CORRECTION DE DISSERTATIONS : ANALYSE DE L'ACTIVITE DE PROFESSEURS ENGAGES DANS UNE APPROCHE PAR COMPETENCES

Marc Vantourout, Rémi Goasdoué

Université Paris Descartes – SHS Sorbonne
Laboratoire EDA – Education et apprentissages
45, rue des Saints-Pères
75006 Paris
marc.vantourout@parisdescartes.fr
remi.goasdoue@parisdescartes.fr

Mots-clés : *activité de l'évaluateur, approche par compétences, dissertation*
Résumé. *L'approche par compétences en évaluation est-elle un moyen de réduire les divergences entre évaluateurs ? C'est à cette question, entre autres, que ce travail de recherche tente de répondre en comparant l'activité de professeurs corrigeant des dissertations de SES, certains le faisant de manière « classique », d'autres avec leur grille de compétences. Nous leur avons proposé de corriger les mêmes copies en verbalisant leurs remarques à propos de la copie. Nous retrouvons les constats établis par la docimologie sur les écarts de notation, mais il semble que l'utilisation de la grille ait quelques incidences sur les notes attribuées. Au-delà de ce constat chiffré, l'analyse des verbalisations révèle une variété de sources de divergences. Si les correcteurs ont tendance à s'intéresser aux mêmes passages des copies, ils ne paraissent pas nécessairement les appréhender et les juger de la même façon.*

---

1. Position du problème, cadre théorique, questions abordées

1.1 Position du problème : évaluer des compétences ?

En France selon les prescriptions institutionnelles, les enseignants du primaire et du secondaire se doivent désormais d'exercer leur métier dans le cadre d'une approche par compétences (APC). L'APC est indissociable d'un questionnement sur l'évaluation, ce qui se comprend aisément lorsque l'on sait que sa diffusion est étroitement liée à la mise en œuvre de dispositifs d'évaluation externes, dont l'emblématique PISA (Crahay, 2006). De nombreux spécialistes de l'évaluation (voir, par exemple, De Ketele, 2006 ; Carette, 2007) considèrent que la compétence constitue un nouvel objet pour l'évaluateur ; certains (Mottier-Lopez & Laveault, 2008) remarquent – à juste titre selon nous – que l'opérationnalisation de l'APC soulève un défi, en particulier quand il s'agit d'évaluer des compétences dites complexes (ou en lien avec des tâches complexes). Globalement, au regard de la quantité de travaux et publications qui peuvent s'inscrire dans une thématique que l'on nommera « L'évaluation dans une approche par compétences », on s'accordera à considérer que cette thématique s'avère une « question vive » dans le champ des recherches sur l'évaluation en éducation.

Les épreuves écrites constituent l'une des modalités canoniques d'évaluation lors de la scolarité. Elles sont utilisées pour évaluer des compétences. C'est ce que font, par exemple, B. Rey et coll. (2003) et V. Carette (2007) avec leur modèle permettant la construction d'épreuves selon trois phases. Ces auteurs, bien qu'ils fournissent des grilles critériées pour évaluer chacune des phases, reconnaissent leurs difficultés lorsqu'il s'agit de proposer des modalités pour évaluer la phase (ou tâche) impliquant une compétence complexe. Par ailleurs, ils ne se sont pas intéressés (jusqu'à maintenant en tout cas) à l'activité de celui qui est chargé de corriger ou d'évaluer leurs épreuves. Il existe d'autres publications, orientées également vers la formation pédagogique, où l'on trouve

des indications pour corriger ou évaluer des copies renvoyant à des taches complexes. Ces publication s'inscrivent dans des perspectives formative et formatrice (Veslin & Veslin, 1992 ; 2001) ou dans une APC (Roegiers, 2004 ; Scallon, 2004). Leur lecture laisse apparaître peu de différences, voire aucune différence, entre les perspectives formative et formatrice et l'évaluation dans une APC (en dehors des dates de parution plus récentes pour l'APC). Dans ces publications, l'explicitation des critères est centrale et les apports, le plus souvent judicieux, relatifs à l'activité de correction ou d'évaluation reposent pour une bonne part sur l'expérience, la capacité à innover et l'introspection des auteurs. Autrement dit, l'activité de l'évaluateur n'est pas vraiment étudiée, mais le lecteur est informé des « bonnes pratiques » afin de pouvoir progresser professionnellement, en particulier dans le domaine de l'évaluation. D'une manière générale, bien que l'expression « activité évaluative » se soit imposée dans la littérature (Figari & Achouche, 2001), les recherches centrées sur l'activité de l'évaluateur sont rares, que celui-ci s'inscrive ou non dans une APC.

Le travail présenté porte sur l'activité de professeurs de sciences économiques et sociales (SES) engagés dans une APC et que nous confrontons à une tâche de correction de dissertations. En étudiant – effectivement – l'activité de correction de ces professeurs, nous abordons un point aveugle des recherches sur l'évaluation en éducation, paradoxalement dissimulé par la question et vive mentionnée précédemment. L'enjeu principal est alors de mieux connaître et de comprendre le déroulement de l'activité de ces correcteurs, en nous centrant sur sa dimension cognitive.

1.2 Un nouveau cadre théorique pour appréhender l'activité de correction de copie

D'un point de vue historique et dans une perspective psychologique, les travaux sur la correction de copies renvoient à la docimologie (pour une synthèse, voir Piéron, 1963) puis à la docimologie expérimentale (pour une synthèse, voir Noizet & Caverni, 1978). En nous intéressant à la correction de dissertations, nous prolongeons en quelque sorte ces deux courants, en les renouvelant toutefois. En effet, les travaux mentionnés n'ont jamais véritablement porté sur l'évaluateur (ou correcteur), ni sur son activité (Vantourout, Goasdoué & coll., à paraître). Mais surtout, nous remettons en cause la place centrale accordée au processus de comparaison ainsi que les termes mêmes de cette comparaison, tels qu'ils figurent dans le « modèle explicatif des comportements d'évaluation » des docimologues expérimentaux (Noizet & Caverni, 1978). Ces derniers considèrent le jugement évaluatif comme une comparaison entre la copie à évaluer et un modèle de référence. De notre point de vue, l'activité déployée pendant la correction de copies ne peut être assimilée à un prélèvement-accumulation d'indices positifs ou négatifs. Nous appréhendons la correction de copie comme une tâche particulière de lecture-compréhension de texte et analysons l'activité évaluative (en particulier, lorsqu'il s'agit de dissertations) comme un processus de compréhension au sens conféré à ce concept dans les travaux sur la compréhension de texte, c'est-à-dire comme un processus interactif et dynamique impliquant un lecteur (qui élabore une représentation du texte) et un texte (où « figurent » des indices, dont la saisie dépend de la représentation et qui en retour contribuent à son évolution). Le jugement évaluatif ne se limite pas à un prélèvement d'indices et l'enjeu n'est plus d'identifier ce qui pèse sur le processus de comparaison, mais de comprendre comment s'élaborent les représentations issues de la lecture de la copie. Bref, le jugement évaluatif s'appuie plus vraisemblablement sur cette représentation que sur une simple accumulation d'indices. Par ailleurs, on peut également s'interroger sur l'existence a priori du second terme de la comparaison, c'est-à-dire du modèle de référence. A l'occasion de travaux sur l'évaluation formative en mathématiques, l'un de nous (Vantourout, 2007) a montré que certains professeurs élaborent leur modèle de référence au fil de l'activité de correction[1]. Si le modèle de l'évaluation-comparaison peut sembler satisfaisant lorsque l'on s'intéresse principalement au produit de la correction, à la note, et quand il s'agit de situations d'évaluation relativement « simples » ou « fermées » (comme la correction d'une dictée), sa valeur

[1] J. Ardoino et G. Berger (1986) ont évoqué l'existence d'un système d'attentes qui n'est ni complètement prédéfini, ni figé. A notre connaissance, il ne s'agit chez ces auteurs que d'une hypothèse judicieuse n'ayant jamais conduit à des travaux empiriques afin de la vérifier.

heuristique semble plus limitée pour étudier l'activité d'évaluation ou de correction en tant que processus et lorsque les tâches ou épreuves à évaluer sont complexes.

Le changement de cadre théorique de référence proposé permet d'étudier le déroulement de l'activité de l'évaluateur en s'intéressant aux caractéristiques de ce dernier et à un nouveau type de déterminants internes à l'évaluateur[2] : les connaissances[3] et représentations impliquées dans les processus de lecture-compréhension. La plupart des modèles élaborés dans le domaine de la compréhension de texte (Van Dijk & Kintsch, 1983 ; pour une synthèse, voir Blanc & Brouillet, 2003) s'accordent sur l'idée que comprendre un texte consiste à intégrer la succession des propositions dans une représentation cohérente. Les inférences produites lors de la lecture jouent ainsi un rôle déterminant dans la compréhension. Celle-ci est alors conçue comme un processus interactif dans lequel les connaissances et représentations du lecteur occupent une place prépondérante. Dans le cadre de l'activité de correction, les inférences produites peuvent être un moyen privilégié d'accès aux connaissances sollicitées lors de la lecture. On peut également supposer que la finalité évaluative de cette activité de lecture suscite la construction d'une représentation non seulement du texte, mais aussi l'élaboration d'une représentation évaluative du texte, voire de l'élève évalué.

1.3 Questions abordées

Dans cette présentation de nos travaux, nous chercherons à :

1. Vérifier si l'APC permet, en tout cas dans la présente situation d'évaluation, de limiter les écarts entre correcteurs, comme l'avance G. Scallon (2004, p. 180).
2. Dépasser le simple constat chiffré des divergences entre correcteurs et « comprendre », à travers l'analyse de l'activité, où, quand et comment celles-ci se produisent. En corollaire, estimer la valeur heuristique du nouvel ancrage théorique (appréhender l'activité de correction de dissertations comme un cas particulier de lecture-compréhension de texte), à l'origine des catégories permettant d'analyser l'activité des correcteurs.
3. Apporter quelques éléments relatifs au débat « Intérêts et limites d'une APC en évaluation ».

2. Présentation de la recherche

Les données présentées ici proviennent d'une recherche de plus grande ampleur, en cours. Dans la recherche mentionnée, douze professeurs expérimentés corrigent tous les mêmes copies manuscrites (photocopies), celles-ci ayant été produites par des élèves de terminale ES confrontés lors d'un baccalauréat blanc à une épreuve de dissertation avec étude de documents. Nous nous intéressons à la dissertation en SES car il s'agit d'une épreuve particulièrement adaptée aux enjeux théoriques de ce travail. En effet, les copies sont très riches car on demande aux élèves de disserter en prenant appui sur des documents – 3 textes et 3 documents statistiques – qui accompagnent le sujet. Le sujet est un sujet de sociologie figurant dans les annales du baccalauréat ES : *« Dans quelle mesure les mutations du travail peuvent-elles expliquer la crise du syndicalisme ? »*.

En nous appuyant sur l'expertise d'une collègue, nous avons extrait trois copies (copies X, Y, Z) d'un lot provenant d'une classe de terminale. Deux de ces copies sont estimées « moyennes », sachant que de telles copies sont celles qui poseraient le plus de difficultés aux évaluateurs. Plus précisément, nous voulions une copie (copie X) qui se lise aisément – nous parlons ici du plan rédactionnel –, mais plutôt « creuse » conceptuellement. A l'inverse, nous voulions une copie (copie Y) plus difficile à lire (maladresses) que la copie X, mais beaucoup plus aboutie sur le plan conceptuel. Enfin, nous avons choisi une copie estimée « mauvaise » (copie Z).

---

[2] Alors que les docimologues expérimentaux ont identifiés des déterminants externes auxquels seraient sensibles tous les évaluateurs, l'effet d'ordre par exemple (Noizet & Caverni, 1978).

[3] L'intérêt porté aux connaissances de l'évaluateur n'a pas pour fin de juger de sa légitimité, mais de mieux comprendre le rôle de celles-ci dans les processus de correction d'une copie.

La tâche proposée aux professeurs est une tâche de correction avec verbalisation concomitante. Lors de la consigne, nous demandons, d'une part, de corriger, d'annoter, d'attribuer une note aux copies et de rédiger une appréciation, d'autre part, de commenter à haute voix leur travail de correction sans hésiter à formuler toutes les remarques qu'ils jugent nécessaires et utiles à notre compréhension. L'ordre de correction des copies est imposé : X, Y, Z. L'activité de chaque professeur est entièrement filmée, le zoom de la caméra fixant la copie pour coordonner les annotations portées sur la copies et les verbalisations. Nous recueillons également les copies corrigées et annotées.

Parmi les douze professeurs qui ont été observés, quatre (A, B, C, D) utilisent pour corriger lors de l'expérimentation une grille de compétences commune à leur académie. Ils s'en servent depuis deux ans lors de l'épreuve de dissertation au baccalauréat et en classe. Elle est connue et « maîtrisée » par ses utilisateurs. Les huit autres professeurs (identifiés 1, 2, …), constituant un second groupe, ne s'inscrivent pas dans une APC et n'utilisent pas la grille pour corriger.

3. Premiers résultats : les notes attribuées

3.1 Constats classiques en docimologie

Dans le tableau 1, on présente les notes attribuées par chaque professeur aux trois copies ainsi que d'autres informations docimologiques.

	N'évaluant pas dans une APC (n'utilisent pas de grille de compétences)								**Evaluant dans le cadre d'une APC (utilisent la même grille de compétences)**			
Correcteurs-évaluateurs	1	2	3	4	5	6	7	8	**A**	**B**	**C**	**D**
Note attribuée à la copie X	9	14	10	11	15	8	8	12	16	12	14	12
Rang de la copie X (/ 3 copies corrigées)	2	1	1	1	1	2	2	2	1	2	2	2
Note attribuée à la copie Y	16	12	10	9	13	13	10	13	13,5	12,5	15,5	16,5
Total des points attribués aux copies X et Y	25	26	20	20	28	21	18	25	29,5	24,5	29,5	28,5
Note attribuée à la copie Z	9	4	5	5	5	6	5	7	5	6	5,5	6
Ecart entre note la + haute et note la + basse	7	10	5	6	10	7	5	5	11	6	10	12,5

Tableau 1 : Notes attribuées par les 12 correcteurs aux 3 copies – informations docimologiques

Nous serons brefs sur les constats docimologiques « classiques » des divergences de notation entre correcteurs. La copie X obtient des notes allant du simple au double, de 8 à 16, avec une moyenne de 11,75 et un écart type de 2,67. Tous les correcteurs, à l'exception du correcteur 4 (9/20), attribuent une note égale ou supérieure à 10 à la copie Y ; Les notes vont de 9 à 16, avec une moyenne de 12,83 et un écart type de 2,39. On peut avancer que pour l'ensemble des correcteurs la copie Y est mieux appréciée que la copie X. Les rangs des copies soulignent également des divergences : cinq correcteurs placent la copie X au premier rang[4], alors que sept le font pour la copie Y. Au regard des notes attribuées, les copies X et Y sont bien considérées comme des copies « moyennes » et la copie Z comme « une mauvaise » copie.

3.2 Incidences de l'utilisation de la grille de compétences

L'utilisation de la grille de compétences s'accompagne de finalités pédagogiques (voire de revendications) exprimées par les correcteurs A, B, C et D à l'occasion d'entretiens post-corrections. Parmi les principales :

1. Réduire les divergences entre évaluateurs, corriger plus « juste » en s'accordant sur les attentes de la dissertation.
2. Parvenir à mieux utiliser l'ensemble de l'échelle de notation.

[4] Lors de l'entretien individuel post-correction, nous demandons aux correcteurs s'ils ont eu des hésitations quant à la hiérarchisation des copies.

3. Valoriser ce qui doit l'être, en attribuant d'excellentes ou de mauvaises notes au baccalauréat aux copies qui le méritent, comme cela se fait dans d'autres disciplines (les notes en mathématiques tenant lieu ici de référence).

Dans le tableau 2, nous avons rassemblé des informations qui complètent celles figurant dans le tableau 1 et qui peuvent constituer des apports supplémentaires pour discuter de l'atteinte des finalités annoncées. Reposant souvent sur des effectifs réduits (4 correcteurs), les moyennes présentées ne sauraient constituer des « preuves » ; cependant, ces informations nous permettent de développer nos réflexions.

	Moyenne des notes des 12 correcteurs	Moyenne des notes avec grille (correcteurs A B C D)	Moyenne de notes sans grille (correcteurs 1, …, 8)	Moyenne des 4 notes les plus élevées sans grille (4 parmi corr. 1, …, 8)	Moyenne des 4 notes les plus basses sans grille (4 parmi corr. 1, …, 8)
Copie X	11,75	13,50	10,88	13,00	8,75
Copie Y	12,83	14,50	12,00	13,75	10,25
Copie Z	5,71	5,63	5,75	6,75	4,75

Tableau 2 : Comparaison des moyennes des notes attribuées par les correcteurs qui utilisent la grille de compétences et ceux qui ne l'utilisent pas

Avec les correcteurs A, B, C et D, on trouve également des écarts dans les notes attribuées aux copies X et Y, respectivement 4 points (de 12 à 16) et 3 points (de 13,5 à 16,5 ; voir tableau 1). En revanche, pour la « mauvaise » copie (copie Z), l'écart se limite à 1 point. On remarquera, en prenant cette dernière copie, qu'il est possible de trouver quatre correcteurs (les correcteurs 3, 4, 5 et 7) qui n'utilisent pas la grille de compétences et qui attribuent tous la note 5 à la copie Z. Pour la copie X, on peut trouver quatre correcteurs, les correcteurs 2, 4, 5 et 7, dont l'écart des notes est de 3 points (de 12 à 15), c'est-à-dire inférieur à celui des correcteurs utilisant la grille (4 points, de 12 à 16). L'utilisation de cette grille ne semble pas produire de meilleurs résultats en matière d'accords inter-correcteurs que ceux obtenus en regroupant les notes de certains professeurs (ou, pour dire les choses autrement, il faudrait avoir davantage de correcteurs avec grille pour pouvoir conclure sur ce point, ce qui sera fait prochainement).

En revanche, l'utilisation de la grille pourrait servir de « garde-fou » et avoir des incidences sur l'utilisation de l'échelle de notation. Pour les copies moyennes (copies X et Y), en se référant au tableau 2, on remarque que les moyennes des notes attribuées avec la grille sont supérieures d'environ 1,7 point à la moyenne des notes des douze correcteurs (13,50 – 11,75 et 14,50 – 12,83) et de plus de 2,5 points à la moyenne des correcteurs sans grille (13,50 – 10,88 et 14,50 – 12). Les correcteurs qui utilisent la grille auraient donc tendance à donner plus de points aux copies « moyennes », comme cela apparaît dans le tableau 1 où, avec 29,5 et 28,5 points attribuées aux copies X et Y, les correcteurs A, C et D occupent les trois premières places en terme de « générosité ». Les correcteurs 5 et 2 (28 et 26 points) sont quasiment aussi généreux. Toutefois, ces derniers font un peu figure d'exception au sein de leur « groupe » : pour arriver à ce total, ils ne donnent jamais de note égale ou inférieure à 10 aux copies X ou Y, contrairement à ce que font tous les autres (correcteurs 1, 3, 4, 6, 7 et 8). Tous ces éléments vont dans le sens de ce que nous avons écrit en début de paragraphe : l'utilisation de la grille, qui empêcherait de « sous-noter » les copies moyennes, servirait alors de garde-fou. Pour essayer de résumer et d'appuyer cette conclusion, tout en faisant des comparaisons avec des groupes de professeurs de même effectif, nous avons calculé, pour chaque copie et pour les correcteurs n'utilisant pas la grille de compétences, la moyenne des quatre meilleures notes (M4MN) et celle des quatre notes les plus basses (M4BN). Pour les copies X et Y, s'il apparaît que la M4MN est très proche (bien qu'inférieure) de la moyenne des correcteurs avec grille (copie X : 13 et 13,50 ; copie Y : 13,75 et 14,50), la valeur de la M4BN est relativement basse, en particulier pour la copie X (8,75 ; copie Y : 10,25). Faute de place, nous ne développerons pas ici le même type d'analyse pour la copie Z. Pour cette copie, on pourrait également penser que la grille joue un rôle de garde-fou, mais cette fois il fonctionnerait dans deux directions opposées en empêchant de sous-noter et de sur-noter. Quant à une utilisation plus large de l'échelle de notation, les correcteurs A et D sont ceux dont l'écart entre la note la plus basse et la note la plus haute est le plus important (respectivement 11 et

12,5 points). Il y a cependant des correcteurs (2 et 5) dont l'écart, 10 points, est le même que celui du correcteur C. On peut avancer, mais avec prudence, que l'utilisation de la grille de compétences s'accompagnerait d'une utilisation d'un empan plus important de l'échelle de notation.

4. Analyse de l'activité de correcteurs engagés dans une APC pour évaluer

Nous nous intéressons dans cette partie d'abord aux correcteurs A, B, C et D, puis uniquement aux trois premiers cités (l'analyse de l'activité du correcteur D étant en cours).

4.1 Présentation de la grille de compétences

Faute de place, nous ne nous étendrons pas sur la description et l'analyse de la grille d'évaluation dont un extrait figure en annexe 1. A l'origine, il s'agissait au sein d'une académie d'approfondir les critères d'évaluation en termes de compétences. Les concepteurs de la grille se sont appuyés sur les textes institutionnels relatifs aux épreuves de SES au baccalauréat et ont puisé leurs références théoriques chez des spécialistes de l'évaluation des compétences.

La grille comporte sept compétences (Introduire, Mettre en œuvre des séquences argumentatives pour répondre au sujet…) évaluées à l'aide sept échelles descriptives globales[5]. Pour chaque compétence, ces échelles comportent quatre niveaux d'exigence (non, plutôt non, plutôt oui, oui) auxquels correspondent plusieurs indicateurs. Le barème qui permet de valoriser certaines compétences n'accorde pas à chacune d'entre elles le même nombre de points. Dans les limites fixées par le barème (et par le bon sens) et selon la copie, le correcteur bénéficie d'une certaine autonomie pour décider du nombre de points correspondant au niveau d'exigence qu'il juge atteint[6]. Un mode d'emploi est destiné aux utilisateurs. Cette grille – sophistiquée – est donc proche des outils que l'on trouve dans les ouvrages pédagogiques mentionnés précédemment. Sa diffusion a donné lieu à des stages de formation et il existe des commissions d'harmonisation ; les correcteurs A, C et D ont été et restent fortement impliqués dans ces dispositifs. Quant au correcteur B, qui utilisait auparavant sa propre grille, il déclare n'avoir participé à aucune action de formation et être un collègue lambda qui a reçu la grille par la voie hiérarchique et qui se l'est appropriée seul.

4.2 Attribution des points et grille de compétences

La comparaison du détail des points attribués par les correcteurs pour les différentes compétences (voir annexe 2) montre que s'il peut y avoir parfois accord quant au degré d'exigence atteint et au nombre de points attribués, ce n'est pas le cas le plus fréquent : pour la copie Y, la compétence « Mettre en œuvre des séquences argumentatives… » est créditée unanimement de 3 points (sur 4 possibles), alors que, pour la copie X, les correcteurs A, B, C, D attribuent respectivement à cette compétence 4, 2,5, 4 et 2 points. Dans ce dernier cas, les niveaux d'exigences jugés atteints diffèrent selon les correcteurs. Autre forme de divergence : pour une même note finale à une copie, des correcteurs attribuent un nombre de points différents à une ou plusieurs compétences. Par exemple, les correcteurs B et D notent 12/20 la copie X en attribuant un nombre de points différent à cinq compétences sur sept. Bien qu'une différence de points puisse s'expliquer par la part d'autonomie du correcteur lors de l'opérationnalisation de la correspondance « niveau d'exigence atteint ⇔ points attribués », certains écarts renvoient à des jugements différents à propos du niveau d'exigence atteint, comme le confirment nos analyses[7] (voir ci-après).

---

[5] Pour une présentation de ces échelles, se référer à G. Scallon (2004). Pour une illustration, voir l'annexe 1.

[6] Par exemple, pour la compétence « Argumenter » (barème sur 4 points), combien de points un correcteur va-t-il accorder à une copie lorsqu'il estime que le niveau d'exigence atteint correspond à « plutôt oui » : 3 points ? moins de 3 points ? plus de 3 points ?

[7] Pour la compétence « Introduire », notée sur 2 points, un écart de 1/2 point a de fortes chances de renvoyer à des niveaux d'exigences différents. Il n'en va pas obligatoirement de même pour la compétence « Argumenter », notée sur 4 points (voir également la note précédente).

Ce constat des divergences entre correcteurs ne peut conduire à les incriminer et à douter de leur compétence : il doit être dépassé. Une voie pour y parvenir consiste à analyser leur activité afin de mieux la connaître et la comprendre.

4.3 Inférences et jugements évaluatifs : convergence/divergence et accord/désaccord entre correcteurs

Nous allons présenter les résultats de l'analyse comparative de l'activité des correcteurs A, B et C, confrontés à la copie X. Les analyses portent sur la transcription intégrale des vidéos recueillies. Pour pouvoir analyser l'activité des correcteurs, nous mettons en relation lors de la transcription ce que dit le correcteur (verbalisations concomitantes à la correction) et/ou ce qu'il fait (lit la copie, annote la copie, utilise la grille, consulte le dossier documentaire…) avec l'objet concerné par son discours et/ou son action (phrase de la copie, partie de la copie, document…).

4.3.1 Première illustration : jugement évaluatif et accord entre correcteurs

Commençons par un cas « simple » d'accord entre correcteurs. Ici, l'accord concerne à la fois la compétence évaluée et l'orientation (positif, négatif) du jugement évaluatif. Dans l'introduction figure la phrase suivante (phrase 2 de la copie) : *Les salariés n'adhèrent donc plus aux syndicats, c'est pourquoi on parle de crise du syndicalisme*. Les correcteurs A et B déclarent respectivement que « l'élève fait un effort pour préciser ce qu'est la crise du syndicalisme » et que « « l'élève fait un effort de définition des termes du sujet ». Le correcteur C, bien que plus nuancé, va dans le même sens : « La crise du syndicalisme est définie de façon acceptable ». En nous appuyant sur la grille et les verbalisations, nous remarquons que les trois correcteurs, à la lecture de cette phrase, la relèvent (« en disent quelque chose ») et évaluent tous la même compétence, « Introduire », à l'aide de l'indicateur « Cadrage du sujet » pour lequel il est attendu que les élèves définissent les termes du sujet. Les jugements évaluatif exprimés concordent : ils sont positifs.

4.3.2 Deux autres illustrations : inférences, convergence/divergence et accord/désaccord entre correcteurs

Dans nos travaux, nous adoptons la définition communément admise de la notion d'inférence : toute information non explicitement présente dans le texte que le lecteur construit à partir de ses connaissances. Ailleurs (Vantourout, Goasdoué & coll., à paraître), nous avons distingué trois formes d'inférences que nous ne pouvons présenter ici en les détaillant : inférences de compréhension, inférences évaluatives sur le contenu, inférences évaluatives sur l'élève[8]. Dans la suite du texte, nous proposons des illustrations des deux premières et invitons le lecteur curieux à consulter la référence mentionnée ci-dessus.

Dans l'introduction, on trouve également (phrases 3 et 4) : *Ce déclin puise une partie de ses causes dans les mutations du travail. Les mutations sont des transformations, des changements du monde du travail.* Les trois correcteurs continuent d'évaluer la compétence « Introduire » et expriment tous un jugement évaluatif négatif ; le correcteur A, à titre d'illustration, déclare : « Il manque un exemple, définir mutations par changements est un peu facile. ». Il y a accord sur ce point entre les correcteurs. Le correcteur C pense de plus que « l'élève a sans doute mal compris la notion de mutations » (il le dit en fin de correction en revenant sur l'introduction). Pour avancer cela, il a produit lors de sa lecture une inférence de compréhension. Comme, il est le seul à verbaliser cette inférence à cet endroit de la copie, nous relevons une divergence entre ce correcteur et les deux autres. Cette inférence de compréhension s'accompagne d'une inférence évaluative sur le contenu et le correcteur C est le seul qui évalue « ici » la compétence « Mobiliser

[8] Faute de place, nous ne proposerons pas d'illustration pour les inférences évaluatives sur l'élève. Un correcteur produit une telle inférence lorsqu'il dit : « A-t-elle vraiment repris l'un des éléments du document 2 ou le fait-elle pour faire plaisir au correcteur ? ». A cet endroit de la copie, le correcteur évalue négativement la compétence « Mobiliser des informations des document… », alors que d'autres valident l'utilisation du document 2.

des connaissances… » (à l'aide de l'indicateur « Maîtrise les notions du programme »). Son jugement évaluatif est négatif.

Troisième page de la copie (phrases 24 et 25) : *Les syndicats sont devenus des institutions, c'est pourquoi les salariés ne se reconnaissent plus dans les syndicats et n'y adhèrent donc plus. Les représentants des syndicats ne sont plus des salariés c'est pourquoi on peut se demander si les syndicats défendent leurs propres intérêts ou ceux des adhérants* (orthographe de la copie). En lisant ce passage, le correcteur C ne tarit pas d'éloges envers l'élève : « Très bon mécanisme, très bonne argumentation. La chose est loin d'être simple à comprendre pour des élèves de terminale. Ce n'est pas évident pour l'élève de comprendre que l'institutionnalisation provoque une sortie des délégués syndicaux trop fréquente de l'entreprise pour qu'il y ait un contact important entre les collègues ». Son jugement évaluatif est extrêmement positif et il valide un « mécanisme », indicateur de la compétence « Mobiliser des connaissances… ». Le correcteur B réclame une définition pour le terme institution et signale que l'on parle plutôt « d'institutionnalisation des syndicats ». Mais surtout, il annote le début de la phrase 25 (*Les représentants des syndicats ne sont plus des salariés*) ainsi : « Ne sont plus des salariés présents dans l'entreprise. Donner un exemple pour construire la démonstration ». Ces deux correcteurs n'ont pas la même interprétation de ce passage, en particulier quant au niveau de compréhension de l'élève. Le correcteur C (en lisant ces phrases) produirait une inférence de compréhension qui l'amènerait à créditer l'élève « de la sortie des délégués syndicaux de l'entreprise » (tout se passe comme s'il pensait que l'élève s'exprime mal) puis une inférence évaluative sur le contenu le conduisant à penser que l'élève a compris et à dire qu'il est en présence d'un « mécanisme » pouvant être validé. Le correcteur B, quant à lui, produit certainement une inférence de compréhension qui est proche de celle du correcteur C ; il est ainsi en mesure de rétablir de qui est erroné dans la phrase de l'élève. Mais contrairement au correcteur C, il prend « l'élève au pied de la lettre », considère que ce dernier n'a pas compris (ou ne montre pas qu'il a compris) et qu'il n'y a pas là de séquence argumentative qui puisse être validée.

Des analyses inter-correcteurs, nous retenons ici trois résultats :

1. Un premier, attendu : une compétence peut-être évaluée différemment dans une même copie par des correcteurs, et ceux-ci peuvent attribuer à la copie la même note globale.
2. Un second, plus surprenant : une compétence peut-être évaluée globalement de la même manière par des correcteurs – accord sur le niveau d'exigence atteint et sur les points attribués –, alors que ce ne sont pas obligatoirement les mêmes passages de la copie qui ont été pris en compte lors de l'évaluation de la compétence, ou encore, si ce sont bien les mêmes passages, pour certains d'entre eux l'orientation des jugements évaluatifs a pu différer.
3. Un troisième, que nous venons d'illustrer et qui témoigne de la richesse des analyses : des inférences de compréhension analogues peuvent donner lieu à des jugements évaluatifs différents.

4.4 Résumer les convergences/divergences et accords/désaccords entre correcteurs

Longue de plus de six pages, la copie X comporte quarante-trois phrases. Quand on compare des correcteurs, les points de convergence/divergence et les accords/désaccords sont très nombreux. Aussi, est-il difficile d'en tirer une conclusion. Nous avons alors cherché à résumer toutes ces informations, à disposer d'une vue plus synthétique : pour chacune des phrases, nous comparons trois paires de correcteurs et proposons de calculer trois taux pour mesurer la convergence et l'accord inter-correcteurs.

Le taux de convergence de « l'intérêt/désintérêt »[9] des paires de correcteurs est élevé, autour de 80% (voir, ci-après, le tableau 3). Les correcteurs montrent de l'intérêt pour les mêmes parties de

[9] Nous relevons, pour une phrase donnée, une convergence de l'intérêt/désintérêt entre deux correcteurs quand tous deux « disent quelque chose » relativement à cette phrase (verbalisation et/ou annotation) ou l'ignorent. Si un seul des deux correcteurs verbalise ou annote une phrase, nous relevons une divergence inter-correcteurs. Ce taux est calculé de la manière suivante : $\sum$ conv / $\sum$ (conv + div).

la copie (il y a très peu de phrases faisant l'objet d'un désintérêt commun). Il reste cependant 20% des phrases pour lesquelles il y a une divergence d'intérêt. Nous avons peut-être là un premier élément permettant de comprendre les divergences entre correcteurs.

	Comparaison entre		
	Correcteurs A et B	Correcteurs B et C	Correcteurs A et C
Taux de convergence de "l'intérêt/désintérêt"	81%	83%	77%
Taux de convergence des inférences : TCI	49%	37%	29%
Taux d'accord des jugements évalautifs : TAJE	37%	27%	29%

Tableau 3 : Taux inter-correcteurs de convergence de l'intérêt/désintérêt, de convergence des inférences (TCI) et d'accord des jugements évaluatifs (TICE)

Le tableau 4 permet d'illustrer la méthode de calcul des taux inter-correcteurs de convergence des inférences (TCI) et d'accord des jugements évaluatifs (TAJE). Il contient également des informations complémentaires pour appréhender les divergences entre correcteurs.

INFERENCES	Corr. A (16//20)	Corr. B (12/20)	Conv	Div
Inf. de compréhension	21	21	15	13
Inf. éva. sur les contenus	13	8	6	8
Inf. éva. sur l'élève	1	3	1	2
Converg. corr. A et B (conv / conv + div)	**48,89%**			

JUG. EVALUATIFS	Corr. A (16//20)	Corr. B (12/20)	Accord	Désaccord
Positifs	26	16	14	17
Négatifs	13	36	11	27
Autres	4	5	3	3
Accord corr. A et B (acc. / acc. + désacc.)	**37,33%**			

Tableau 4 : Calculs des taux de convergence des inférences (TCI) et d'accord des jugements évaluatifs (TAJE) pour les correcteurs A et B corrigeant la copie X

Tout d'abord, si nous nous référons aux résultats d'analyses antérieures (Vantourout, Goasdoué & coll., à paraître), nous remarquons que les correcteurs observés produisent (verbalisent) des inférences en très grand nombre. Leur stratégie de lecture leur permet d'accéder à une compréhension fine du texte. Ils fondent leur évaluation sur une représentation de la copie issue d'un traitement systématique du texte (plus coûteux cognitivement qu'un traitement heuristique). Il y a convergence lorsque deux correcteurs produisent la même inférence (ou une inférence de même forme) pour une phrase donnée. Il y a divergence si un seul correcteur produit une inférence ou si les inférences ne sont pas de même forme. Les TCI calculés s'échelonnent entre 49% et 29% (voir tableau 3 ; pour la méthode de calcul, voir tableau 4). Moins de la moitié et parfois moins du tiers des inférences (toutes formes confondues) produites par une paire de correcteurs seraient communes. Autrement dit, il peut y avoir jusqu'à plus de deux tiers d'inférences non partagées. Si l'on se limite aux inférences de compréhension, le TCI est à peine supérieur à 50% pour les trois paires comparés. Quand on connait l'importance des inférences dans les processus de compréhension de texte, et si l'on veut bien considérer que le professeur évalue la représentation de la copie (et celle de son auteur) qu'il s'est construite, nous pensons tenir là une explication complémentaire des divergences entre correcteurs-évaluateurs.

Les modalités de calcul du TAJE sont analogues à celles du TCI (voir tableau 4). Les TAJE calculés s'échelonnent entre 37% et 27%. Comme cela a été illustré ci-dessus (voir 4.3.2), d'une inférence de compréhension convergente ne s'ensuit pas obligatoirement un jugement évaluatif de même orientation (voire, de même nature). Nous avons là un nouvel élément qui devrait contribuer à expliquer les divergences entre correcteurs. Il est d'ailleurs remarquable que ce soient les correcteurs A et B, dont l'écart de notation est le plus important (4 points), pour lesquels nous trouvons les TCI et TAJE les plus élevés (voir tableau 3). En quantifiant, nous faisons comme si tous les jugements évaluatifs avaient un poids équivalent, ce qui n'est bien sûr pas le cas. Selon nous, cette limite, qui doit rester présente à l'esprit, ne retire rien aux apports de l'analyse quantitative présentée. De plus et toujours avec la même réserve, les comparaisons inter-correcteurs en matière d'orientation des jugement évaluatifs nous semblent pouvoir expliquer des

écarts de notations : les notes de 16/20 et de 12/20 sont respectivement attribuées par des correcteurs qui expriment 66% (26/39) de jugements évaluatifs positifs et 70% (36/52) de jugements évaluatifs négatifs.

5. Eléments de conclusion

Dans cette conclusion, nous revenons sur les trois questions essentielles qui jalonnent ce travail (voir 1.3).

1) En lien avec le constat maintes fois répété des divergences entre évaluateurs, ce travail avait notamment pour but – en comparant des notes produites par certains professeurs corrigeant avec leur grille de compétences et par d'autres corrigeant « classiquement » sans grille – d'estimer l'effet de l'utilisation d'une grille de compétences sur les écarts de notations. Si nous retrouvons, en considérant tous les professeurs observés, des notes pouvant varier du simple au double pour la même copie, nous pensons avoir identifié une tendance (qui devra être confirmée par des observations supplémentaires) à la réduction de ces écarts pour le groupe de professeurs utilisant la grille de compétences. Mais surtout, nous avons relevé chez ces derniers un consensus plus net dans la différenciation entre des copies considérées comme « moyennes » et une copie « moins bonne ». La grille permettrait donc de valoriser les copies dites « moyennes » tout en ne favorisant pas indûment les « mauvaises copies ». Cet effet semble en accord avec les fonctions assignées à la grille par ses concepteurs (voir 3.2). Nous devons toutefois indiquer qu'une même note globale ou des notes globales proches (notes sur 20) peuvent résulter d'une série d'appréciations assez différentes quant au niveau d'exigence atteint par chacune des compétences évaluées. Par ailleurs, un même nombre de points pour une compétence donnée (note partielle liée au niveau d'exigence atteint) peut résulter de jugements évaluatifs divergents.

2) Ces derniers éléments, qui révèlent combien l'activité de correction de dissertations est complexe, en particulier lorsque les correcteurs-évaluateurs s'inscrivent dans une APC, soulignent l'intérêt d'analyser leur activité. La richesse des verbalisations montre d'abord que l'utilisation de la grille engagerait ces correcteurs dans une lecture approfondie du texte. Leurs nombreuses inférences témoignent d'une compréhension et d'une analyse fines du texte. La mise en relation des inférences avec les phrases de la copie permet de décrire la multiplicité des sources de divergences. Si ces correcteurs ont tendance à s'intéresser globalement aux mêmes passages de la copie, ils ne paraissent pas nécessairement les appréhender et les juger de la même façon. Nous avons observé, en comparant l'activité de paires de correcteurs, que ce ne sont pas obligatoirement les mêmes passages d'une copie qui sont pris en compte pour évaluer une compétence donnée et, par ailleurs, qu'un même passage peut contribuer à la validation de compétences différentes. Ces résultats confirment, selon nous, l'intérêt d'analyser l'activité de correction en mobilisant des concepts issus des travaux en psychologie de la compréhension de texte. La notion d'inférence permet d'accéder conjointement au processus de compréhension, aux connaissances, aux attentes du correcteur et donc d'expliquer localement la production des jugements évaluatifs.

3) Enfin, ce travail apporte des éléments de réponse contrastés au débat sur « les intérêts et les limites d'une APC en évaluation ». Les constats précédents à propos des incidences manifestes de la grille semblent à première vue plaider en faveur de ce type d'approche. Cependant, dans notre travail, rien n'indique que la formulation de critères d'évaluation en termes de compétences soit à l'origine de ces constats. Il pourrait s'agir de l'effet de l'utilisation d'une grille quelle qu'elle soit : celle-ci, en tenant lieu de mémoire externe, favoriserait une certaine « décharge cognitive » et permettrait au correcteur d'être « plus » exhaustif. Dans nos observations, tout se passe comme si le recours à la grille produit un effet cumulatif qui entraîne la valorisation des copies « moyennes » ; la lecture du correcteur est davantage encadrée par la grille et, en quelque sorte, ce qui serait validé dans la grille constituerait un « acquis » ne pouvant être balayé par des maladresses ou des erreurs, (contrairement à ce nous avons pu observer parfois chez des correcteurs corrigeant sans grille). Notons que, pour certains professeurs, cet effet cumulatif – « un acquis est un acquis » – associé à l'absence de points négatifs constitue plutôt une limite, voire un défaut, de l'outil d'évaluation. Bref, en l'état actuel de nos travaux, il ne nous semble pas possible

trancher de manière ferme sur les apports d'une APC en évaluation relativement aux objectifs que se sont fixés les professeurs observés (concepteurs et utilisateurs de la grille). En complément, dans le cadre d'un questionnement plus pédagogique (qui n'est pas celui que nous privilégions dans nos travaux), nous avons remarqué que les informations produites sur la copie sont extrêmement riches et nous semblent pouvoir favoriser la mise en œuvre d'une évaluation véritablement formative. Toutefois, subsistent là aussi des interrogations : si l'on considère l'importance des divergences et désaccords observés chez ces professeurs, quels peuvent être les feedbacks adressés à élève ? Ceux-ci, pour un même passage, pourront être positifs ou négatifs en fonction du professeur qui aura corrigé sa copie ? Enfin cette grille, qui semble être un élément central au sein d'un dispositif d'enseignement de la dissertation, a permis de développer un outil et une démarche d'évaluation qui servent à la fois la formation et la certification, avec une dernière limite que ne manqueront pas de soulever certains détracteurs de l'APC : le risque d'« enseigner la grille ».

6. Références et bibliographie

Ardoino, J. & Berger, G. (1986). L'évaluation comme interprétation. *Pour, n°107*, 120-127.

Blanc, N. & Brouillet, D. (2003). *Mémoire et compréhension - lire pour comprendre*. Paris: In Press.

Carette, V. (2007). L'évaluation au service de la gestion des paradoxes liés à la notion de compétence. *Mesure et évaluation en éducation, vol. 30/*2, 49-71.

Crahay, M. (2006). Dangers, incertitudes et incomplétudes de la logique de compétence. *Revue française de pédagogie, n°154*, 97-110.

De Ketele, J.-M. (2006). Interroger les « objets » - La compétences, In G. Figari & L. Mottier-Lopez, *Recherche sur l'évaluation en éducation* (pp. 17-24). Paris : L'Harmattan.

Figari, G. & Achouche, M. (2001). *L'activité évaluative réinterrogée : Regards scolaires et socioprofessionnels*. Bruxelles : De Boeck.

Mottier Lopez, L. & Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation, vol.31/*3, 5-34.

Noizet, G. & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire.* Paris : PUF.

Piéron, H. (1963). *Examens et docimologie*. Paris : PUF.

Rey, B., Carette, V., Defrance, A. & Kahn, S. (2003). *Les compétences à l'école – Apprentissage et évaluation*. Bruxelles : De Boeck.

Roegiers, X. (2004). *L'école et l'évaluation*. Bruxelles : De Boeck.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.

Van Dijk, T. A. & W. Kintsch. (1983). *Strategies of Discourse Comprehension*. New York : Academic.

Vantourout, M. (2007). Etude de l'activité évaluative de professeurs stagiaires confrontés à des productions d'élèves en mathématiques : quel référent pour l'évaluateur ? *Mesure et évaluation en éducation, vol. 30/*3, 29-58.

Vantourout, M. Goasdoué, R., Lanéelle, X. & Martins, D. (à paraître). Etude de l'activité évaluative de quelques professeurs de sciences économiques et sociales. *Actes du 21e colloque international de l'Admeé-Europe*, Louvain-la-Neuve, Belgique, janvier 2009.

Veslin, O & Veslin, J. (2001). Evaluation formatrice et critères de réalisation, In G. Figari & M. Achouche, *L'activité évaluative réinterrogée* (pp. 89-101). Bruxelles : De Boeck.

Veslin, O & Veslin, J. (1992). *Corriger des copies – Evaluer pour former*. Paris : Hachette Education.

Annexe 1

Extrait de la grille académique d'évaluation de la dissertation

Compétence	Indicateur	Niveau d'exigence				Barème
		Non	Plutôt non	Plutôt oui	Oui	
Introduire	• **Cadrage du sujet :** *sens des notions, champ géographique, champ historique, faits incontournables.* • **Tâche à accomplir :** *termes d'un débat, effets d'une variables sur une autre, causes, mises en relation…* • **Annonce de plan.**	Absence de cadrage ET De tâche à accomplir d'annonce de plan OU Annonce de plan seule	Cadrage incomplet sans définition sans la tâche à accomplir avec annonce de plan OU Cadrage incomplet sans définition avec la tâche à accomplir mais sans annonce du plan	Cadrage incomplet mais avec définitions ET tâche à accomplir ET Annonce de plan	Cadrage complet ET Tâche à accomplir ET Annonce du plan	2
Mettre en œuvre des séquences argumentatives <u>pour répondre au sujet</u>	Présence explicite d'une idée par séquence reliée au sujet Présence d'argument(s) pour la défendre Présence d'une illustration (il s'agit ici d'évaluer seulement la cohérence de l'argumentation)	Seulement un des indicateurs	Idée ET illustration MAIS pas d'argument	Idée ET illustration MAIS partiellement argumenté OU Idée ET argument SANS illustration *(2 à 3 fois)*	Idée ET arguments ET illustration *(3 à 4 fois)*	4

Annexe 2

Points attribués par les correcteurs A, B, C et D aux compétences évaluées dans la copie X

Compétences évaluées	**Barème**	**Corr. A**	**Corr. B**	**Corr. C**	**Corr. D**
Introduction	/2	1,5	1	0,5	0,5
Mettre en œuvre une réponse organisée	/2	1,5	1,5	1,5	1
Conclusion	/1	0,5	1	0	0,5
Utilisation pertinente pour répondre au sujet des termes du sujet	/2	2	1,5	2	1,5
Mettre en œuvre des séquences argumentatives pour répondre au sujet	/4	4	2,5	4	2
Mobiliser des connaissances pour répondre au sujet	/4,5	3	2,5	3,5	2,5
Mobiliser des informations des documents pour répondre au sujet	/4,5	3,5	2	3	3,5
Présentation de la copie	Bonus Malus			-0,5	
Note finale sur 20	/20	16	12	14	11,5 =>12