

UNE COMPOSANTE INDISPENSABLE
DE
LA FORMATION DE L'ENSEIGNANT:
L'INITIATION A LA DOCIMOLOGIE

Par: Denise LOUANCHI
Chargée de cours à l'I.L.E.
Université d'Alger.

Bien qu'elle s'intègre dans le cadre des licences d'enseignement, la formation dispensée par l'Université aux étudiants en langues étrangères se contente trop souvent de doter le futur enseignant d'une maîtrise de la langue et d'un bagage culturel jugés suffisants pour qu'il assure à son tour la transmission des connaissances au niveau du secondaire.

La formation professionnelle de l'enseignant se réduit ainsi à un apprentissage strictement cognitif et la pédagogie prendra au mieux la forme de CM et de TD de didactique, excluant d'ailleurs une donnée capitale de la situation d'apprentissage : le destinataire réel, l'élève du secondaire.

Cet article se propose de montrer à quel point une telle conception de la profession enseignante est aveugle, voire irresponsable, et qu'il est urgent, au moment où une nouvelle licence de langues étrangères envisage de donner au futur enseignant une authentique formation professionnelle, de recenser et développer les capacités requises pour l'exercice de ce métier.

Parmi les apprentissages jusqu'alors négligés, il en est un qui retient tout particulièrement notre attention : l'initiation à l'évaluation scolaire. Le jeune licencié qui se voit attribuer ses premières divisions pédagogiques ne sait pas préparer un test évaluatif et encore moins apprécier les productions de ses élèves parce qu'en trois années de licence on n'a pas cru nécessaire de s'attarder sur cette tâche, comme si l'apprentissage des connaissances n'impliquait pas inmanquablement le contrôle de leur acquisition. Tout se passe comme si on laissait entendre au futur enseignant qu'il suffit qu'un cours soit bien fait par le maître pour qu'il soit acquis par les apprenants. Le contrôle des connaissances, auquel se trouvera vite confronté l'enseignant du secondaire, prend ainsi l'allure d'une routine à laquelle il est convenu de s'adonner en fin de trimestre afin d'attribuer une note aux élèves et une appréciation sur un bulletin. Et il est singulier de constater que le même individu qui, étudiant, savait si bien contester les notes reçues et exiger des justifications, devienne subitement si peu rigoureux lorsqu'à son tour il distribue les notes à ses propres élèves. La même pratique, aussi unique que péremptoire, se reconduit alors, prouvant une fois de plus -gloire à Voltaire!- que rien n'est plus vénérable qu'un vieil abus. Pourtant notes, bulletins trimestriels, livrets scolaires et examens ne sont-ils pas les critères sur la base desquels se décident le maintien ou l'exclusion d'un élève du giron scolaire ? L'évaluation scolaire n'est-elle pas ce qui conditionne la sélection scolaire et, par voie de conséquence, la promotion sociale ?

Qu'il y ait lieu, et moyens, de réfléchir sur l'évaluation scolaire, c'est pourtant ce qu'on ne peut plus ignorer depuis la naissance de la docimologie. A cette fin, nous voudrions relater ici une partie d'un travail de recherches entrepris avec les étudiants du semestre 6 de la licence de français, au cours de l'année universitaire 82 - 83.

Une première séance consiste à soumettre à chacun des 77 étudiants de semestre 6 un même lot de 10 copies qu'il avait pour tâche d'évaluer en leur attribuant notes et observations. Ces copies étaient des "essais" de français élaborés en temps limité par des élèves de 2^e AS-ST (Bilingues) d'un lycée d'Alger. Les étudiants n'avaient reçu aucun conseil préalable pour la correction et il leur était rigoureusement interdit de discuter entre eux des copies. Certains se trouvaient dans la situa-

tion de l'enseignant débutant qui reçoit son premier paquet de copies et ne sait pas comment corriger. D'autres, PEM en détachement pour la plupart, accomplissaient une tâche déjà familière.

Le tableau suivant donne pour chaque copie la note minimale et la note maximale qui lui ont été attribuées par les 77 évaluateurs, ainsi que l'écart entre ces deux notes.

COPIE N°	NOTE MINIMALE	NOTE MAXIMALE	ECART
1	5	16	11
2	5	13,5	8,5
3	4	15	11
4	4	15	11
5	5	15	10
6	4	13	9
7	5	16	11
8	3,5	15	11,5
9	4	14	10
10	5	15	10

(La notation se faisait sur 20)

Quand on pense que 34 des 77 évaluateurs avaient déjà plusieurs années de pratique professionnelle derrière eux et que les 77 devaient être bientôt professeurs de lycée, on ne peut s'empêcher de déplorer que l'avenir des jeunes lycéens dépende des notes scolaires.

Lors de la séance suivante, un débat entre étudiants tenta de dégager un certain nombre de critères communs d'évaluation ainsi qu'un barème. On pouvait en effet supposer que les divergences de notations venaient de ce que les 77 évaluateurs n'avaient pas utilisé les mêmes critères, les uns sanctionnant surtout les idées, l'argumentation ou le plan, les autres accordant plus d'importance à la rectitude de la langue. La discussion, fort animée, révéla qu'à quelques mois d'entrer dans la carrière professorale les futurs enseignants de français étaient

loin d'avoir une idée précise de la finalité de leur fonction-maîtres à penser ou professeurs de langue étrangère ? - et divergeaient sur ce qu'il convenait de sanctionner comme fautes "graves".

Après que les critères eurent été explicités et un barème commun admis ("admis", du moins accepté momentanément pour permettre à l'expérience de se poursuivre car aucun consensus ne se dégagait), une des 10 copies de la séance précédente fut de nouveau soumise à correction individuelle. Pour chacun des 4 groupes de TD, le nouvel écart enregistré entre la note minimale et la note maximale fut de 6, 6, 8 et 6 points ! La comparaison entre la 1ère et la 2ème note données par chaque étudiant montra par la suite que très peu d'étudiants avaient modifié leur première note (quoique celle-ci ne leur eût pas été communiquée lors de la 2ème correction). La fidélité à eux-mêmes semble leur avoir été plus contraignante que le respect des normes décidées en commun.

Comment alors espérer parvenir à une quelconque objectivité d'évaluation aux examens grâce aux réunions de coordination des notes si chacun s'en tient avant tout à ses critères personnels ? Faut-il dire avec H. Piéron que pour prédire la note d'un candidat à un examen, il vaut mieux connaître son examinateur que le candidat lui-même ?

Mais que sont ces critères personnels ? Chacun est-il en mesure de les expliciter ? Lors même qu'ils seraient explicités, est-il bien certain qu'ils s'avèreraient servir de bases exclusives au jugement évaluatif ? Le détail de la 1ère séance d'évaluation apporte quelques éclaircissements sur cette question car, outre les divergences de notes et d'observations, elle se proposait encore de montrer les processus inconscients à l'oeuvre dans l'acte d'évaluation.

Précisons l'expérience :

Les 77 évaluateurs étaient à leur insu répartis en 5 groupes expérimentaux (à distinguer des groupes de TD). 4 groupes expérimentaient des variables informationnelles, le 5ème les effets dits séquentiels. Le principe général de l'expérience peut se résumer en une théorie des "attentes" de l'évaluateur, inspirée de l'ouvrage de M. Noizet

et Caverni. *(1) Suivons ces auteurs : "Les mêmes copies sont évaluées différemment lorsqu'on donne aux évaluateurs des informations différentes sur les élèves censés les avoir rédigées". (2)

Des informations, apportées préalablement à la correction, opèrent une prévention du jugement et créent des attentes chez l'évaluateur qui inconsciemment sélectionnera les informations réellement contenues dans la copie pour retenir principalement celles qui sont conformes à ses attentes et négliger celles qui sont discordantes. Pour reprendre les termes de M. Noizet et Caverni, l'évaluateur chercherait à établir une "consonance" entre les informations anticipées -apportées par l'expérimentateur- et les informations présentes -caractéristiques de la copie- par une "assimilation" de l'actuel à l'antérieur.

Plusieurs types d'informations étaient expérimentés :

- 1er groupe expérimental, variable A : le profit scolaire de l'élève. La même copie était présentée, accompagnée du profil scolaire de son auteur... qui différait d'un évaluateur à l'autre. L'élève était censé être tantôt faible, tantôt fort, tantôt en progression, tantôt en régression. Un trait de comportement commun aux évaluateurs devait se manifester à savoir que, quand les notes précédentes de l'élève sont faibles, l'évaluateur a tendance à sous-estimer la copie présente, tandis qu'au contraire il la surestime quand elles sont fortes. Ce phénomène fut sensible pour la copie n° 4. Qu'on en juge par le tableau suivant :

MOYENNE DES NOTES INDUITES PAR LE PROFIL CONSTRUIT	MOYENNE DES NOTES DONNEES PAR LES EVALUEURS
7	(6
9	(7,2
10,5	(7,8
13	(10,4

*(1) "Psychologie de l'évaluation scolaire" Noizet et Caverni. PUF, 1978.

(2) Idem. p. 97.

On voit certes qu'il ne s'agit pas d'un conditionnement mécanique mais l'effet de suggestion est net pour cette copie : la moyenne des notes données par les évaluateurs a progressé parallèlement à la moyenne des notes induites par l'expérimentateur. Précisons que le phénomène a des limites : l'évaluateur refuse de prendre en considération l'information lorsque le profil scolaire construit s'écarte trop de la copie présente. Ainsi pour les copies trop faibles les évaluateurs ne tiennent pas compte des profils forts.

- 2è groupe expérimental, variable B : pour chaque copie on précisait les notes que 2 évaluateurs lui avaient précédemment attribuées. Ces notes étaient construites et les évaluateurs fictifs. Cette variable s'est révélée très opérante pour les 5 copies sur lesquelles elle a été testée. En effet, la moyenne des notes données par les évaluateurs a varié comme suit :

COPIE N°	MOYENNE INDUITE	MOYENNE OBTENUE
1	$\left\{ \begin{array}{l} - 7 \dots\dots\dots) \\ - 8 \dots\dots\dots) \\ - 11 \dots\dots\dots) \end{array} \right.$	$\left\{ \begin{array}{l} 8,7 \\ 9,4 \\ 11,6 \end{array} \right.$
2	$\left\{ \begin{array}{l} - 6 \dots\dots\dots) \\ - 8 \dots\dots\dots) \\ - 13,5 \dots\dots\dots) \end{array} \right.$	$\left\{ \begin{array}{l} 8,4 \\ 8,6 \\ 10,3 \end{array} \right.$
3	$\left\{ \begin{array}{l} - 7 \dots\dots\dots) \\ - 10,4 \dots\dots\dots) \\ - 14 \dots\dots\dots) \end{array} \right.$	$\left\{ \begin{array}{l} 9,9 \\ 10,9 \\ 12,06 \end{array} \right.$
4	$\left\{ \begin{array}{l} - 6,5 \dots\dots\dots) \\ - 11 \dots\dots\dots) \\ - 12,5 \dots\dots\dots) \end{array} \right.$	$\left\{ \begin{array}{l} 5,6 \\ 8,57 \\ 10 \end{array} \right.$
5	$\left\{ \begin{array}{l} - 7 \dots\dots\dots) \\ - 10,5 \dots\dots\dots) \\ - 13,5 \dots\dots\dots) \end{array} \right.$	$\left\{ \begin{array}{l} 7,6 \\ 10,75 \\ 12,5 \end{array} \right.$

Les étudiants ont dans l'ensemble entériné les notes des évaluateurs fictifs, comme s'ils ne voulaient pas contester le jugement de gens "expérimentés".

- 3è groupe expérimental, variable C : informations sur la filière (Math, Sciences, Lettres), la série (normale ou transitoire), le niveau (1ère AS, 2è AS, 3è AS). Là encore les informations, construites, variaient d'un évaluateur à l'autre. Il s'agissait de préciser si les évaluateurs modifient leurs normes de correction en fonction du type d'élèves. En effet, on ne peut évaluer de la même manière une copie de 3è AS et une copie de 1ère AS, une copie de Lettres ou une copie de Sciences. Or il ne semble pas que les étudiants évaluateurs aient ressenti la nécessité d'une telle adaptation des normes aux types d'élèves. Cette raideur des normes est regrettable.

- 4è groupe expérimental, variable D : l'établissement scolaire d'origine. On se proposait de savoir si la renommée, bonne ou mauvaise, de certains lycées influe sur le jugement de l'évaluateur. Ainsi la même copie fut évaluée en moyenne 10,45 lorsqu'elle était attribuée à un élève du lycée de Bou Ismail, établissement rural, tandis que lorsqu'elle était censée appartenir à un lycéen d'Amara Rachid, établissement de Ben Aknoun assez bien "coté", elle vit sa moyenne de notation s'élever à 12,9. Ce type de variable et d'information peut agir dans des sens opposés. Ainsi un évaluateur avoua par la suite avoir "remonté" la note d'un élève du lycée Et Thaëlibya parce qu'il savait pour y avoir été scolarisé lui-même, que ce lycée d'Hussein Dey accueille une population scolaire défavorisée socialement et culturellement. D'autres évaluateurs ont, à l'inverse, "retrouvé" le mauvais niveau attendu et évalué en conséquence.

- 5è groupe expérimental, variable E : le groupe avait pour but d'étudier les effets de contraste dus à la séquence. Cet effet joue diversement. Dans l'expérience présente on se limitait à l'étude du rôle de la position des copies les unes par rapport aux autres. Le phénomène est connu : une copie "moyenne" par exemple sera notée différemment selon qu'elle suit une copie "forte" ou une copie "faible" (sous ou sur-évaluation). On voit qu'il ne s'agit plus d'informations apportées de l'extérieur par l'expérimentateur. La manipulation est ici l'oeuvre des copies elles mêmes : chaque nouvelle copie corrigée modifie les normes de correction inconscientes de l'évaluateur et conditionne la lecture de la suivante.

L'effet séquentiel a faiblement agi dans l'expérience, neutralisé par des variables "parasites" dues aux conditions défavorables de l'expérience (nombre insuffisant de copies soumises à correction, effectif limité d'évaluateurs, copies de niveaux trop voisins, présence dans le groupe d'évaluateurs de quelques éléments recourant aux portions extrêmes des notes etc...). Donnons cependant deux exemples. En changeant de position dans la série, la copie 8 a vu sa moyenne d'évaluation varier de 7,6 à 8,44 et la copie 10 est passée d'une moyenne de 9 à une moyenne de 10,5.

Les effets séquentiels dus à la fatigue de l'évaluateur en raison d'une longue série de copies n'ont pu être étudiés dans cette expérience.

Ainsi s'expliqueraient -en partie- les écarts entre évaluateurs: les informations données (par l'expérimentateur pour les variables A, B, C, D, ou par la copie précédente pour la variable E) avaient pour but d'orienter leur lecture et leur jugement. Parce que les informations sur une même copie variaient d'un évaluateur à l'autre, l'appréciation de cette copie était déterminée à varier elle aussi.

Se pose alors le problème de la généralisation des résultats de l'expérience.

En TD, les évaluateurs avaient été volontairement manipulés à leur insu. Les conditions de correction des travaux scolaires présentent-elles une analogie avec les conditions artificielles de TD? L'enseignant en situation sera-t-il aussi peu objectif avec les copies de ses élèves que les étudiants-évaluateurs ?

Nous pensons pouvoir répondre par l'affirmative. D'une part, si l'expérience décrite n'a pas donné tous les résultats escomptés en raison des conditions de travail quelque peu "artisanales", elle renvoie néanmoins à des recherches menées ailleurs avec la rigueur et les précautions scientifiques d'usage. D'autre part, les artifices de l'expérience sont, ici et ailleurs, inspirés des situations habituelles d'évaluation en cours d'année scolaire ou aux examens.

Reprenons les cinq variables soumises à l'expérience :

- Variable A, le profit scolaire. : Quel enseignant s'abstient, au moment d'aborder la correction d'une copie, de se dire en identifiant son auteur : "celui-ci EST bon", "celui-là EST mauvais", "cet autre manque de sérieux" etc...? Malgré les efforts d'objectivité et la conviction intime qu'hier ne présume pas totalement d'aujourd'hui, qu'un élève est susceptible de progression (ce qui justifie nos efforts), de régression (d'où notre vigilance), d'accidents etc..., l'enseignant a bien du mal à se faire tout neuf et disponible face aux travaux d'élèves et d'étudiants familiers. Trop souvent les premières notes attribuées à l'élève déterminent à notre insu notre attitude présente et future. Malgré nous, nous "étiquetons", de sorte qu'un des traits du bon pédagogue pourrait se définir par sa capacité de reconstruire sans cesse l'image de ses élèves. Dans quelques cas limites -fort heureusement rares- on trouve des enseignants qui "abonnent" leurs élèves à une catégorie de notes : en-dessous ou au-dessus de 10 pour tel ou tel, entre 7 et 9 pour tel autre etc. Ne faut-il pas d'autre part voir un phénomène semblable dans notre habitude de demander à un collègue ce qu'il pense de X ou Y, comme si les résultats de celui-ci en telle matière entraînaient nécessairement un même type de résultats dans une autre matière ?

- Variable B, les notes données par d'autres; la 3ème correction. La 3ème correction, telle que pratiquée au bac par exemple, se fait dans l'ignorance des 2 premières notes données à une même copie. Elle joue alors pleinement son rôle car le troisième évaluateur s'attend à une copie litigieuse et s'impose une vigilance accrue. Ce n'était pas le cas dans l'expérience présente. On se proposait de mettre en lumière l'hésitation et l'incertitude habituelles de nos jugements évaluatifs, incertitude et hésitation qui se dissipent dès que l'instinct grégaire peut jouer : si deux personnes s'accordent pour noter entre 12 et 14 ou entre 7 et 9, que de soulagement à trouver ce "consensus" qui sécurise ! Le phénomène se rencontre lors d'évaluation en commun, par ex. aux examens quand les enseignants se réunissent pour coordonner leurs évaluations ou lorsqu'un évaluateur, embarrassé par une copie, demande à ses collègues "combien ils mettraient".

- Variable C, rôle des filières et séries; les lieux communs de la profession. On a vu que, contrairement à ce que l'hypothèse laissait prévoir, les évaluateurs ont apprécié les copies sans adapter les normes aux catégories d'élèves. Il y a là une défaillance dans l'évaluation. Mais il est certain que le jeune enseignant se sent fort démuni devant ses premières corrections parce qu'il n'a pas encore élaboré ses étalons de correction qui ne lui viendront qu'avec l'expérience. Prenons garde alors à ce qu'il ne passe pas à une attitude opposée et, tenant trop compte des différences de séries et de filières, il ne tombe dans les lieux communs contestables ! En effet, elles ne sont pas rares les réactions personnelles, en grande partie inconscientes, qui se traduisent par de l'agacement et de la colère ou bien de l'axisme et de l'indifférence : le professeur de français se venge de l'élève de série Mathématiques qui prend ses cours à la légère et est incapable d'écrire trois lignes correctement "parce que ce n'est pas une matière payante" ou bien l'enseignant ne sanctionne pas et ferme les yeux sur les fautes parce qu'il "comprend bien" que l'élève de Mathématiques fasse ses choix.

- Variable D, l'établissement scolaire. L'étude de cette variable mérite un approfondissement car elle est assurément celle que les enseignants reconnaissent le moins volontiers et contestent avec indignation. Ce qui signifie en clair qu'elle est le plus sournoisement agissante.

Oui, il y a des classes et des établissements réputés forts, faibles, élitistes, dépotoires. La population socio-culturelle majoritaire n'est pas la même dans tous les lycées. Cela est connu. Beaucoup moins évidentes, mais tout aussi contestables, sont les réactions sociales des enseignants face à des établissements ou des élèves trop "marqués". Tout se rencontre : on se reconnaît socialement dans la population-élèves et on la favorise inconsciemment (tel l'étudiant mentionné ci-dessus); on ne se reconnaît pas en elle, et les aigreurs remontent, ou le mépris, ou l'incompréhension. Lycées et universités sont des lieux de sourdes luttes de classes et l'objectivité pédagogique en pâtit.

- Variable E, effets séquentiels. Quel enseignant y échappe ? Qui

ne s'est surpris en train de sur ou sous-évaluer une copie qui émergeait, en bien ou en mal, d'une suite monotone, ni bonne, ni mauvaise ? L'effet séquentiel le plus connu de l'enseignant est celui qu'on n'a pu étudier dans l'expérience : l'accroissement de la sévérité en fonction du nombre de travaux corrigés comme si, dans sa lassitude et sa fatigue, l'enseignant avait besoin de s'en prendre à l'élève.

Nous savons que nombre de collègues s'offusqueront de cette généralisation de l'expérience à la pratique enseignante habituelle et se sentiront offensés qu'on puisse soupçonner leur objectivité et leur conscience professionnelle. Qu'on nous comprenne bien : il ne s'agit pas de stigmatiser tel ou tel d'entre nous. Tout enseignant, quelle que soit sa volonté d'objectivité, est guetté par la subjectivité et l'erreur dès qu'il fait office d'évaluateur. Il importe de savoir que l'évaluation est source d'injustices et d'erreurs graves et de rappeler d'autre part l'importance sociale des compositions trimestrielles et des examens scolaires. Et les cinq variables étudiées dans l'expérience relatée sont loin d'épuiser la liste des sources d'erreurs rencontrées dans l'évaluation scolaire :

Faut-il en conclure que l'acte d'évaluation exclut toute objectivité ? Nous ne le pensons absolument pas quoiqu'il soit exclu de prétendre parvenir un jour à des évaluations incontestables, surtout dans des épreuves comme les essais. La suite des recherches effectuées en TD avec la collaboration des étudiants a montré qu'une rationalisation de cette tâche était envisageable et qu'on pouvait espérer mettre un terme aux aberrations actuelles de la notation. C'est pourquoi il nous semble impératif de sensibiliser aux problèmes de docimologie les étudiants des licences d'enseignement, qu'ils se destinent au secondaire ou au supérieur. Ils doivent connaître les sources principales d'erreurs à l'oeuvre dans l'évaluation afin d'être vigilants et d'assurer une plus grande objectivité aux sanctions scolaires. Il faut aussi qu'ils réfléchissent sur les types d'évaluation et leurs finalités possibles. On peut encore aller plus loin, leur apprendre à fabriquer des tests et des barèmes d'évaluation, à se servir de l'échelle des notes ou classes, à formuler des observations constructives, etc.

D'autres TD, organisés autour de ces objectifs, se sont révélés

très positifs. Mais surtout il convient de donner à l'évaluation sa vraie dimension : plutôt que d'être sommative et de sanctionner épisodiquement, l'évaluation a une vocation formative et fait partie intégrante de l'acte d'apprentissage.

Pour conclure rappelons, avec les auteurs déjà cités, la nécessité que "l'évaluation soit définie comme un triple moyen de situer l'enseigné par rapport au comportement terminal à acquérir, de repérer la nature et la position des obstacles cognitifs à contourner , d'éprouver l'efficacité des procédés pédagogiques".(1)

(1) "Psychologie de l'évaluation scolaire" Noizet et Caverni, PUF, 1978. PP; 193 - 194.