



<https://bulletin.vals-asla.ch> · e-ISSN: 2813-7973 · p-ISSN: 1023-2044

North, B. (1992). Activités de communication et évaluation de l'oral. *Bulletin CILA*, 55, 97-108.
<https://doi.org/10.5169/seals-978062>

Cet article est publié sous la licence *Creative Commons Attribution 4.0 International* (CC BY):
<https://creativecommons.org/licenses/by/4.0>



© Brian North, 1992

Activités de communication et évaluation de l'oral

Cet article décrit un système d'évaluation de l'oral développé à l'Eurocentre de Bournemouth, l'un des six Eurocentres d'Angleterre entre 1986 et 1988 et qu'on est actuellement en train d'introduire pour l'enseignement de l'anglais dans les autres Eurocentres. C'est un système d'évaluation en groupes, dans la salle de classe, et non par test individuel.

Contexte:

Le système a été développé dans le contexte d'un projet qui s'appelle LAC/NEC «Levels, Assessment & Certification», ou «Niveaux, Evaluation et Certification», et qui consiste en un système intégré de planification et d'évaluation sur la base d'une échelle de compétences de neuf niveaux. Cette échelle existe en versions différentes pour définir les objectifs dans chacune des langues enseignées par les Eurocentres, et en versions séparées pour l'évaluation à l'oral et l'évaluation à l'écrit. On est en train de développer également des versions pour l'attestation d'un nouveau certificat qui sera introduit en 1993.

Ce projet LAC/NEC essaie d'introduire avec l'échelle des compétences un cadre de références transparent et cohérent:

- transparent, c'est à dire que l'apprenant peut se situer dans le parcours de son apprentissage, et se rendre compte de son progrès par rapport à l'échelle avant de négocier les mesures efficaces pour améliorer ses résultats. On a essayé de développer un système d'évaluation transparent pour que les stagiaires puissent le comprendre, prendre part au processus, et puissent développer la possibilité d'auto-évaluation.
- cohérent, c'est à dire que l'interprétation des niveaux est plus ou moins semblable dans tous les Eurocentres qui offrent les stages dans la même langue, et les différentes versions développées pour chaque langue. On a essayé de développer un système cohérent qui explique de façon claire la relation entre les niveaux des Eurocentres et les examens officiels, pour que les stagiaires puissent comprendre les niveaux sur l'échelle comme étapes réelles dans leur parcours vers l'objectif choisi.

Pour un système d'évaluation transparent, nous voulions une évaluation simple et directe qui permette aux stagiaires de se sentir partenaires et pas seulement objets évalués. Cette approche conduisait à une évaluation de la classe dans la salle de classe au moyens d'activités de communication.

Au début des années 80, il y avait un débat sur la faisabilité de l'évaluation directe de la communication; on a parlé de «L'évaluation *de* la communication» (l'évaluation directe) en opposition à «L'évaluation *pour* la communication» (l'évaluation indirecte) (ALDERSON and HUGHES, 1981; HUGHES and PORTER 1983). Bien sûr on ne peut pas simuler la vraie communication dans la salle de classe, mais, d'autre part, on peut envisager des activités de communication qui se déroulent de façon structurée et pourtant autonome: structurée pour que tout le monde ait quelque-chose à dire, et, à un certain moment, le droit à la parole; autonome, décentralisée, avec un scénario assez large, assez libre, pour que chacun puisse dire ce qu'il veut, en utilisant toutes les ressources linguistiques dont il dispose.

De plus, on peut soutenir que, – évaluation mise à part – des activités qui regroupent les acquis les plus importants de l'unité d'enseignement et qui donnent à l'étudiant la liberté de s'exprimer avec toutes ses ressources, sont en soi non seulement valables, mais également indispensables dans un programme d'apprentissage qui porte l'étiquette «communicative».

Ainsi, la base de notre système d'évaluation est l'exploitation de ces activités de communication dans le programme, d'une part pour l'évaluation formative continue en faisant un diagnostic – parfois en enregistrant les performances sur magnétoscope (chaque semaine ou tous les quinze jours) – et d'autre part pour l'évaluation sommative, en situant les stagiaires sur l'échelle par rapport à la grille des critères (tous les quinze jours ou tous les mois). L'utilisation de ces activités comme base du système, amène à une évaluation tout à fait intégrée dans le programme, et met l'accent sur l'importance de l'interaction communicative entre les stagiaires.

L'évaluation en groupes est une évaluation sur la base d'un discours naturel et égalitaire mené par le groupe et qui n'est pas dominé par le professeur. Le professeur, l'examineur, joue le rôle d'observateur et n'intervient pas pendant le déroulement de l'activité, sauf pour donner des instructions ou réorganiser les groupes pour une nouvelle phase. De cette façon, il ne fait pas partie du discours à évaluer et par conséquent, il conserve une objectivité plus grande et également plus de recul, pour se référer à sa grille de critères.

Mais que peut-on utiliser comme types d'activités? Ce qui est le plus important c'est que les activités donnent aux étudiants la liberté d'utiliser toutes leurs ressources. Il y a des activités qui consistent en un entraînement à la communication, avec la répétition de quelques phrases dans un contexte restreint, par exemple un petit jeu de rôles dans un magasin. A mon avis une telle tâche est trop restrictive pour donner un exemple représentatif. On a besoin d'un scénario plus large, avec plusieurs phases.

Deuxièmement, pour arriver à un exemple de production représentative, il faut qu'il y ait des phases différentes qui génèrent des discours différents – une heure d'activité du même type de discours n'est pas représentative. Dans une analyse faite à l'université de Birmingham en 86, sur plusieurs types d'activités, et comportant à peu près trente catégories (NORTH 1986), j'ai discerné trois phases principales dans les activités de communication:

- D'abord, une phase que j'appellerai la phase de collaboration, pendant laquelle des groupes de trois à cinq étudiants discutent de la façon dont ils vont exécuter la tâche donnée, trouvent l'information nécessaire et développent leurs points de vue. Cette phase est caractérisée par des interventions courtes mais relativement lentes, et par une utilisation relativement intensive des stratégies de communication.
- Ensuite, une phase d'échange d'information et d'opinions, normalement dans un nouveau groupe formé d'étudiants des autres groupes. Cette phase est caractérisée par une participation beaucoup plus formelle, plus longue, plus cohérente et plus lente.
- Enfin, une discussion, soit après chaque échange, soit comme phase distincte. Cette phase est plus difficile à définir, mais les interventions sont relativement rapides, surtout au niveau supérieur, et tout le monde ne prend pas toujours une part active.

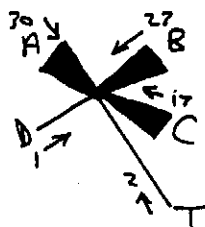
Quand on analyse un échantillon représentatif du discours pendant les trois phases selon le modèle classique de SINCLAIR et COULTHARD, on obtient ces formes graphiques qui représentent chacune trois minutes d'un exemple de chaque phase principale (SINCLAIR and COULTHARD 1975). Les chiffres indiquent les «Initiations» suivant le modèle «Initiation, Response, Feedback». Les «Initiations» ce sont les interventions de celui qui prend l'initiative.



EXCHANGE



DISCUSSION



- Dans la phase de collaboration, on peut voir qu'il y a eu participation de tous les étudiants et que tout le monde a pris l'initiative.
- Dans la phase d'échange, chaque étudiant a la parole, l'un après l'autre. La diagramme montre deux types de réactions extrêmes. Dans le premier, «B» a trouvé la contribution de «A» si difficile qu'il l'a interrompue onze fois, et les deux autres ont commencé à intervenir pour que «A» puisse garder la parole! Dans le deuxième exemple c'est l'inverse - personne n'a interrompu. Ces exemples sont deux extrêmes, normalement on trouve un discours entre les deux: quelqu'un a la parole, mais de temps en temps les autres posent des questions, et ajoutent des avis ou des informations supplémentaires.

- Dans la phase de discussion, A, B et C ont été très actifs, mais «D», qui est Japonais, n'a presque rien dit. («T» est le professeur, qui est intervenu deux fois pour dire que l'activité devait prendre fin dans quelques minutes et est reparti.) C'est un discours typique: une discussion libre donne l'occasion aux étudiants les plus forts de prendre davantage la parole et de montrer ce qu'il peuvent faire.

On peut dire que l'exemple de discours est plus représentatif si on a les trois phases. Si une activité a toujours la même phase, il n'y a pas assez de variété pour faire une évaluation véritable: s'il s'agit seulement d'une phase de collaboration, le discours manque de cohérence; si on a seulement un échange, - une présentation de la part d'un étudiant - il n'existe pas d'interaction spontanée; et si on a seulement une discussion libre, il y aura toujours des étudiants qui ne seront qu'auditeurs, ne disant presque rien.

Nous avons donc sélectionné et élaboré des activités qui se déroulent en trois phases. Il y a toute une gamme d'activités qu'on peut utiliser, mais nous avons trouvé utile de nous concentrer sur cinq types seulement:

Type 1: Etudes de cas

Ce type regroupe des activités au cours desquelles on échange des informations, raconte des anecdotes et trouve des solutions aux problèmes donnés. On a quatre groupes dans la classe, A,B,C,D, qui discutent entre eux. Après la phase de collaboration les groupes se mélangent, et on forme de nouveaux groupes

Exemple: Lettres à un magazine pour demander des conseils sur des problèmes personnels. Chaque groupe a une lettre différente: Première phase: collaboration en groupes pour trouver des solutions, des conseils; phase secondaire avec les nouveaux groupes: explication des problèmes et des conseils proposés par les premiers groupes, (phase d'échange); évaluation des conseils proposés, formulation de nouvelles solutions, et élargissement de la discussion sur des problèmes similaires (phase de discussion).

Autre exemple: Le mystère de la disparition du Marie Céleste. Chaque groupe initial reçoit des théories différentes pour expliquer l'événement. Quand on forme les nouveaux groupes, on explique des théories différentes (phase d'échange) et on les évalue (discussion).

Type 2: Réunions de travail informelles

Dans ces activités, les trois phases se déroulent normalement dans le même groupe, qui simule un groupe de travail ou de recherche. Tout le monde joue le même rôle pour résoudre certains problèmes. Les rôles ne font que

créer le contexte; il n'y a pas de confrontation programmée comme avec le Type 4.

Exemple: Réunion de professeurs pendant laquelle on discute de la façon dont on va organiser quelque chose, ou dont on va utiliser une donation (phase de collaboration), avec par la suite une série d'échanges, d'explications et d'avis lorsque chaque professeur prend la parole pour informer les autres d'un problème dans sa classe. (phases d'échange et de discussion)

Type 3: Conférences de Presse

Avec ce type, il s'agit de la présentation d'un projet développé par chaque groupe dans une phase de préparation.

Exemple: Développement d'un nouveau journal ou autre produit comme une boisson: première phase: collaboration en groupes pour rechercher les idées, organiser la présentation; phase secondaire: présentation aux autres groupes (phase d'échange) qui peuvent poser des questions (discussion).

Autres exemples: Groupe de jeunes qui veulent transformer une vieille maison en club de jeunesse; ou les survivants d'un accident d'avion dans le désert qui expliquent leurs aventures.

C'est peut-être le type le plus créatif du point de vue des étudiants et le plus facile à évaluer parce que chaque groupe l'un après l'autre fait une présentation à la classe entière (phase d'échange). Mais d'un autre côté, cette méthode de présentation formelle peut réduire les autres à un rôle d'auditeurs, si on n'a pas donné d'instructions pour préparer les questions à poser dans le but d'amorcer une discussion intéressante.

Type 4: Réunions publiques contradictoires:

Les «simulations», les «jeux de rôles» classiques: un festival de musique pop: pour ou contre? une autoroute, pour ou contre? Chaque groupe représente des intérêts différents: préparation des cas (collaboration); explication des arguments (échange); puis débats, parfois confrontations (phase de discussion).

Cette sorte d'activité marche mieux si la réunion finale se fait en groupes de 6 à 8 étudiants seulement.

Type 5: Les entretiens

Les exemples classiques sont un entretien d'embauche avec deux ou trois étudiants qui posent des questions au candidat, un entretien avec une pop star, ou avec un réalisateur de films. Normalement, l'activité consiste seulement en une phase de préparation (collaboration) et un entretien (échange),

mais si les journalistes/employeurs abordent un problème avec la pop star/candidat qui doit s'exprimer spontanément, on peut arriver à une discussion.

Organisation:

Le but est d'assurer une vraie interaction entre les étudiants dans laquelle tout le monde peut parler, et c'est pour cette raison que le professeur prend du recul et reste observateur. Si au lieu d'être dominée par le professeur, la classe est dominée par un étudiant, cela revient au même - car celui-ci a tendance à manquer de la sensibilité qu'on attend d'un vrai professeur. Il y a un certain danger de reproduire ce qu'on veut éviter: une salle de classe formelle et centralisée où personne n'a le droit à la parole!

Voilà le résultat le plus tangible de l'analyse faite en 1986 dont j'ai parlé (NORTH 1986):

Grand groupe + structure de séance formelle

+ jeu de rôles

+ président dominant

= discours stérile dans lequel la plupart des étudiants garde le silence et où ceux qui ont l'habitude de parler en classe, prennent la parole.

C'est un argument contre l'utilisation des scénarios du Type 4 (réunion publique simulée, sous la direction d'un «président») en groupe de classe - ou alors il faut choisir soigneusement un président qui puisse jouer un rôle d'animateur plutôt qu'un rôle autoritaire. On peut expliquer ce rôle pendant la phase de collaboration pour que l'engagement soit vrai et l'interaction animée.

Deuxième précaution: Si le professeur prend du recul, et si on a évité qu'un «président» ne domine, c'est l'activité en soi qui entraîne la production linguistique et qui définit les limites de ce qu'on peut dire. Avec une tâche mal choisie, surtout avec une tâche trop simple, les étudiants ne peuvent probablement pas montrer leur vraie compétence. Dans ce cas, il est difficile d'arriver à une évaluation sûre parce que on a l'impression que la classe ou que certains étudiants, montreraient plus de compétence s'ils étaient confrontés à une tâche plus exigeante.

Autrement dit, pour être positif: le système sert à évaluer non seulement les étudiants, mais aussi les exigences proposées par chaque activité, - et tente d'aider le professeur à découvrir s'il sous-estime ou non les capacités de sa classe.

L'évaluation

La méthode d'évaluation suit les trois phases d'activités (collaboration, échange et discussion), avec une quatrième étape de négociation entre les deux examinateurs. L'idée première était que le professeur responsable de la classe fasse une évaluation formative, en notant les fautes et les difficultés (comme d'habitude), tandis qu'une autre personne (en général celle qui a la responsabilité de toutes les classes d'un stage) ferait une évaluation sommative avec des critères spécifiques. Mais, aujourd'hui on a tendance à utiliser en général les deux personnes à la fois pour l'évaluation sommative.

Les quatre étapes de l'évaluation

L'Impression Générale: Dans un premier temps, pendant la phase de collaboration, les examinateurs observent individuellement les étudiants, à partir de leur échelle d'évaluation pour l'interaction orale, et décident du niveau que présente la classe. En général, ce niveau occupe deux ou trois points sur l'échelle. Dans un deuxième temps ils observent chaque étudiant et – toujours en consultant leur échelle de compétences – lui donnent, rapidement, une note d'impression générale. Cette première phase prend à peu près 15 minutes.

L'Analyse: Quand la phase d'échange commence, les examinateurs font une analyse des performances de chaque étudiant. On utilise pour cela les critères, qu'on appelle en anglais RANGE, ACCURACY, DELIVERY, INTERACTION R; A; D; I; et qui avec la note Overall a donné le nom de RADIO au système.

RANGE: c'est à dire l'étendue des connaissances montrées par l'apprenant: la richesse de l'expression, le choix du vocabulaire, la complexité des tournures de phrases et la grammaire.

ACCURACY: c'est à dire la gravité des fautes grammaticales, mais aussi la justesse de l'expression et des mots par rapport à l'intention du discours.

DELIVERY: la qualité d'élocution de la performance, la facilité de se faire comprendre, les difficultés de prononciation, les hésitations, etc.

INTERACTION: les stratégies de discours, la faculté d'intervenir, de laisser la parole, d'interpeller et de se référer à ce qu'ont dit les autres.

Et **OVERALL:** l'évaluation globale: En bref, après cette analyse, quelle note générale peut-on donner, qui corresponde à ces quatre critères?

Voici un exemple:

Class	Date	Level norm						
17	4.5.88		1	2	3	4	5	6
			7	8	9	10		
Student	R	A	D	I	O	Grade		
	Range	Accuracy	Delivery	Inter-action	Overall	Im-pression	Final	Moderat
Kjell	9+	9	9	0+	9+	9	9+	
Cony	8/9	8	9	9	8+	9	8+	
Toos	7	7	7+	7+	7+	7	7+	
Olivier	7	7	7	7	7	7/8	7	
Andrea	7/8	8?	8?	8?		8	8	
Margarita	7	7-	8	8	7+	8	7+	
Claudia	8	8/9	8/9	9	8+	9	8+	
Lars	10	9+	10	9	9/10	10	10	
Rolf	7+	7+	7/8	7/8	7+	8	7+	
Helene	9	8/9	9	8/9	9-	8	8/9	
Ursula	7(8)	7+	7	8	7/8	7/8	7+	
Carlo	9	8	10	9	9	9+	9	

À gauche on a le numéro de la classe, la date, et à droite, le niveau général de la classe – c'est la première décision. Ici, il s'agit d'une classe de niveau supérieur, et on pense que le niveau des étudiants oscille entre 8 et 9.

À droite, on peut voir la note d'impression générale pour chaque étudiant avant l'analyse, et à gauche les notes données pendant la phase d'échange dans laquelle l'examineur peut se concentrer sur l'étudiant qui – à ce moment-là – a la parole. Il faut dire que c'est quelquefois difficile de donner une note à chaque étudiant pour chaque critère, comme c'est le cas ici avec Andrea.

La Comparaison: La troisième étape est de comparer la note d'impression initiale, et la note de l'analyse. Si les deux notes sont semblables cela suffit. Mais si, comme dans la plupart des cas, il y a une différence, on peut prendre le temps de réfléchir. Parfois on arrive à une décision rapide, mais en général, on peut se concentrer sur les étudiants dont on n'est pas certain pendant la phase de discussion, – pour pouvoir décider plus facilement. On fait cela en comparant les notes données aux autres étudiants. Dans ce cas précis, à la fin, on n'a pas pu décider si Hélène méritait 8, 8+ ou 9, mais pour les autres, ça allait.

La Négociation: La dernière étape est de discuter les notes avec l'autre examinateur. Deux appréciations donnent un résultat beaucoup plus fiable qu'une seule évaluation.

En résumé, on est parti d'une appréciation de la classe en tant que groupe pour arriver à une évaluation concertée pour chaque étudiant grâce à une méthode en quatre étapes, comportant les appréciations de deux professeurs différents, – l'un qui connaît la classe et l'autre qui connaît bien le système et les niveaux de toutes les classes. C'est donc un système transparent et relativement économique, intégré dans le programme, bien apprécié par les étudiants et qui a un effet positif sur l'enseignement. C'est à dire que c'est un système qui a une validité évidente, mais est-ce fiable?

Comme je l'ai dit, au début des années 80, on s'interrogeait pour savoir si «l'évaluation communicative», était «l'évaluation *de* la communication», ou «l'évaluation *pour* la communication». S'agissait-il d'évaluer un échantillonnage d'activités communicatives, ou d'évaluer la capacité linguistique dont on a besoin pour communiquer? Si l'évaluation se borne à tester une seule activité (même si elle comporte plusieurs aspects), peut-on pour cela en déduire une compétence générale à communiquer, dans n'importe quelle situation qui pourrait se présenter? Et si l'évaluation se borne à tester la connaissance théorique de l'outil de communication (la grammaire, le vocabulaire, les tournures de phrases), peut-on pour cela conclure à la capacité pratique d'utiliser cet outil dans toutes les situations communicatives qui se présentent?

Aujourd'hui, la plupart des linguistes appliqués anglais sont de l'avis de Pauline REA-DICKENS pour dire que les tests directs sont nécessaires, mais pas forcément suffisants: une évaluation précise des connaissances est également indispensable pendant le processus d'apprentissage pour prouver l'assimilation du contenu du stage, donner des conseils aux étudiants et fournir une orientation pour la planification du cours (REA-DICKENS, 1991).

Au lieu de faire un compromis entre les tests directs et indirects, nous avons développé à l'Eurocentre de Bournemouth un modèle binaire. La note qu'on donne dans cette activité est complétée pour établir la note finale par un «C-Test», qu'on peut utiliser pour vérifier les notes données par les professeurs à la fin du cours pour chaque apprentissage spécifique (compréhension orale et écrite, expression orale et écrite).

Un «C-Test» est une espèce de test «Cloze» dans lequel on supprime la dernière moitié d'un mot sur deux dans une série de textes courts. C'est une technique simple mais fiable d'origine allemande qui évite certains problèmes associés au Cloze simple. (KLEIN-BRALEY & RAATZ 1985)

Les deux tests offrent une évaluation complémentaire: L'activité de communication offre une évaluation subjective et directe de la perfor-

mance active à l'oral, et donne une bonne garantie de validité; on évalue la communication en relation avec l'échelle de compétences, et les critères spécifiques. Le C-Test donne un contrôle des connaissances passives – ou compétences théoriques si l'on veut – qui est objectif, indirect, et écrit – pour donner une garantie supplémentaire de fiabilité.

Pour établir si cette note finale a un tel degré de validité et de fiabilité, j'ai fait des expériences à Noël 87 et à Noël 88 avec 45 et 48 étudiants. Les deux fois la corrélation entre la note finale et les résultats du Cambridge First Certificate était de plus de 0,7, ce qui n'est pas mal avec un système aussi économique.

Developpements en cours:

Mais nous ne sommes pas tout à fait satisfaits du C-test. Il prend peu de temps (20-30 mins), est très facile à corriger, fiable – mais que peut-on faire avec le résultat? Il ne donne pas d'information détaillée à l'étudiant, pas de «feed-back». Il apporte une cohérence au système, mais apporte-il une transparence?...

Pour l'anglais, nous sommes actuellement en train de mettre au point un nouveau test qui remplacera les C-tests et qui s'appelle «Itembanker». Comme son nom l'indique c'est «une banque» de 1,000 questions de plusieurs types, et pour tous les niveaux de compétence, sur l'outil de communication (la grammaire, le vocabulaire, la cohérence et cohésion du texte, les tournures de phrases etc.) (JONES, à paraître). Ces points linguistiques dans cette «banque» de questions, découlent directement ou indirectement des définitions de l'échelle de compétences, par le biais d'une analyse des tâches communicatives impliquées par ces définitions et des structures nécessaires pour les réaliser. Au niveau élémentaire, ce lien est clair et direct; au niveau supérieur on a une liste de tâches et une liste de structures, mais il faut accepter que les liens entre les deux ne soient pas toujours évidents.

Les questions dans cette banque sont étalonnées entre elles, en relation avec l'échelle des neuf niveaux, grâce à une méthode américaine de statistiques, qu'on appelle en anglais «Item Response Theory» ou «Latent Trait Measurement» (HENNING 1987). Avec Itembanker, le professeur peut compléter l'évaluation à l'oral, pour laquelle il a choisi l'activité communicative la plus conforme à son programme, par un test qui lui fournit des questions sur les points de grammaire et de vocabulaire qu'il veut, au niveau qu'il veut.

Le système d'évaluation à l'oral «RADIO» se complète plus efficacement avec Itembanker qu'avec les C-tests:

Activité de communication	Itembanker
Subjective	Objectif
Directe	Indirect
Critères généraux	Points spécifiques
Orale	Ecrit
Active	Passif
Performances	Connaissances
Validité	Fiabilité

De cette façon on est en mesure de situer chaque étudiant sur l'échelle de compétences par rapport à ses performances, ou ses connaissances – ou pour la moyenne des deux, en obtenant une note qui est encore plus fiable que l'évaluation séparée des deux. En même temps, on est en mesure de fournir, en cours de stage, des renseignements précis sur les performances dans les deux tests (critères de communication, points spécifiques de grammaire) et d'aider ainsi les étudiants à améliorer leurs résultats dans les domaines de leur choix, avant de repartir chez eux. Ce système devrait être mis en place en Angleterre dès octobre 1992.

Eurocentres
Seestrasse 247
CH-8000 Zürich

BRIAN NORTH

Références

- ALDERSON, J. C. and HUGHES, A. eds.: «*Issues in Language Testing*», ELT Document 111, The British Council 1981
- HENNING, G.: Chapter 8 «Latent Trait Measurement» and chapter 9 «Item Banking» in «*A Guide to Language Testing: Development, Evaluation, Research*» Newbury House, Cambridge Massachusetts, 1987
- HUGHES, A. and PORTER, D. eds.: «*Current Developments in Language Testing*» Academic Press, 1983
- KLEIN-BRALEY, C. & RAATZ, E.: «C-Tests in der Praxis», in *Fremdsprachen und Hochschule*, AKS-Rundbrief 13/14, Bochum: Arbeitskreis Sprachenzentrum (AKS) 1985
- JONES, N.: «*An Item Bank to Measure Language Learning*» Phd thesis, University of Edinburgh, forthcoming.
- NORTH, B.: «*Activities for Continuous Communicative Assessment*», unpublished MA Phase 3 project, English Language Research Dept., University of Birmingham, 1986
- SINCLAIR, J.McH. & COULTHARD, M.: «*Teacher Talk*» Oxford University Press, 1982.
- REA-DICKENS, P.: Colloquium reply to Harrison's paper: «Language Assessment as Theatre: ten years of communicative testing», in ALDERSON, J. C. and NORTH, B. eds. «*Language Testing in the 1990s*» Modern English Publications, British Council, Macmillan, 1991