
L'évaluation des compétences linguistiques

Une analyse des relations entre les niveaux du CECR et les notes attribuées par les enseignants

Gieruc Gabriella *, Irene Lys**

** Unité de recherche pour le pilotage des systèmes pédagogiques (URSP)*

Signal 11

1014 Lausanne-CH

gabriella.gieruc@vd.ch

*** Haute école pédagogique*

Cour 33

1014 Lausanne-CH

irene.lys@hepl.ch

RÉSUMÉ. Notre thématique s'inscrit dans la récente évolution de politique éducative en matière d'évaluation des compétences linguistiques en langues secondes à l'école (introduction du Portfolio Européen des langues (PEL) et de standards sur la base du Cadre européen commun de référence pour les langues (CECR)).

En réponse à ces projets, dans le cadre d'une étude visant à évaluer l'efficacité d'un modèle de filière bilingue (baccalauréat bilingue français-allemand) réalisé dans un canton de la Suisse francophone, nous avons voulu explorer la relation entre le CERC et l'évaluation courante des performances linguistiques dans le milieu scolaire. Pour ce faire, nous avons basé l'évaluation de 175 textes produits à l'examen écrit de baccalauréat sur les descripteurs attachés aux différents niveaux du CECR, opérationnalisés par le PEL. Cela nous a permis d'analyser la relation entre les niveaux de compétence que nous avons attribués aux travaux des élèves et les notes données par les enseignant(e)s. Grâce à une série d'indicateurs de performance linguistique, nous avons pu mettre en évidence la relative proximité entre ces deux démarches évaluatives

MOTS-CLÉS : Evaluation - textes - indicateurs de performance linguistique - Cadre européen commun de référence pour les langues

1. Introduction: contexte et objectifs de la recherche

La toile de fond de notre thématique est constituée par l'évolution récente, en Suisse, des politiques éducatives en matière d'évaluation des compétences linguistiques à l'école. En effet, le *Portfolio européen des langues* (PEL) sera introduit, dans un futur proche, comme outil d'évaluation et d'enseignement des langues secondes dans la scolarité obligatoire et post-obligatoire. De plus, dans une visée d'harmonisation de l'enseignement, des standards seront fixés au niveau national, très vraisemblablement sur la base du CECR, pour déterminer des niveaux "seuil" en langues secondes à différents moments de la scolarité. Ces innovations ne manqueront pas de susciter débats et interrogations au sein du corps enseignant et de l'institution, puisqu'elles remettent en question les pratiques et les objectifs actuels de l'enseignement des langues.

Dans le cadre d'une étude visant à évaluer l'efficacité d'un modèle d'enseignement bilingue (français-allemand) en filière de baccalauréat dans la Suisse francophone, nous avons voulu apporter des éléments de réponse aux questions suivantes:

- le *Portfolio pour jeunes et adultes* se prête-t-il à l'auto-évaluation des élèves de baccalauréat?
- les descripteurs attachés aux différents niveaux du CECR peuvent-ils constituer un point de départ pour l'élaboration de critères d'évaluation au secondaire supérieur, ou ce cadre précis nécessite-t-il d'autres descripteurs plus circonstanciels ?
- dans le domaine de l'expression écrite, quelles relations y a-t-il entre les niveaux du CECR, l'échelle de notation utilisée par les enseignants pour ce même genre de productions et une série d'indicateurs de performance linguistique que nous avons utilisés pour analyser les textes des élèves? Les différents outils d'évaluation mis en oeuvre aboutissent-ils à des classements des textes cohérents entr eux ?
- les performances des élèves testés correspondent-elles aux compétences attendues ?

2. Méthodologie

2.0. Population

Notre population est constituée de deux groupes d'élèves de troisième année de lycée, inscrits en filière de baccalauréat¹ bilingue (élèves de Maturité bilingue: dorénavant MBIL) ou régulière (élèves de filière régulière: dorénavant MREG). Ces

¹ Dans le canton de Vaud, la filière menant aux études universitaires est appelée filière de maturité.

derniers constituent le groupe de contrôle. Les élèves ont été ultérieurement partagés, en fonction de leur famille d'origine: une première analyse de la population a en effet montré qu'un élève de filière bilingue sur trois environ est issu d'une famille où l'allemand est la langue (ou une des langues) de l'un ou l'autre de ses parents. Nous aboutissons ainsi à quatre groupes d'élèves:

Elèves	Allemand à la maison (+ all)	Français (ou autre) à la maison (-all)	Total
MBIL	22	42	64
MREG	8	103	111

Tableau 1. Répartition en 4 groupes

2.1. Corpus et analyses

Le corpus que nous présentons ici est constitué de deux séries de textes, écrits par les élèves au temps 1 (après deux ans de cursus de baccalauréat) et au temps 2 (lors de l'examen de baccalauréat en allemand, une année plus tard). Les productions écrites des élèves ont été analysées de différentes manières :

- elles ont été évaluées sur la base des descripteurs du CECR, opérationnalisés dans le PEL, et classés d'après cette échelle (évaluation de la qualité globale) ;
- elles ont été analysées sur la base de critères linguistiques quantitatifs et qualitatifs (indicateurs de performance) ;
- des traitements statistiques (régression linéaire simple et régression multiple) ont été réalisés, pour déterminer d'une part les relations entre notes et niveaux CECR, et d'autre part les relations entre ces deux classements et les indicateurs linguistiques.

3. Résultats

3.0. Adéquation du PEL pour jeunes et adultes pour l'évaluation des productions écrites au niveau baccalauréat

L'une des premières questions qui nous intéressait était de déterminer si les outils proposés par les PEL étaient exploitables pour l'évaluation dans le contexte particulier où notre observation se déroulait.

Lors du premier volet de notre recherche (après deux ans d'immersion), le *PEL pour jeunes et adultes* avait été utilisé pour le classement des textes. Or, en fin de parcours de baccalauréat, une année plus tard, nous avons constaté qu'il n'était pas aisé de se baser sur ses listes de repérage. Plus particulièrement, les trois premiers niveaux de compétence (A1, A2, et partiellement B1) ne se prêtaient pas à

l'évaluation des productions écrites que nous avons à étudier. Pour ces niveaux, le PEL décrit en effet des activités langagières authentiques, en dehors du contexte scolaire. L'inadéquation constatée peut avoir deux origines: soit les descripteurs utilisés ne sont pas adaptés aux tâches demandées dans le contexte d'une école de baccalauréat, soit le niveau de compétence décrit est trop élémentaire par rapport aux types de tâches demandées. En l'occurrence, certains des textes produits semblaient correspondre au niveau de compétence très élémentaire des descripteurs A2 et partiellement B1; le type de descripteur utilisé, par contre, ne semblait pas adapté aux contenus demandés. Cette constatation met en évidence une des difficultés à prévoir au moment de généraliser l'utilisation des PEL dans les écoles de baccalauréat : les descripteurs utilisés doivent être adaptés aux tâches demandées aux élèves. En ce qui concerne notre recherche, nous avons résolu ce problème en changeant d'outil de référence : pour nos analyses finales, nous nous sommes basées sur les descripteurs du *Portfolio européen – éducation supérieure*. L'échelle de descripteurs utilisée va de A2 à C2.

3.1. Niveaux de compétence attendus et niveaux constatés

Le tableau 2 résume le classement des textes en niveaux de compétence CECR.

Elèves / niveaux du Cadre commun	MBIL+all (22)	MBIL-all (42)	MREG+all (8)	MREG-all (103)	total (% de 175)
A2			1	19	11.4%
B1.1			1	23	13.7%
B1.2	1	3	3	32	22.3%
B2.1	1	11		15	15.4%
B2.2	2	10		10	12.6%
C1.1	5	3	1	2	6.3%
C1.2	4	8		2	8%
C2.1	3	6	1		5.7%
C2.2	6	1	1		4.6%

Tableau 2. *Elèves et niveaux CECR*

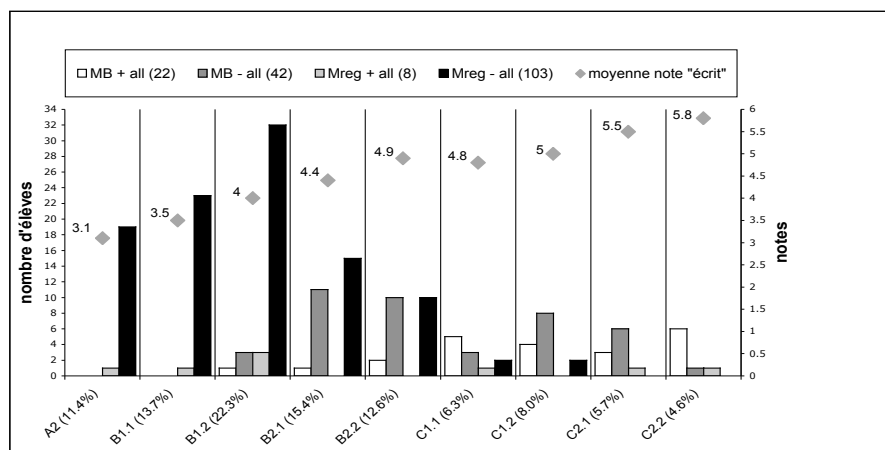
Au vu de ces résultats, une première constatation s'impose : les élèves de filière bilingue, qu'ils aient bénéficié ou non d'un contact avec l'allemand à la maison, produisent des textes d'une meilleure qualité globale. Mais, par rapport à nos prévisions ou aux éléments qui donnent un aperçu des attentes institutionnelles en voie de définition, les conclusions sont moins positives.

Après deux ans d'immersion, lors de notre première prise d'informations, les niveaux de compétence constatés nous avaient amenés à émettre des pronostics pour la fin de la troisième année : nous nous attendions à ce que les élèves MBIL

atteignent le niveau C1 et les MREG B2. Nos attentes ont donc été réalisées en ce qui concerne les élèves MBIL+all (plus de 80% de ces élèves se situent au niveau C1-C2), mais pas par les élèves MBIL-all (seuls 43.1% des travaux se trouvent au niveau préconisé ou au-dessus). Quant aux élèves MREG-all, seulement 27.2% des textes se trouvent au niveau repère B2 ou au-dessus. On peut attribuer la non réalisation de nos attentes (en particulier en ce qui concerne les élèves MREG) à plusieurs facteurs : situation d'apprentissage qui induit une stagnation dans les apprentissages, palier difficile à franchir entre les niveaux de compétence ou simplement attentes inadaptées au niveau réel des élèves. La comparaison du classement en niveaux avec les notes attribuées par les enseignants devrait nous permettre d'approfondir la réflexion sur cette question.

3.2. Niveaux de compétences du CECR et notes

Au cours de notre recherche, nous avons pu constater que l'absence d'attentes institutionnelles clairement établies empêche la standardisation des pratiques. Dans ces conditions, on peut se demander si les critères pour obtenir la note seuil sont réellement comparables d'un établissement scolaire à l'autre. Notre travail d'analyse nous a permis de comparer le classement des textes en niveaux du CECR (effectué dans le contexte de la recherche, par des chercheurs et des experts) et les échelles de notation appliquées par les enseignants (dont nous ne savons pas sur quels critères elles reposent exactement). Le graphique ci-dessous offre une image synthétique de la convergence entre ces deux échelles d'évaluation. Ce graphique complexe visualise le pourcentage d'élèves qui ont atteint un niveau donné : pour le niveau A2, par exemple, il s'agit de 11.4% des 175 élèves. De plus, il nous renseigne sur le nombre d'élèves de chaque groupe à avoir obtenu ce niveau (1 des 8 élèves MREG+all et 19 des 103 élèves MREG-all). Finalement, le graphique nous indique la moyenne des notes obtenues (3.1) par ces 20 élèves. Pour plus de clarté, la note seuil est 4.



Graphique 1. Relation entre la moyenne des notes obtenues à l'examen et les niveaux CECR

D'après les résultats illustrés ici, nous pouvons constater que la progression entre les deux échelles est parallèle et équilibrée (différence de 0.4 à 0.6 points entre un niveau et l'autre). Ce parallélisme indique que les deux manières d'évaluer n'entrent pas en contradiction, ce qui est encourageant. Remarquons cependant qu'il est interrompu en ce qui concerne les niveaux B2.2, C1 et C1.2. Il semble que ces trois niveaux ne correspondent pas à des paliers significatifs pour les enseignants. Pour cette zone de compétence, on dirait que le classement pratiqué par les enseignants utilise des critères moins fins que les nôtres. L'existence de ce « plateau », en tous les cas, confirme la nécessité d'un travail d'explicitation des critères sous-jacents aux deux échelles.

Les moyennes occultent la dispersion des notes réelles. Le tableau 3 montre que cette dispersion est plus prononcée pour les niveaux les moins exigeants.

Note / Niveau (+moy notes)	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6	TOT élèves
A2 (3.1)	1	1	6	2	7	2	1				20
B1.1 (3.5)			1	7	10	1	5				24
B1.2 (4)		1	1	2	9	12	11	2		1	39
B2.2 (4.9)						3	6	5	8		22
C1.1 (4.8)						2	3	3	3		11
C1.2 (5.0)							5	5	2	2	14
C2.1 (5.5)							1	2	3	4	10
C2.2 (5.8)								1	2	5	8
TOTAL élèves	1	2	10	11	28	29	37	23	22	12	175

Tableau 3. Notes obtenues par niveau CECR (nombre d'élèves par note/niveau)

Pour le niveau A.2 (20 élèves), nous trouvons des notes allant de 1.5 à 4.5. Dans ce cas, il est clair que les évaluations des différents enseignants ne s'harmonisent pas. De même, la dispersion est très prononcée pour le niveau B1.2 (moy: 4). A quoi cela est-il dû ? A l'hétérogénéité des élèves ou à celle des critères des enseignants ?

3.3. Seuil de compétence attendu versus note seuil

Il nous a semblé intéressant de pousser l'analyse en ce qui concerne la relation en niveaux CECR et la note seuil, le 4. A quel niveau cette note correspond-elle ? Peut-on discerner des relations claires entre le 4 et l'un des niveaux ? Les attentes minimales sont-elles différenciées par filières ? Le tableau 4 montre qu'en réalité les attentes minimales n'ont pas été différenciées : pour les quatre groupes d'élèves, le 4 correspond au même niveau.

Moyenne note/ niveau	MBIL + all (moy: 5.0)	MBIL – all (moy: 4.6)	MREG + all (moy: 4.6)	MREG – all (moy: 3.9)
A2			4	3
B1.1			3.5	3.5
B1.2	4	4	4	4
B2.1	4	4.4		4.4
B2.2	4.8	4.9		5
C1.1	4.6	4.7	5.5	5.3
C1.2	5	4.9		5.8
C2.1	5.5	5.4	6	
C2.2	5.7	6	6	

Tableau 4. *Notes moyennes et niveaux*

D'après les correspondances observables ici, les enseignants ne situent pas le niveau seuil exprimé par la note 4 à B2 (comme nous l'avions envisagé pour les élèves MREG), mais à B1.2 pour tous les élèves. Cela appelle quelques commentaires.

Les enseignants ne différencient pas leurs attentes minimales en fonction des filières (bilingue ou régulière). Cette pratique peut s'expliquer par le fait que le cadre institutionnel ne fixe pas d'attentes spécifiques aux filières d'enseignement bilingue au niveau national. Cependant, cet état de choses devrait être interrogé, puisque les conditions d'enseignement ne sont pas les mêmes dans un cas et dans l'autre. Dans l'un des établissements, la filière bilingue offrait presque 5 fois plus de périodes enseignées en allemand que la filière régulière. A notre avis, cette disparité devrait aboutir à une évaluation plus exigeante en filière bilingue. Par ailleurs, nous constatons que les exigences des enseignants augmentent effectivement par rapport aux productions des élèves MBIL, en ce qui concerne les niveaux de compétence plus élevés : pour un texte de niveau C1.1, les élèves MBIL obtiennent une moyenne de 4.6/4.7 et les élèves MREG une moyenne de 5.5/5.3. Le même effet peut être observé pour les niveaux avoisinants.

La comparaison entre les notes moyennes et les niveaux permet de mettre en évidence la stabilité des évaluations des enseignants, ce qui vient contredire l'argumentation que nous avons développée plus haut. Même en l'absence d'attentes formulées de manière objectives ou de conditions de passation des examens standardisées, nous observons une identification claire du niveau de compétence associé à la note 4 (B1.2). Il est intéressant de noter, à ce propos, que cette appréciation relativement précise du niveau seuil de la part des enseignants, fondée

vraisemblablement sur une base empirique, ne semble correspondre ni aux niveaux de compétences actuellement déjà en vigueur², ni à ceux préconisés³. Cette constatation incite à la prudence dans la fixation de standards, d'autant plus s'ils sont certificatifs, sous peine de voir les échecs augmenter drastiquement.

3.4. Niveaux CECR, notes et indicateurs de performance linguistique

Outre l'évaluation en niveaux du CECR (d'après leur qualité globale), les productions écrites des élèves ont également été analysées en fonction d'indicateurs de performance linguistique. L'un d'entre eux est qualitatif, comme la qualité globale: il s'agit de la textualité (évaluation de la cohérence et de la construction du texte). D'autres, sélectionnés pour leur pertinence après le premier volet de notre recherche, sont quantitatifs: longueur du texte, richesse lexicale, comptage des erreurs de mots pleins/idiomaticité, des erreurs de valence (construction du verbe), calcul de la densité des problèmes lexicaux observés, du taux d'erreurs concernant la place du verbe/des éléments du groupe verbal, calcul de la densité globale des erreurs. Ces critères quantitatifs, reliés à des éléments observables, offrent l'avantage d'une évaluation qui échappe à la subjectivité. On ne peut pas en dire autant en ce qui concerne le classement des textes selon les niveaux du CECR. Ces derniers, malgré leur précision et la cohérence avec laquelle ils ont été construits, laissent toujours une marge à l'interprétation personnelle. Les notes des enseignants, quant à elles, sont complètement opaques quant aux critères qui les sous-tendent. La mise en relation de paramètres quantitatifs avec les deux échelles d'évaluation analysées devrait permettre de préciser la nature des attentes des enseignants et de voir dans quelles mesure elles convergent avec les critères du *Cadre commun*. Dans notre démarche, nous partions d'ailleurs de l'hypothèse que les résultats des élèves par rapport aux différents indicateurs de performance linguistique utilisés correspondraient autant aux notes attribuées par les enseignants qu'aux niveaux CECR attribués par les chercheurs. Pour tester ce postulat, nous avons eu recours à plusieurs démarches complémentaires. Une régression linéaire simple pour chaque couple de variables nous a permis d'examiner le lien entre indicateurs et notes d'un côté, niveaux de l'autre.

Variable explicative (indicateur de compétence)	Pourcentage de la variance du score expliquée	
	Qualité globale (maturité)	Note examen écrit (maturité)
Textualité	41%	38%
Qualité globale	-	55%
Densité des problèmes lexicaux	53%	49%

² Pour la Maturité (baccalauréat) professionnelle commerciale: B2 (niveau général).
Entrée en filière tertiaire de formation des futurs enseignants : B2.

³ B1.2 en compréhension orale à la fin de la 9e année.

Erreurs de valence	26%	27%
Erreurs mots pleins et idiomaticité	49%	43%
Erreurs de position du verbe	36%	37%
Densité globale des erreurs	(35%)	(36%)
Richesse lexicale	20%	19%
Longueur du texte	10%	10%

Tableau 5. Régressions simples (indicateurs linguistiques de compétence/ notes et niveaux)

Ce tableau présente, sous la forme d'un pourcentage, la variance du score qui peut être expliquée par un seul indicateur de compétence. Les performances des élèves dans les différents domaines varient de manière comparable dans les deux types de classement des textes. Nous pouvons ainsi vérifier que les critères qui sous-tendent les deux échelles d'évaluation ne sont pas fondamentalement différents. De légères différences existent bien (+ ou - 1%), mais il nous est difficile de les interpréter. Dans deux cas, par contre, l'impact de l'indicateur est plus important pour l'évaluation de la qualité globale (niveaux du CECR) : pour les erreurs de mots pleins (+6%) et pour la densité des problèmes lexicaux (+ 4%). Cette différence pourrait indiquer que l'évaluation basée sur le *Cadre commun* donne davantage d'importance à la maîtrise du vocabulaire que l'évaluation faite par les enseignants. Dans les deux échelles, cependant, les performances lexicales ont une influence prépondérante par rapport à l'évaluation, la maîtrise du vocabulaire primant sur son étendue.

A l'aide de calculs de régression multiple, nous avons tenté d'identifier le groupe d'indicateurs qui permet la meilleure explication de la note et du score obtenu.

Modèle 1	
Variables	Contribution à l'explication / la prédiction de la note
<i>Qualité globale</i>	55.1%
Toutes les variables	67.3%
Modèle 2	
Variables	Contribution à l'explication / la prédiction de la note
<i>Densité des problèmes lexicaux + richesse lexicale</i>	59.2%
Toutes les variables	66%
Modèle 3	

Variabes	Contribution à l'explication / la prédiction de la qualité globale
<i>Densité des problèmes lexicaux + richesse lexicale</i>	54.8%
Toutes les variables	71%

Tableau 6. Régression multiple (modèle explicatif le plus puissant)

Les modèles 1 et 2 concernent les relations entre indicateurs et notes : quels indicateurs contribuent le mieux à expliquer - ou prédire - les notes attribuées ? Le modèle 3, par contre procède à la même analyse mais par rapport au classement selon les niveaux CECR (qualité globale). Les variables prise en compte sont des indicateurs quantitatifs (densité des problèmes lexicaux, richesse lexicale, erreurs de position du verbe) et des indicateurs qualitatifs (textualité et qualité globale).

En comparant les modèles explicatifs des notes, nous constatons qu'un critère qualitatif (qualité globale, autrement dit le classement en niveaux) peut expliquer une grande partie (55.1%) de la variance de la note. Le recours aux critères quantitatifs (densité des problèmes lexicaux et richesse lexicale) aboutit à un résultats comparable, 59.2%, dans le modèle 2. Ce premier résultat met en exergue qu'une évaluation qualitative des textes, basée sur les descripteurs du *Cadre commun* a une force explicative comparable à celle basée essentiellement sur des critères linguistiques quantifiés, dont on peut penser qu'ils sont plus fiables. Autrement dit, l'évaluation sur la base du *Cadre* n'est encore une fois pas trop éloignée de l'évaluation couramment pratiquée par les enseignants. Ce constat est d'ailleurs confirmé par la comparaison des modèles 2 et 3, où la force explicative des indicateurs liés au lexique (densité des problèmes lexicaux, richesse lexicale) est montrée par rapport aux notes et aux niveaux : ces critères expliquent une grande partie des évaluations attribuées dans les deux cas (59.2% pour les notes, 54.8% pour les niveaux). Dans ces deux modèles, la densité des problèmes lexicaux a le poids prédictif le plus grand: en prenant en compte ce seul indicateur, on explique environ 48% des notes et 53% de la qualité globale.

Ces résultats démontrent donc le parallélisme entre les deux échelles d'évaluation comparées. Ils rendent également plus transparents les critères sur lesquels elles se fondent : la prise en compte des erreurs lexicales et, en moindre partie, de la richesse lexicale. Ainsi, nous pouvons montrer que les deux systèmes ne sont pas aussi éloignés l'un de l'autre que l'on pourrait le croire.

4. Conclusion

A travers cette recherche, nous avons mis en évidence plusieurs aspects touchant aux attentes et à l'évaluation dans le cadre d'une filière bilingue de baccalauréat. Les compétences attendues n'étant pas clairement définies par l'institution scolaire, nous assistons à une forme d'autorégulation des pratiques évaluatives.

En tout premier lieu, il faut signaler que le niveau atteint ne correspond ni aux attentes formulées par les chercheurs, ni aux niveaux déduisibles du contexte curriculaire. De plus, le niveau seuil en filière bilingue est visiblement le même qu'en filière régulière, de manière plutôt contradictoire, puisque les conditions d'enseignement/apprentissage ne sont pas les mêmes.

Pour contrebalancer ces constatations plutôt négatives, nous avons montré que les échelles en niveaux de compétence du CECR ne sont pas très éloignées des pratiques évaluatives des enseignants. Dans les deux démarches, les aspects lexicaux présentent une importance prépondérante. Des travaux sont sans doute encore nécessaires afin de rendre ces échelles de niveaux opérationnelles pour l'évaluation dans le cadre du baccalauréat, pour les outiller en indicateurs linguistiques, comme nous l'avons fait. Cependant, notre analyse montre l'applicabilité et la fiabilité de cet outil d'évaluation.

Enfin, à partir de notre analyse, des pistes pourraient être dégagées dans le sens d'une pondération entre des critères qui restent en partie intuitifs malgré la sophistication de leur élaboration (les descripteurs des niveaux CECR), mais qui sont très puissants, et des critères quantifiables (les indicateurs linguistiques), objectifs mais d'une portée limitée.

Bibliographie

DFJ (2001). *Préparation d'une maturité avec mention bilingue français-allemand dans les écoles de maturité des gymnases du Canton de Vaud*. Lausanne : DFJ.