moi. Et quand j'ai un problem il m'aide.</p> <p>C'est parce que j'aime mon père.</p>

Communication	4	Langue	4
---------------	---	--------	---

Pour rappel, l'échelle de notes va de 1 à 5 (cf. chapitre 2.2)

Communication

Les 2 points sont abordés et le sujet est développé. Il se termine par une conclusion. Ton amical, mais pas de prénom dans la formule d'appel et pas de formule de congé.

Langue

Les points positifs :

- Plusieurs connecteurs ou adverbes : souvent, quand (3 fois), aussi, parce que
- Bon emploi du présent et du passé composé. Bonne conjugaison (à part « mangons »)
- Vocabulaire simple mais adéquat

- Bon emploi du pronom indirect (« nous nous entendons », « il m'aide »)
- Emploi de l'infinifit pour 2 verbes qui se suivent.

Aucune erreur qui entrave la compréhension. Mais moins de prise de risques au niveau du lexique que dans la copie-étalon 4.

Autres erreurs : « a » à la place de « la » devant « télévision », oubli de l'accord à « sport »

Exemples de productions en allemand commentées

Tâche A2 : Neues Hobby

Hey Peter, wie geht's?
 Hör mal, ich habe ein neues Hobby – Postkarten zu sammeln! Ich beschäftige mit es schön für zwei Wochen. Ich mag es sehr.
 Aufwiedersehen,
 Linze

Aufgabenbewältigung/ Kommunikation	2	Sprachliche Ausführung	2
---------------------------------------	---	---------------------------	---

Aufgabenbewältigung/ Kommunikation

Inhaltspunkte 1 und 2 klar, Inhaltspunkt 3 (Was gefällt dir an dem Hobby) nicht behandelt.

Der Textsorte teilweise angemessene formale Gestaltung (*Hey Peter, wie geht's?* angemessen; *Hör mal, Aufwiedersehen* nicht akzeptabel)

Sprachliche Ausführung

Beispiele für gelungene Ausdrucksweise:

Grammatische Korrektheit: *Ich mag es sehr.*

Wortschatz: *Postkarten sammeln*

Fehlgriffe, die möglicherweise die Kommunikation behindern:

Grammatische Korrektheit: *für zwei Wochen; Ich beschäftige mit es.*

Rechtschreibung: *schön statt schon*

Weitere Fehler, die das Verständnis nicht beeinträchtigen: *Postkarten, Aufwiedersehen*

Tâche B1 : Familienmitglied

Hallo, Jakob
 Ich erzähle dir über meine Familie. Sie ist nicht besonders groß. Meine Familie besteht aus drei Personen. Das sind ich, mein Vater und meine Mutter. Wir wohnen im Zweizimmerwohnung. Ich mag beide Eltern. Wir leben uns vertige. Sie sind teuer für mich. Wir verbringen sehr viel Freizeit zusammen. Z.B. wir gehen ins Kino. Sehr oft fahren wir in den Wald und machen Picknick. Am Wochenende fahren wir ins Dorf. Dort wohnt meine Oma. Das ist auch eine wichtige Person in meiner Familie. Sie ist ältere und kluge, als andere Person. Sie gibt uns immer gute

Ratschläge. Ich liebe sehr meine Familie.
Aufwiedersehen

Aufgabenbewältigung/ Kommunikation	4	Sprachliche Ausführung	4
---------------------------------------	---	---------------------------	---

Aufgabenbewältigung/Kommunikation

Inhaltspunkte 1 und 2 klar behandelt. Inhaltspunkt 3 (*Warum versteht ihr euch gut?*) nicht klar
Aufwiedersehen als Grußformel nicht akzeptabel

Sprachliche Ausführung

Beispiele für gelungene Ausdrucksweise:

Grammatische Korrektheit – Wortstellung: *Sehr oft fahren wir, Am Wochenende fahren wir*

Wortschatz: *besteht aus, Zweizimmerwohnung, Picknick machen, Wir verbringen sehr viel Freizeit zusammen.*

Kohärenz: Variation der Satzstrukturen, wenngleich hauptsächlich kurze Sätze: *Ich ... Meine Familie ... Das sind ... Wir ... Sehr oft ...*

Fehlgriffe, die möglicherweise die Kommunikation behindern:

Grammatische Korrektheit: - *Komparativ: Sie ist ältere und kluge als andere Persone.*

Wortschatz: *Wir leben uns vertige, teuer für mich*

4. La définition des scores de césure par standard setting

Dans son article « A comparative approach to constructing a multi-lingual proficiency framework », Neil Jones (2009) fait une analogie avec le thermomètre. La première étape pour mesurer la température a été l'invention d'un instrument de mesure : le thermomètre. La deuxième étape a été l'adoption d'une mesure commune : le degré Celsius. C'est seulement lorsqu'une échelle commune de mesure est adoptée par tous qu'on peut comprendre ce que signifient différents degrés de fièvre. De la même façon, le CECR est un instrument de référence pour toutes les langues européennes, et la définition des niveaux doit se faire à partir d'une échelle de mesure commune. Il ne suffit pas pour un organisme de certification de définir des scores de césure entre les différents niveaux pour que ceux-ci soient fiables. Encore faut-il que le niveau B1 en espagnol soit comparable au niveau B1 en anglais, allemand, français, etc.

Pour permettre une évaluation fiable des niveaux commune aux cinq langues, une première étude a été menée avec 125 correcteurs. Ceux-ci devaient classer des exemples de productions écrites en deux langues, c'est-à-dire, décider quelles étaient les meilleures et quelles étaient les moins bonnes. Cette étude a permis de comparer l'interprétation des niveaux à travers les différentes langues.

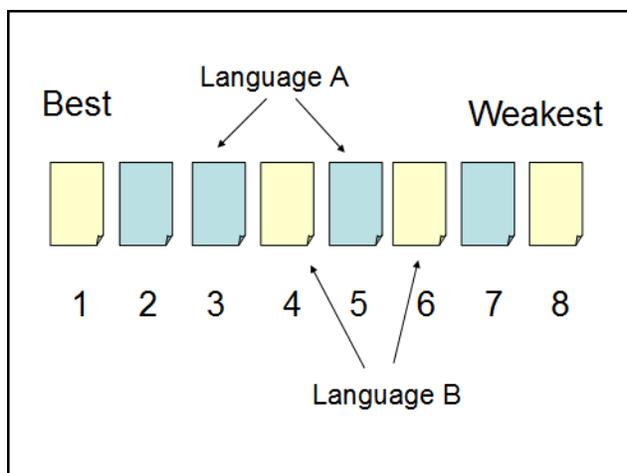


Schéma 5 : Classement des productions en deux langues de la meilleure à la plus faible pour une tâche donnée.

Les données de cette étude ont été analysées selon le modèle de Rasch¹ afin de classer toutes les productions sur une seule échelle de compétence, quelle que soit la langue. Cette étude a aussi été corrélée avec les résultats du séminaire de définition des scores de césure. En septembre 2011, ce séminaire a réuni 70 experts au total pour les cinq langues de l'enquête ESLC. L'objectif de ce séminaire était de déterminer quelles productions étaient représentatives de chaque niveau pour chaque langue. Lors des deux premières phases, les participants ont dû classer, pour chaque tâche, une série de productions écrites et déterminer si chacune d'entre elles atteignait ou non le niveau requis par la tâche. Pour la troisième phase, les participants ont évalué l'ensemble des productions écrites de 30 élèves et attribué un niveau du CECR à chacun. Les évaluations de cette troisième phase, croisées avec celles des deux premières et l'étude de classement préalable, ont permis de recouper tous les jugements qui avaient été portés sur les niveaux.

Réconcilier les données des différentes évaluations faites dans chaque groupe de langue n'a pas été une tâche facile concernant la production écrite. Les analyses ont cependant pu fournir des exemples de production écrites, illustrant les niveaux en cinq langues pour les mêmes tâches. Ceux-ci sont disponibles dans le rapport final (2012: 108-116) et reproduits en annexe pour les niveaux A2 et B1.

5. Comparabilité de l'évaluation des langues

L'évaluation des langues étrangères obéit à des règles propres à chaque langue et à chaque système éducatif. Dans cette enquête européenne, l'objectif a été de surmonter ces différences et ce cloisonnement entre les différentes pratiques. Les élèves en fin de scolarité obligatoire devaient être testés de la même manière, quel que soit leur pays et quelles que soient leurs langues étrangères de manière à établir des comparaisons précises et justes entre les résultats.

Par ailleurs, l'équipe *SurveyLang* chargée de l'élaboration du test en 5 langues a décidé de ne pas traduire les tâches d'une langue à l'autre mais plutôt de les adapter de façon à préserver les aspects culturels propres à chaque langue. Ainsi les tests en allemand, anglais, espagnol, français et italien ont été conçus de manière à ce qu'ils soient aussi semblables que possible, tant par les procédures utilisées que par le contenu de chaque test, qu'il s'agisse du type de tâche ou du niveau de langue. L'évaluation de la production écrite a également été standardisée (cf. chapitre 2.3.4.).

Les performances recueillies pour chaque tâche d'une langue à l'autre se sont révélées assez proches, comme le démontre cette comparaison directe des copies répondant à la tâche A1 « Carte postale de vacances » :

Sylvie Lepage, Roselyne Marty, Michaela Perlmann-Balme & Claudia Stelter (2014), Évaluation de la production écrite dans le cadre de l'Enquête européenne sur les compétences linguistiques (ESLC). *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 58-77. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Lepage_et_al.pdf.

Tableau 5: Exemples de productions écrites de la tâche A1

<p>Dear Simone Hallo Simone, I have a good time. I stay in a fantastik hotel, the weather is sunny. They people in the photo they are my friends, by Giovanni</p>	<p>J'ai été dans l'hôtel Bellevue. L'hôtel est sur la plage. Je joue au volleyball. J'ai suis très heureuse dans ce hôtel.</p>	<p>Lieber Peter Du siehst Ich bin mit meiner Familie in Belleve Hotel in Antalya. Hier ist es gut. Die Speisen sind frisch und heiß. Hotel ist am strand. Nach dem Mittagessen spielen Volleyball Turnior. Bis später Kübra</p>
---	--	---

Le lexique et les structures grammaticales de la langue n'ont pas joué un rôle aussi important qu'attendu, sans doute principalement parce que les types de tâches se concentrant sur la connaissance de mots ou de structures spécifiques ont été évités.

Les exigences concernant l'apprentissage de la grammaire en français et en allemand augmentent très rapidement et de ce fait, la correction grammaticale est traditionnellement mise en avant dans l'apprentissage scolaire. Dans l'enquête ESLC cependant, la correction grammaticale, l'exactitude orthographique et le choix du vocabulaire n'ont pas joué un rôle dominant. Le consortium a privilégié l'efficacité communicative, telle qu'elle est prônée par le CECR. Les erreurs présentes dans les productions ont été évaluées en fonction de leur impact sur la communication. Elles n'ont été pénalisées que si elles entravaient la communication. Les autres erreurs n'ont pas été prises en compte.

Les réactions des correcteurs ont été diverses. Lors des deux séminaires de formation à l'évaluation de la production écrite qui se sont tenus au CIEP, les correcteurs ont perçu l'avantage d'évaluer les productions écrites de niveau A par comparaison avec des copies-étalon. Le sentiment général était que cette approche présentait des avantages car elle permettait d'obtenir un haut degré de corrélation entre les évaluateurs. La notation sur une échelle de 3 points (moins bon, aussi bon ou meilleur que la copie-étalon) n'a pas suscité de controverse. La notation sur 5 points pour les niveaux B s'est révélée plus complexe. Lors du premier séminaire qui a suivi la phase d'expérimentation sur le terrain, les niveaux B ont été évalués à partir de 4 critères : la réalisation de la tâche (seul critère commun avec les niveaux A), la compétence communicative (c'est-à-dire le lexique, l'aspect sociolinguistique et l'effet sur le lecteur), la compétence discursive (degré d'élaboration des phrases, cohérence et cohésion) et la correction linguistique (orthographe et grammaire). Les correcteurs ont trouvé ce système d'évaluation très complexe. Les analyses psychométriques qui ont été menées pour évaluer le degré de corrélation inter et intra-jury ont montré une certaine variabilité dans le degré de générosité ou de sévérité car les correcteurs ont eu tendance à interpréter certains critères différemment pour les niveaux B. Plus le nombre de critères est élevé, plus il est difficile d'identifier des profils de performance et d'établir une comparaison pour juger si une production est meilleure, aussi bonne ou moins bonne que la copie-étalon. C'est pourquoi le nombre de critères a été réduit à deux pour l'enquête finale (cf. chapitre 2.2., tableau 2) :

✓ A. Communication

- ✓ Combien de points sont réalisés ? (clarté)
- ✓ De quelle manière les points sont-ils développés ?
- ✓ Correction sociolinguistique (ton – registre)

✓ B. Langue

- ✓ lexique
- ✓ correction
- ✓ cohérence
- ✓ cohésion



Schéma 6 : Critères d'évaluation, niveau B, deuxième version

Il s'avère donc indispensable de réduire le nombre de critères pour pouvoir mettre en œuvre cette méthodologie, et d'arriver ainsi à un niveau satisfaisant de standardisation et de comparabilité en évaluation.

Comme indiqué au début de cet article, des exercices de familiarisation ont été nécessaires avant de passer à la pratique : le classement des productions écrites par rapport à des copies-étalon. Il fallait transmettre un concept important : chaque production devait être évaluée par rapport à une copie-étalon (meilleure, aussi bonne, moins bonne que celle-ci). Ceci a permis de réduire la complexité des prises de décision. Il était donc superflu de définir si les erreurs étaient d'ordre lexical ou grammatical. Cette méthode a également eu un impact décisif sur la manière d'évaluer : au lieu de compter les erreurs, l'évaluation s'est faite en tenant compte de la communication. Cela a permis de mieux comparer les performances dans les différentes langues. Alors que dans les enquêtes de l'OCDE, dites PISA (Programme international pour le suivi des acquis des élèves), les tests consistent en tâches identiques, traduites dans les différentes langues des pays participants, la Commission européenne a pris la décision de créer des tests parallèles mais non identiques pour les 5 langues testées. Ainsi donc, la comparabilité s'est avérée cruciale pour l'acceptation des résultats de l'enquête ESLC.

On pourra se référer en annexe aux exemples de productions en français et en allemand publiées dans le rapport final comme caractéristiques des niveaux A1, A2 et B1 du CECR. Ces « benchmarks » ont été sélectionnés suite aux analyses psychométriques qui ont suivi le séminaire de définition des scores de césure.

Pour davantage d'informations, on se reportera à l'article de Perlmann-Balme cité ci-dessous (2013).

Références

Conseil de l'Europe (2001), *Cadre européen commun de référence pour les langues: Apprendre, enseigner, évaluer*. Paris: Didier.

First European Survey on Language Competences – Technical Report (2012), [Online http://ec.europa.eu/lan-guages/policy/strategic-framework/documents/language-survey-technical-report_en.pdf. 28.07.2014].

First European Survey on Language Competences – Final Report (2012), [Online http://ec.europa.eu/lan-guages/policy/strategic-framework/documents/language-survey-final-report_en.pdf. 28.07.2014].

Informations sur l'Enquête européenne sur les compétences linguistiques et matériel de familiarisation en 5 langues pour les tâches de compréhension orale et écrite et de production écrite [Online <http://www.surveylang.org/>. 28.07.2014].

Sylvie Lepage, Roselyne Marty, Michaela Perlmann-Balme & Claudia Stelter (2014), Évaluation de la production écrite dans le cadre de l'Enquête européenne sur les compétences linguistiques (ESLC). *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 58-77. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Lepage_et_al.pdf.

Jones, Neil (2009), A comparative approach to constructing a multi-lingual proficiency framework - Standard setting research and its relevance to the CEFR. In: Figueras, Neus & Noijons, José (Eds.), *Linking to the CEFR Levels: Research Perspectives*. Cito, Council of Europe, EALTA [Online: http://www.coe.int/t/dg4/linguistic/EALTA_PublicatieColloquium2009.pdf. 28.07.2014].

Manuel pour Relier les examens de langues au Cadre européen de référence pour les langues [Online http://www.coe.int/t/dg4/linguistic/manuel1_fr.asp. 28.07.2014].

Perlmann-Balme, Michaela (2013), European Survey on Language Competences – Comparability of A1 level competences across five languages. In: Galaczi, E.D. & Weir, C.J. (Ed.), *Exploring Language Frameworks: Proceedings of the ALTE Krakow-Conference, July 2011*. *Studies in Language Testing (SiLT)* 36. Cambridge: Cambridge University Press, 85-102.

Annexe

Exemples de production écrite en français et en allemand, sélectionnées pour l'exercice de classement

TÂCHE 1 Photo de vacances



Tu es en vacances. Tu envoies un **email** à un ami avec cette photo de tes vacances.

Tu utilises la photo pour parler de :

- l'hôtel
- le temps
- les activités

Tu écris 20–30 mots.

Production 1

Je suis à un hôtel qui s'appelle Hôtel Bellevue. Le temps est beau et très chaud. Je dois beaucoup de sport parce que je joue au volley avec des amis tous les soirs.

Production 2

La hôtels nom est Hôtel Bellevue.

des enfants jouer de la football.

Production 3

L'été. Les enfants jouent au volley. Il fait tres shot. Derrier les enfants il y a la mer, devant hôtel Bellevue. J'aime l'été. Vive l'été !

Production 4

Salut Per,
 Qu'est-ce que tu fair dans tes vacances ?
 Je voyage à l'Italie.
 Le italiens sont très sympatique.
 J'ai habitent dans le Hotel Bellevue.
 Il fait du soleil et il est très chaud.
 Les garçons sont très mignon.
 Repondes-tu.

Lors de la mise en commun, les participants ont indiqué leur classement et justifié leur choix. Le rôle du formateur était de guider la discussion pour aboutir au classement prédéfini au préalable par l'organisme de certification en charge de la formation.

Ces productions ont été classées ainsi de la meilleure à la plus faible :

1, 4, 3, 2

Ces mêmes productions, regroupées en livret, devaient être évaluées par la suite à partir de copies-étalon.

TÂCHE 1 Photo de vacances

Exemples de productions en allemand classées de la meilleure à la plus faible :

Hallo !
 Ich habe jetzt Ferien und ich habe viel Spaß in Frankreich. Das Wetter ist warm, aber es ist nicht so heiß. Ich bin in Hotel Bellevue und es ist ein super Hotel ! Die Leute auf diese Foto sind meine Freunde. Wir spielen Volleyball am morgen.
 Bis bald.
 Dein Robin

Lieber Peter
 Du siehst Ich bin mit meiner Familie in Belleve Hotel in Antalya. Hier ist es gut. Die Speisen sind frisch und heiß. Hotel ist am strand. Nach dem Mittagessen spielen Volleyball Turnior.
 Bis später
 Kübra

Liebe Florence,
 Ich bin Hana. Hotel name ist Bellevue. Hotel ist seher gut. Wir spilen Volleball spilen. Ich habe sonne Tag.
 Viele GruÙe,
 deine Hana

Gut Tag
 Ich habe eine hotel « Bellevue ». Hotel ist sehr schon. Das Wetter ist OK, ist sonnig und heiÙ.
 Die Leute machen ok.
 Mit groÙe Leben !

Exemples de productions en franais et en allemand publiées dans le rapport final illustrant les niveaux A1-A2 et B1 du CECR

Tableau 6 : Exemples en franais, First European Survey on Language Competences - Final report – p.113

	Niveau A2 acquis	Encore au Niveau A1
A2 Nouveau passe- temps	Salut ! J'ai commenc� une nouvelle activit� ! J'ai r�alis�e une activit� de la lecture en semaine pass� et j'ai aim� parce que je peux �tudier les langues. Bisous !	Salut me ami. Je as commenc� une nouvelle activit�, est football. Je commenc� cette activit� en septembre. Je adore fait cette activit� pourque je adore sport. Adeus mi ami.
	Niveau B1 acquis	Encore au NiveauA2
B1 Membre de la fa- mille	Le membre de ma famille qui j'aime beaucoup c'est mon p�re. Il est sociable, un vraiment ami, amusant et sympathique. Ensemble, nous jouons au football, volleyball... Nous allons au th�atre, au cin�ma et nous allons vu le SLBenfica, au stadium. Il est du FCPorto et je suis du SLBenfica, et quand existe un Porto-Benfica, nous allons au stadium. Psicologiquement, nous sommes passives, amustants et intelligent. Nous nous entendons tr�s bien parce que, simplement, nous sommes p�re et fills."	Je suis Beata Schmidt, j'ai 16 ans, le membre de ma famille que j'aime tr�s beacoup est ma cousine Magda. Je l'aime beacoup parce-que je et elle nos entendons tr�s bien, parce-que elle m'aime et je l'aime. Nous aime etudi� group� et nous sortons a promener. En fin nous aime beacoup entre nous. Il y a 19 ans et elle etude anas-thesie en ovida, en la universit�. Fin, j'aime ma cousine"

Tableau 7 : Exemples en allemand First European Survey on Language Competences - Final report – p.114

	Niveau A2 acquis	Encore au Niveau A1
A2 Neues Hobby	<p>Liebe Rose, Ich liebe Kino! Das ist mein neue Hobby!</p> <p>Ich gehe ins Kino wenn gibt es Gute Film zu sehen. Mein Lieblings-Film ist „Harry Potter“ oder „Some like it hot“ mit Marilyn Monroe! Meine Lieblings-schauspielerin sind Marilyn Monroe, Rose McGowen und Shannen Doherty. Ich mage Kino, weil du kannst Film sehen.</p> <p>Angela</p>	<p>An: Andrzej Von: Maria</p> <p>Ich habe ein neues Hobby. Meine neues Hobby ist spielen Volleyball. Ich spielen Volleyball seit drei Jahre. Ich mag spielen Volleyball.</p> <p>Lieben Grußer</p>
	Niveau B1 acquis	Encore au niveau A2
B1 Familienmitglied	<p>Hallo, wie geht es?</p> <p>Ich schreibe über meinen kleinen Bruder Tadek. Er ist 6 Jahre alt. Er ist sehr komisch und magt Spongebot Swammkopf. Er sieht ihn auf Deutsch. So hat er Deutsch gelernt. Er geht in den Kindergarten und hat Freunde. Wir haben nicht sehr viel gemeinsames veil wir nicht die selbe Generation sind, aber wir verstehen uns sehr gut.</p> <p>Er errinert mich (an) auf mich wan ich klein war. Nicht mit dem aussehen, aber mit gedanken. Ich mag meinen kleinen Bruder und mag es Zeit mit ihnen zu verbringen.</p>	<p>'Liebe Darin!</p> <p>Ich ferbringe viele Zeit mit meiner Familie. Meine Mutter und ich spielen viele spiele und sie erzelt mir geshiste uber seine kindertage. Meine Oma ist immer auf der vardets und ich helfe sie. Mit meine Tante ferbringe ich nicht viele zeit aber unser Zeit zuzamen ist lustisch. Mit meinen Onkel ferbringe ich die ganze vochenende. Wir turnen und spielen fußball. Unsere Zeit zuzamen ist lustich aber argern aber der streight ist nicht groß. Wir versteht uns gut veil wir volen z u verstanden uns. Wir sind eine lustige Familie.</p> <p>Liebe gruse, Filip</p>

Note

¹ Le modèle de Rasch est une méthode statistique d'analyse psychométrique notamment employée pour mesurer les propriétés des items de façon indépendante des groupes d'individus auxquels ils ont été administrés. Avec ce modèle il est possible de classer toutes les productions sur une seule échelle de compétence, quelle que soit la langue.