



ASp
La revue du GERAS

34 | 2001
Actes de la journée d'étude sur l'évaluation

La nécessaire interaction entre évaluation et processus d'apprentissage en langues

Sylvia Sommer



Édition électronique

URL : <https://journals.openedition.org/asp/1706>

DOI : 10.4000/asp.1706

ISBN : 978-2-8218-0386-2

ISSN : 2108-6354

Éditeur

Groupe d'étude et de recherche en anglais de spécialité

Édition imprimée

Date de publication : 1 décembre 2001

Pagination : 35-41

ISSN : 1246-8185

Référence électronique

Sylvia Sommer, « La nécessaire interaction entre évaluation et processus d'apprentissage en langues », *ASp* [En ligne], 34 | 2001, mis en ligne le 21 septembre 2010, consulté le 23 novembre 2024.

URL : <http://journals.openedition.org/asp/1706> ; DOI : <https://doi.org/10.4000/asp.1706>

Ce document a été généré automatiquement le 23 novembre 2024.



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

La nécessaire interaction entre évaluation et processus d'apprentissage en langues

Sylvia Sommer

- 1 Bien qu'activement engagée dans la réflexion théorique sur l'évaluation par des travaux de recherche qui ont abouti au doctorat, et très impliquée dans la pratique de l'évaluation par de récentes fonctions de formateur d'examineurs et de responsable de centre d'examens pour le dispositif des examens de l'*University of Cambridge Local Examinations Syndicate*, c'est surtout en tant qu'enseignante face à l'évaluation en langues que je m'adresse à vous ici. Avant de rentrer dans le vif du sujet, je voudrais rappeler quelques notions indispensables relatives à l'évaluation.
- 2 L'évaluation, comme nous le savons, est intrinsèquement liée à tout le processus de l'enseignement. Bien que H. Douglas Brown (1987 : 7) nous rassure en précisant qu'elle « est tout simplement un moyen de mesurer la compétence dans un domaine donné », en langues, elle sert à formuler un jugement à partir d'un échantillon de comportements langagiers. Ces comportements peuvent faire l'objet de différents types d'évaluation, selon l'étape à laquelle ils sont soumis à l'acte d'évaluation. Car notre enseignement est jalonné d'actes d'évaluation. Cependant, l'évaluation ne doit pas être un but en elle-même, mais doit toujours s'articuler autour du contenu pédagogique et se situer par rapport aux dispositions pédagogiques mises en place. Le didacticien en langues se doit de faire la différence entre la perspective psychométrique et la perspective édumétrique. L'objectif de l'évaluation prime. « Sommairement », nous dirons que nous utilisons les deux catégories d'évaluation suivantes :
 - l'évaluation pronostique ou formative, comme les tests d'aptitude ou de niveau,
 - l'évaluation diagnostique ou sommative, les tests de contrôle ou de progrès.
- 3 La journée d'étude met l'accent sur la certification. L'intérêt se porte donc sur l'évaluation sommative, comme l'explique Bloom :

Alors que l'évaluation formative revêt en principe un caractère privé (sorte de dialogue particulier entre l'éducateur et son élève), l'évaluation sommative est

publique : classement éventuel des élèves entre eux, communication des résultats aux parents par un bulletin scolaire, attribution d'un certificat ou d'un diplôme.
(1979 : 115)

- 4 En évaluation, nous sommes obligés de prendre en compte certains critères posés sur les fondations scientifiques de la validité et de la fiabilité. Autrement dit, traditionnellement, sur le plan psychométrique, toute évaluation devrait mesurer, schématiquement, ce pour quoi elle a été mise au point. Cette validité signifie que l'outil d'évaluation doit correspondre au matériel, au cours ou au programme considéré (validité du contenu). L'outil d'évaluation doit être en accord avec la théorie linguistique de la langue visée et de la linguistique acquisitionnelle (validité du construit). L'outil d'évaluation doit fournir les mêmes résultats si l'évaluation est répétée dans les mêmes conditions (fiabilité). Nous pouvons donc dire que, idéalement, toute évaluation doit comporter une validité de construit et de contenu par rapport aux compétences visées, et être fiable.
- 5 Ces trois critères sont à la base de la réflexion scientifique sur l'évaluation. Cependant, pour valider un outil d'évaluation, il faut également tenir compte d'un autre critère : **l'effet rétroactif**, ou l'impact de l'évaluation sur le processus d'apprentissage en langues. Ce critère, domaine de recherche actuellement peu exploité en didactique des langues, est au cœur de la problématique posée dans le titre. Avant de l'exposer, examinons rapidement les trois critères précités.
- 6 Celui de la fiabilité est probablement le mieux connu, voire reconnu, car il renvoie à la confiance que nous portons à l'évaluation, c'est-à-dire à la réponse que nous apportons à la question : jusqu'à quel point les résultats sont-ils exempts d'erreurs de mesure ? En effet, nous croyons que l'évaluation est fiable parce que les paramètres de constance et de sûreté sont présents, et souvent nous reléguons cette notion à la seule correction objective des épreuves, de type QCM. Cependant, sur le plan scientifique, la fiabilité comporte plusieurs aspects. Elle concerne autant la conception des formats utilisés, la prestation des examinateurs, que les barèmes et les échelles de correction et de notation des épreuves. Nous reviendrons sur la fiabilité et les difficultés causées par la poursuite acharnée de celle-ci.
- 7 Nous allons maintenant expliciter la validité du construit et celle du contenu. L'évaluation en langues, pour être valide en ce qui concerne le construit, doit s'appuyer sur les fondements théoriques de ce qu'est la langue et le langage. Aujourd'hui nous optons pour le paradigme communicatif, fondé sur une définition consensuelle de nos savoirs, savoir-faire et savoir-être d'une langue. Le cadre théorique et analytique prend appui sur les apports de la linguistique, la sociolinguistique, la psycholinguistique et la pragmatique pour fournir le construit de base : les compétences de communication.
- 8 Pour être valide au niveau du contenu, l'évaluation doit inclure une gamme d'items, de tâches, qui représentent le domaine cible de la performance de l'utilisateur de la langue : le format de l'évaluation représente une constellation de tâches langagières que l'on est censé retrouver en dehors du contexte de l'évaluation, voire même de l'apprentissage. La procédure et le format comptent autant que les compétences mesurées. Il faut donc des tâches comportant un défi lancé à l'emploi interactif de la langue, tout comme le font les situations de performance langagière dans la vie réelle. D'où la nécessité d'inclure la compétence orale comme composante dans un dispositif d'évaluation.

- 9 Ces trois critères sont à la base de la réflexion scientifique sur l'évaluation. Nous sommes cependant obligés de capituler devant d'autres difficultés, tout à fait indépendantes de cette nécessité de consensus de théories et de définitions quant au construit et au contenu de l'évaluation. Il en résulte des choix et des compromis qui font que l'outil d'évaluation mis en place n'est jamais à 100 % conforme aux critères psychométriques.
- 10 Je fais allusion ici à toutes les difficultés qui vont mettre en danger un critère que nous n'avons pas encore cité, mais qui joue un rôle capital : la faisabilité. Ces difficultés se rencontrent au niveau du cadre général dans lequel s'inscrit l'évaluation. Des considérations pratiques, institutionnelles, politiques et surtout économiques, gênent la mise en place ou le maintien d'un dispositif d'évaluation. Les contraintes des différentes forces en jeu limitent le nombre de critères théoriques souhaitables qui pourraient effectivement être respectés. Très souvent, c'est la réticence devant de nouvelles façons de tester, plus lourdes et plus coûteuses sur le plan logistique, qui freine leur acceptation. Un exemple en a été la décision de supprimer la composante orale prévue dans le test TOEFL à l'époque de sa première mise sur le marché. Malgré l'exhortation des psychométriciens de l'époque, tels que J. B. Carroll, qui souhaitaient l'inclusion de cette composante en vue de renforcer la validité du dispositif, l'institution a mis en avant des paramètres purement économiques et commerciaux. Comme le rappellent Alderson & Clapham :
- Les concepteurs de tests prennent nécessairement en compte les autres variables associées à l'objectif de l'évaluation, telles que le public, la logistique impliquée lors de la conception et de l'administration, et qui ne font pas du tout partie des préoccupations des théories linguistiques et psychométriques. (1992 : 149)
- 11 Mais même sur le plan psychométrique des obstacles persistent. Un exemple en est la poursuite acharnée de la fiabilité que j'ai déjà évoquée, fortement influencée par des considérations docimologiques, visant une objectivité maximale en ce qui concerne la correction. Cette attention accordée à l'obtention de la fiabilité, domaine de recherche très prometteur quant à la validation objective de l'évaluation, devient un domaine réducteur quand il s'agit de la réalisation d'une évaluation holistique comportant une part de subjectivité, que préfère le didacticien en langues. Au nom de la fiabilité, nous continuons à utiliser des outils tels que le TOEFL et le TOEIC, même en l'absence de validité du contenu et du construit. À quoi sert la fiabilité en dehors du fait qu'elle constitue un bon argument de vente lors du marketing du produit ?
- 12 Il subsiste d'autres problèmes liés à la fiabilité, tels que les variations de la correction entre examinateurs, dans le temps de correction d'un même examinateur, etc., mais aujourd'hui on parvient à trouver des solutions ou plutôt des améliorations grâce aux apports de la recherche pluridisciplinaire. Il est possible d'améliorer la fiabilité par le recours à la correction à l'aide d'échelles incorporant les critères linguistiques et pragmatiques et dégagant un profil langagier qui allie l'objectivité de la notation et la validité du construit communicatif interactif. On peut également faire appel à la formation des examinateurs, étayée par un contrôle de qualité de leur performance à l'écrit et à l'oral.
- 13 Mais revenons aux deux critères de validité, si différenciés à l'époque de l'évaluation psychométrique. Aujourd'hui, nous préférons tenir compte de l'interaction entre ces deux concepts, surtout si l'on évoque l'effet rétroactif de l'évaluation sur le processus de l'enseignement, l'autre validité, essentielle par son renforcement du lien

curriculaire entre enseignement et évaluation. En ce qui concerne cette interaction, je citerai Weir :

Étant donné que les enseignants dans un cadre communicatif essaient d'équiper leurs étudiants des compétences langagières jugées être en adéquation avec leurs besoins actuels ou futurs, et étant donné que les outils d'évaluation sont censés être conçus pour en être le reflet, plus cet outil d'évaluation ressemble aux dispositions pédagogiques qui les préparent, plus il est probable que l'outil est valide au niveau de son construit. (1990 : 27)

- 14 J'ajouterai que si l'outil d'évaluation est effectivement valide sur le plan de son construit, l'enseignement deviendra lui aussi valide par rapport à ce construit de la langue cible. Nous retrouvons ici le concept de l'effet rétroactif de l'évaluation sur l'enseignement, le « *washback effect* », terme consacré en anglais dans le milieu de la didactique des langues. En renvoyant à cette validité, je renoue avec la problématique de mon titre.
- 15 Dès lors, les critères de base se trouvent modifiés. Comme l'écrivait Swain en 1993, l'évaluation communicative est fondée sur nos savoirs, savoir-faire et savoir-être d'une langue ou sur les compétences de communication, les actes de parole, actes interdépendants et interactifs. Ainsi, les critères pour l'évaluation en langues deviennent :
 - un construit communicatif,
 - un contenu intégratif et interactif,
 - une recherche de la meilleure performance de la part du candidat, ou la notion de l'évaluation positive remplaçant la notation négative, le « syndrome de la dictée de l'institutrice », d'après Clay,
 - l'effet rétroactif positif sur l'enseignement des langues.
- 16 Swain ajoutait que la complexité cognitive, l'effet rétroactif, l'équité, la qualité et l'étendue du contenu, parmi d'autres critères, constituent le fondement de l'évaluation en langues.
- 17 Alors que voulons-nous dire par l'expression « *washback* » ? Dans les ouvrages spécialisés, le phénomène est défini comme l'effet direct ou indirect d'un outil d'évaluation sur la pédagogie, lequel produit un effet sur les attitudes et le comportement de tous les acteurs de l'apprentissage, les enseignants et les apprenants, et sur les supports choisis. Cette définition traditionnelle du phénomène fige l'effet dans le temps. Le terme « impact » utilisé par les didacticiens en France conforte cette idée. Il comporte en outre une connotation de conflit, au détriment de la négociation. Il serait préférable d'introduire une notion de flux constant, de va-et-vient, entre évaluation et processus d'apprentissage, d'où le choix dans mon titre de « la nécessaire interaction ». Il me semble en effet que le phénomène ressemble beaucoup plus à celui de la marée descendante, qui se retire en laissant derrière elle les traces de son passage, tout en emportant des grains de sable de la plage.
- 18 Je suis convaincue que l'évaluation non seulement joue mais aussi doit jouer un rôle important dans l'amélioration de la pédagogie et de l'apprentissage des langues. Ce rôle peut être « négatif » et avoir un « effet pervers » comme le signale Madaus (1990). Dans ce cas, le contenu pédagogique et le procédé éducatif se résument alors à la préparation à une certaine forme d'évaluation. Nous avons tous le souvenir des séances de bachotage pour les QCM et pour certains tests institutionnels. Mais l'effet peut devenir positif s'il s'agit d'une préparation à une évaluation dite communicative, étant donné

que l'enseignant s'efforce toujours d'ajuster l'enseignement au contenu et au format de l'outil d'évaluation. Si ce contenu, l'objet de l'évaluation, reflète une vision étendue des compétences de communication, alors l'outil d'évaluation produira forcément un effet positif sur l'enseignement, dans le sens où l'enseignement sera lui aussi fondé sur une approche communicative.

- 19 Il est courant aujourd'hui de penser que les cours de langues ont adopté l'approche communicative. Mais la méthodologie appliquée dépend presque entièrement de l'attitude de l'enseignant, de sa propre volonté ou aptitude à se plier aux changements ou, au contraire, de sa tendance à y résister. Elle dépend également des moyens donnés pour y parvenir. Pour qu'un processus d'enseignement soit modifié par un nouveau mode d'évaluation, il faudrait des directives explicites pour une préparation appropriée. D'autres facteurs interviennent, tels que le statut de l'outil d'évaluation dans un contexte donné, les ressources disponibles dans le contexte d'apprentissage et surtout la compétence de l'enseignant. Ce dernier facteur, pour nous enseignants, est le plus crucial. Sans la compréhension du fonctionnement de l'effet rétroactif, l'enseignant n'appliquera pas une nouvelle approche pédagogique. Le manque d'analyse critique de l'enseignant face à l'outil mis en place et une mauvaise compréhension de l'objectif et des moyens réduit considérablement l'effet. En revanche, il y a effet positif sur l'enseignement si l'enseignant est formé à la nouvelle méthodologie à mettre en place. Celle-ci sera alors en phase avec les nouvelles techniques et le contenu de l'évaluation. D'où la nécessité d'un encadrement afin d'établir le lien curriculaire entre l'évaluation et l'enseignement qui la précède.
- 20 Nous touchons, ici, à l'idée d'interaction, non seulement entre évaluation et enseignement, mais aussi entre enseignant et apprenant. Tout un travail de sensibilisation doit accompagner le processus d'apprentissage de l'évaluation, surtout l'évaluation sommative. Une interaction constante entre enseignant et apprenant doit s'instaurer pour que s'établisse une compréhension mutuelle quant à l'objectif de l'enseignement et de l'évaluation. Trois domaines de sensibilisation peuvent ainsi être définis :
- une sensibilisation au sens du langage et à l'approche métalinguistique, l'accent étant mis sur la communication, ses stratégies et ses enjeux,
 - une sensibilisation au processus d'apprentissage, voire d'acquisition, en langues,
 - une sensibilisation au format et au contenu de l'outil d'évaluation préparé avec une compréhension de la philosophie et de la culture évaluatrice.
- 21 Ainsi, le processus d'apprentissage s'appuiera sur les mêmes caractéristiques qu'une évaluation valide. Il sera communicatif, incorporant l'apprentissage des compétences linguistiques ainsi que des compétences pragmatiques. Les acteurs, enseignants et apprenants, seront face à l'activité humaine la plus complexe, la communication, un processus de production langagière subjective et non pas un produit objectif.
- 22 L'effet potentiel de l'évaluation sur la préparation pédagogique préalable nous interpelle fortement, en tant qu'enseignants. C'est une considération primordiale dans la mesure où l'évaluation doit refléter les théories de la didactique des langues, et être en mesure d'influencer notre enseignement. Certains auteurs, tels qu'Alderson et Wall (1993), restent sceptiques quant au phénomène, arguant d'un manque de vérification empirique. En effet, la recherche scientifique dans ce domaine reste embryonnaire mais, d'instinct, je suis convaincue qu'elle nous donnera un autre éclairage sur les comportements des acteurs dans le processus d'apprentissage face aux dispositions

méthodologiques mises en place pour la préparation d'une évaluation valide. Ce critère de validité qu'est l'effet rétroactif positif sur l'enseignement des langues reste incontournable pour un enseignant. Inversement, si une équipe d'enseignants opte pour l'apprentissage de telle ou telle compétence spécifique, il est indispensable que cette compétence figure parmi les compétences évaluées. De même, si l'on choisit de développer des compétences plus itemisées ou spécialisées, l'évaluation doit s'adapter pour n'évaluer que celles-ci et rien d'autre.

- 23 Je voudrais donc proposer que la formation des enseignants comporte un module consacré non seulement à l'évaluation, mais également à son effet sur l'enseignement. Il me semble tout aussi important pour un enseignant d'appréhender cette interaction que d'acquérir des techniques d'enseignement. Il nous faut une approche holistique, non seulement dans la façon de concevoir l'évaluation, mais aussi dans la formation même de l'enseignant. L'évaluation fera donc partie intégrante de la prestation pédagogique. Sinon, le processus d'apprentissage lui-même perd en validité. Pour conclure, j'oserai dire que « l'on ne badine pas avec l'évaluation ». Les enjeux sont trop importants. D'où la nécessité, voire l'obligation éthique, de bien choisir l'outil d'évaluation par rapport à l'objectif visé par l'acte d'évaluation. La question « pourquoi l'évaluation ? » est toujours à poser en premier lieu. S'il faut évaluer un profil langagier, il faut évaluer toutes les compétences. Si c'est la production orale ou écrite qui est à évaluer, la situation d'évaluation doit fournir la possibilité de fournir une production orale ou écrite, et il en va de même pour la compréhension de l'écrit et de l'écoute. En situation d'évaluation, les tâches à accomplir doivent être du même type que celles que l'on trouve hors situation d'évaluation. Les savoirs, savoir-faire et savoir-être sont ainsi tous mis en évaluation. Ceci permettra non seulement de respecter la langue, l'objet de l'évaluation, au moyen de critères scientifiques de validité et de fiabilité, mais aussi le sujet de l'évaluation, l'être humain, grâce à des critères d'équité et de justesse.
-

BIBLIOGRAPHIE

Alderson, J. C. & C. Clapham. 1992. « Applied linguistics and language testing: A case study of the ELTS Revision Project ». *Applied Linguistics* 13/2, 149-167.

Alderson, J. C. & D. Wall. 1993. « Does washback exist ? ». *Applied Linguistics* 14/2, 115-129.

Bloom, B.J. in G. De Landsheere. 1979. *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses universitaires de France.

Brown, H.D. 1987. *Principles of Language Learning and Teaching*. Regents, NJ : Prentice Hall.

Madaus, G. F. 1990. *Testing as a Social Technology*. Boston : Boston College.

Swain, M. 1993. « Second language testing and second language acquisition: is there a conflict with traditional psychometrics? ». *Language Testing* 10/2, 193-207.

Weir, C. 1990. *Communicative Language Testing*. Hemel Hempstead : Prentice Hall.

RÉSUMÉS

Traditionnellement la validation d'un bon outil d'évaluation repose sur les critères scientifiques de fiabilité et de validité. L'effet rétroactif est effectivement un critère de validité incontournable pour l'enseignant en langues face au choix de l'outil d'évaluation. Car si l'outil se fonde sur le construit de la compétence de communication, l'approche et les supports pédagogiques feront de même. Ainsi le processus d'apprentissage sera marqué d'autres considérations, telles que la sensibilisation au sens de la communication, du langage et de la langue.

In the language classroom one of the most important criteria for a teacher to use in order to establish the validity of a testing tool is to ascertain the degree to which the content and the format of the test or exam can produce a real positive effect on the teaching prior to the test. If the exam content assesses active communicative use of the foreign language in question, it would be logical that the teaching/learning process involves the same use. As a result, the choice of the testing tool becomes a primary issue in the curriculum itself, involving a more metalinguistic approach to communication, language and language learning.

INDEX

Mots-clés : critère de fiabilité, critère de validité, effet rétroactif, interaction, lien curriculaire

Keywords : curriculum, interaction, reliability criterion, retroactive effect, validity criterion

AUTEUR

SYLVIA SOMMER

Sylvia Sommer est maîtresse de conférences à l'Université Lyon 2 où elle s'occupe du secteur de la Formation continue en langues. Depuis trente ans, elle enseigne l'anglais dans l'enseignement supérieur, en formation initiale et continue, avec des missions en classe bilingue dans le secondaire. Lors de son travail de recherche doctorale, elle a analysé les modes d'évaluation de la compétence de communication à l'oral. Cette théorisation à partir de la pratique en didactique des langues a contribué à la conception et l'élaboration de la partie orale de l'outil d'évaluation, le DCL. En didactique des langues, elle porte un intérêt particulier au lien curriculaire entre évaluation et enseignement. Sylvia.Sommer@univ-lyon2.fr