

ÉVALUATIONS ET JUGEMENTS DES ENSEIGNANTS

Marc **VANTOUROUT**

Université Paris Descartes

marc.vantourout@parisdescartes.fr

Rémi **GOASDOUÉ**

Université Paris Descartes

remi.goasdoue@parisdescartes.fr

Résumé

Sous l'impulsion des travaux pionniers des docimologues, l'étude des pratiques d'évaluation a durablement tourné autour de la question des divergences entre évaluateurs tout en négligeant l'effet des spécificités des contenus en jeu dans les évaluations. L'enjeu principal de l'approche, que nous qualifions de psycho-didactique des évaluations est précisément de combiner l'apport des didactiques disciplinaires et des travaux en psychologie sur l'analyse de l'activité pour mieux comprendre les activités d'évaluations et remettre au centre des préoccupations la question de la validité. Nous présentons des travaux en mathématiques et en sciences économiques et sociales (aspects méthodologiques et résultats) qui montrent l'intérêt de ces approches pour étudier l'activité de correction ou d'évaluation proprement dite.

Mots clés

Évaluation, jugement des enseignants, didactique, analyse de l'activité, situations aménagées

QUELQUES REPERES

Préambule : travaux relatifs à l'évaluation au sein de l'équipe EDA¹

Au sein de l'équipe EDA, depuis plus de quinze ans, des recherches portent sur l'activité des évaluateurs, avec la volonté d'articuler évaluation et didactique de la discipline impliquée. Plusieurs thèses ont été soutenues, principalement sous la direction de Sylvette Maury (Vantourout, 2004 ; Nabbout, 2006), d'autres vont l'être prochainement (Smith, 2015 ; Blanc, 2016). La thématique de l'évaluation permet actuellement de caractériser cette équipe où nous sommes de plus en plus nombreux à nous intéresser à ce domaine (cf., par exemple, Roditi & Salles, 2015). Les premiers travaux, consacrés aux mathématiques, ont permis développer un regard didactique sur les questions liées à l'évaluation des élèves et de leurs apprentissages. Puis, nous avons travaillé sur les sciences économiques et sociales, épaulés par des spécialistes

¹ Le laboratoire EDA (EA 4071) – Éducation, Discours, Apprentissages – est une unité de recherche pluridisciplinaire de la Faculté des Sciences Humaines et Sociales-Sorbonne de l'Université Paris Descartes.

de ce domaine (Goasdoué, Vantourout & Bedoin, 2015, à paraître). Enfin, depuis maintenant cinq ans, plusieurs d'entre nous s'intéressent à l'évaluation en lecture et en compréhension de texte (Blanc, Goasdoué, Smith, Vantourout & Maury).

Au sein de l'activité des évaluateurs, nous distinguons plusieurs facettes : la conception d'épreuves et leur « correction », ainsi que la facette « communication » avec des travaux prenant en compte les livrets des élèves en primaire. Ce texte, comme notre intervention lors du séminaire national, portera sur l'activité de correction ou d'évaluation proprement dite.

Correction-évaluation et docimologie(s)

D'un point de vue historique, c'est dans le cadre de la docimologie puis de la docimologie expérimentale (cf., respectivement, Piéron, 1963, et Noizet & Caverni, 1978) que les premiers travaux consacrés aux examens et à l'évaluation des élèves ont été réalisés. Toutefois, dans ces perspectives, on ne s'est jamais véritablement intéressé à l'activité des correcteurs, mais uniquement à la note produite. Rappelons brièvement que, dans la première moitié du 20^e siècle, les docimologues ont fait le constat des divergences et écarts de notes entre correcteurs, puis que, entre 1960 et la fin des années 70, les docimologues expérimentaux ont montré l'existence de biais systématiques lors de la notation d'un lot de copies (effets d'ordre, de contraste et d'assimilation). La question traitée relève alors de la fidélité. L'importance accordée à celle-ci, aux dépens de la validité, a fait l'objet de critiques fondées, dont celles de Cardinet dès 1973 et de Bodin en 2006. Nous rejoignons ces auteurs et estimons que l'étude renouvelée de la validité, c'est-à-dire sous un angle didactique, peut constituer l'un des apports majeurs des didactiques disciplinaires au domaine de l'évaluation, cette étude pouvant concerner les épreuves (voir la contribution de Grugeon & Grapin dans ce séminaire ; Vantourout & Goasdoué, 2014) et les jugements (Goasdoué, Vantourout & Bedoin, 2015 à paraître). En quelque sorte, nos travaux prolongent ceux des docimologues expérimentaux tout en les renouvelant. En effet, contrairement à nos prédécesseurs, grâce aux apports des didactiques disciplinaires et de la psychologie des apprentissages, nous nous intéressons au fonctionnement cognitif des évaluateurs, aux processus de construction des jugements et prenons en compte les spécificités des contenus en jeu dans les évaluations. Cette prise en compte de ces organisateurs de l'activité d'évaluation nous a permis de montrer notamment que les divergences de notations ne recouvrent pas les divergences de jugements (voir infra).

REMISE EN CAUSE DU MODELE « REFERENT/REFERE »

Le modèle « référent/référé »

Les travaux en docimologie expérimentale ont abouti à la proposition d'un modèle explicatif (cf. Annexe 1) où les auteurs font figurer deux opérations cognitives qui caractérisent l'essentiel des comportements d'évaluation de productions scolaires : « l'opération de sélection des produits attendus parmi les produits possibles et l'opération de comparaison de la copie – produit réel – et du modèle de référence » (Noizet & Caverni, 1978, p. 116). En accord avec ce modèle, l'évaluation, d'un point de vue psychologique, est conçue et définie comme « une comparaison entre un référent et un référé » (cf. par exemple, Barbier, 1985 ; Hadji, 1992, 1997). Toute évaluation suppose donc la construction par l'évaluateur d'un référent, système d'attentes jugées légitimes (ou d'un ensemble des critères au nom desquels on va se prononcer), et d'un référé, « modèle réduit » de l'objet évalué, c'est-à-dire ce à partir de quoi on pourra porter le jugement de valeur (cf. Hadji, 1997, p. 42-45 ; 1992, p. 31-36). Pour Amigues et les

docimologues expérimentaux (1975), la tâche d'évaluation ne peut être exécutée sans que l'évaluateur ne dispose d'un modèle de référence, inscrit dans ses structures cognitives. C'est ce modèle « comparaison référent/référé » que retiennent très majoritairement les spécialistes de l'évaluation pour appréhender l'activité des évaluateurs. Élaboré à partir de tâches de corrections de lots de copies, il est convoqué indépendamment de formes d'évaluation et est en quelque sorte considéré comme universel.

Résultat d'une recherche

L'un des termes de cette comparaison, le référent (ou modèle de référence), principalement ses caractéristiques et sa construction a priori, ont été remis en cause : il a été montré que certains professeurs élaborent leur référent alors qu'ils évaluent et que ce dernier peut renfermer des réponses et solutions erronées (Vantourout, 2004 et 2007). Il est possible de faire un rapprochement entre ce phénomène évaluatif et l'évaluation « appréciative sans modèle prédéterminé » postulée par Ardoino et Berger (1986, cf. Hadji, 1989 et 1992), démarche dans laquelle le référent n'a pas été construit avant toute prise d'information : « Évaluer c'est se poser la question du sens de ce qui se fait (...). C'est interpréter, c'est-à-dire construire au cours de l'opération d'évaluation, un 'référant', qui loin d'être un préalable à l'identification du référé, est le produit de l'activité d'évaluation » (Hadji, 1992, pp. 37-38). Nos recherches ont permis de donner une consistance empirique à ces propositions et d'attester l'existence de tels phénomènes évaluatifs et de constater que, parmi les évaluateurs qui ne disposent pas de référent a priori, figurent ceux qui maîtrisent mal, voire pas du tout, les aspects disciplinaires du problème servant à évaluer (cf. également, Nabbout, 2006).

Double orientation didactique et apports de la didactique des mathématiques pour étudier l'activité des évaluateurs

Cette première recherche menée entre 2000 et 2004 s'inscrit très clairement dans une double orientation didactique : didactique des mathématiques et didactique professionnelle. Sur le plan méthodologique, sa principale source d'inspiration est la didactique professionnelle (Pastré, 2011). L'idée de « situations aménagées » (Vantourout, Goasdoué, Maury & Nabbout, 2012), qui sont un type particulier de simulations, est inspirée des ingénieries de la didactique professionnelle. Ces situations sont qualifiées « d'aménagées » au regard de la distance qui les séparent de la situation dite de « référence », c'est-à-dire avec celles qui existent sur le terrain et que l'on cherche à reproduire, lors de cette recherche des situations d'évaluation formative interactive (Allal, 1979). Dans nos situations aménagées, « ce ne sont pas les appariements de surface qui sont recherchés, mais prioritairement des équivalences fonctionnelles » (Vantourout & al., 2012, p. 191-192). Ces situations permettent, grâce à la manipulation de variables, de proposer des situations d'évaluation où sont modélisés des comportements d'élèves représentatifs d'un point de vue didactique. « On associe au contrôle des variables une analyse de l'activité qui offre une description du comportement des évaluateurs qui s'avère plus riche que celle habituellement conduite dans un cadre expérimental où, pour les travaux sur l'évaluation, l'on prend uniquement en compte la note attribuée par l'évaluateur » (Idem). Avec les situations aménagées, nous réduisons « la complexité des situations de référence dans des proportions que nous estimons « raisonnables », au sens où, du point de vue du fonctionnement cognitif des évaluateurs, nous faisons l'hypothèse qu'il existe des similarités importantes entre les deux situations » (Vantourout & al., 2012, p. 195).

L'autre spécificité de ces situations, par rapport aux copies construites (Noizet & Caverni, 1978) est l'appui sur des travaux issus des « didactiques disciplinaires ». L'ancrage dans le champ de la didactique des mathématiques se manifeste à deux niveaux. D'abord, certainement

le plus important, celui qui renvoie à l'élaboration du matériel expérimental (cf. Annexe 2) présenté lors des situations aménagées². Il s'agit des réponses et des verbalisations de binômes d'élèves « fictifs³ » ayant résolu des problèmes de proportionnalité impliquant des représentations graphiques. Le matériel à évaluer ne peut faire l'objet d'un avis tranché. Son élaboration implique des analyses a priori, réalisées selon deux plans. Elles concernent, d'une part, les problèmes retenus, c'est-à-dire les tâches censées être proposées aux élèves fictifs, et, d'autre part, l'invention des productions et des comportements attribués à ces élèves. Ces analyses a priori font intervenir des connaissances en didactique relatives à la proportionnalité et aux graphiques. La présence de graphiques vise à engendrer des situations d'évaluation rendues relativement complexes en raison des difficultés que soulève l'utilisation de ce type de représentations (Maury, 2002). Bref, nous accordons de l'importance aux contenus mathématiques impliqués dans les productions à évaluer, ainsi qu'aux connaissances sur les apprentissages relatifs à ces contenus. La constitution du matériel expérimental fait appel à des éléments théoriques qui sous-tendent l'analyse a priori de la tâche et l'analyse de l'activité de l'évaluateur. Plus précisément, la construction du matériel (copies, productions d'élèves, échanges verbaux) repose sur des hypothèses quant aux conceptions et raisonnements des élèves et des évaluateurs. Le fait de construire les productions à évaluer nous permet de retenir et d'aménager « des comportements d'élèves qui correspondent à des procédures caractéristiques, en faisant de sorte que celles-ci véhiculent également des éléments pouvant favoriser l'apparition chez certains évaluateurs de conceptions relatives aux notions mathématiques en jeu, et, éventuellement, d'autres conceptions plus générales relatives à l'évaluation » (Vantourout & al., 2012, p. 199). Le second niveau de l'ancrage didactique « touche à des aspects plus généraux de la didactique des mathématiques (que l'on retrouve d'ailleurs en didactique des sciences) qui consistent, dans une double inspiration constructiviste et bachelardienne, à considérer l'erreur comme un indicateur du fonctionnement d'une connaissance et donc à s'attacher, à partir des observables, à identifier et à analyser les procédures et leur signification au niveau des connaissances. En ce sens, notre travail peut être rapproché des recherches en didactique qui, sur la base de cette conception de l'erreur, s'intéressent, plus ou moins directement, à la manière dont les enseignants de mathématiques prennent en compte les erreurs et les réussites des élèves » (Vantourout & Maury, 2006, p. 772). Soulignons que c'est, entre autres, autour de la notion d'erreur et de son analyse que se réalise la jonction entre évaluation formative et didactique.

Analyse de l'activité des évaluateurs : le rôle des connaissances

Bien que le terme « connaissances » figure dans le modèle « référent/référé » (cf. Annexe 1, « Corps de connaissances »), les connaissances des évaluateurs n'ont jamais été prises en compte par les docimologues et docimologues expérimentaux qui au fond se sont peu intéressés au fonctionnement cognitif des évaluateurs si ce n'est en recherchant des déterminants très généraux. Or, les connaissances mobilisées, ou non, constituent pour nous des déterminants de l'activité d'évaluation internes à l'évaluateur. Précisons que l'intérêt porté aux connaissances

² Lors des expérimentations, les situations aménagées sont présentées aux évaluateurs via un CD-Rom. Phase 1, ils travaillent seuls face à un ordinateur (un dispositif permet d'enregistrer leur activité) et doivent répondre à la consigne suivante : « En quoi les productions et réponses de ces élèves sont-elles satisfaisantes et/ou non satisfaisantes ? ». Phase 2, en binôme, ils doivent discuter et s'accorder sur une évaluation commune. Phase 3, ils participent individuellement à un débriefing. Toutes les phases sont filmées. Transcription des verbalisations des phases 2 et 3.

³ Les professeurs évaluent des travaux attribués à des élèves absents et qu'ils ne connaissent pas. Ces travaux ont été élaborés, pour les besoins de l'expérimentation, à la suite d'observations de binômes d'élèves « réels » auxquels nous avons demandé de résoudre le même problème. Notons que ces élèves ne sont représentés que par des traces de leur activité (simulation fonctionnelle).

de l'évaluateur n'a pas pour finalité de juger de sa légitimité en tant qu'évaluateur, mais de mieux comprendre le rôle de celles-ci dans les processus d'évaluation.

En étudiant l'activité de professeurs⁴ confrontés aux situations aménagées, nous avons identifié trois grands pôles de connaissances : pôle « disciplinaire », « évaluation » et « socio-psycho-pédagogique » (Vantourout & Maury, 2006). Au sein du pôle disciplinaire, nous distinguons les connaissances liées aux mathématiques et aux représentations graphiques et les connaissances liées à la didactique des mathématiques⁵. Ces dernières s'observent (sont inférées) quand les enseignants se concentrent sur l'identification et l'étude des procédures mises en œuvre par les élèves. Si la majorité des enseignants observés marque de l'intérêt pour les procédures et reconnaît les situations où une analyse didactique est pertinente (16 enseignants sur 18, mais avec une différence quantitative individuelle très importante), ce n'est qu'exceptionnellement qu'un professeur parvient à conduire une véritable analyse didactique, c'est-à-dire à mener une analyse explicative de l'activité de l'élève en mobilisant des connaissances valides liées à la didactique des mathématiques. Ces dernières représentent moins de 0,25% des occurrences de connaissances que nous avons relevées, bien que ces professeurs, anciens étudiants et stagiaires en IUFM, aient reçu une formation en didactique des mathématiques. Ce résultat renvoie à la question de la pragmatisation des savoirs en connaissances opérationnelles qu'aborde la didactique professionnelle (Pastré, 2011).

De fait, comme cela apparaît dans les résultats, mener une véritable analyse didactique requiert la mobilisation de connaissances dites de « niveau conceptuel supérieur » (Vantourout & Maury, 2006). Au sein des connaissances disciplinaires liées aux mathématiques et aux représentations graphiques, nous distinguons les connaissances de ce niveau de celles dites « de base ». Ces dernières, aux occurrences très nombreuses, sont mobilisées par tous les évaluateurs : elles ont comme principale caractéristique de permettre de « faire » le problème, elles suffisent pour répondre à l'ensemble des questions contenues dans les énoncés de l'expérimentation. Parmi leurs fonctions, on trouve celle de permettre de se prononcer sur la justesse, la fausseté, les imperfections et les erreurs des réponses numériques ou graphiques ou encore celle de permettre l'énonciation et la reconnaissance de notions en jeu dans le problème. Les connaissances de niveau conceptuel supérieur permettent de modéliser la situation, de comprendre au plan mathématique et d'expliquer les procédures des élèves. Elles permettent de traiter la relation « signifiant/signifié » (par exemple, le lien entre la pente d'une droite et le coefficient directeur d'une fonction linéaire ou affine) ou en encore d'amener une solution « experte » (par exemple, un graphique qui représente dans un problème de concentrations à la fois les quantités et les proportions). Enfin, nous avons relevé des connaissances « erronées ou inexactes », mobilisées uniquement par des professeurs des écoles de formation « non-scientifique »⁶.

Le principal résultat de cette recherche est de montrer que :

- des jugements évaluatifs identiques ou (très) proches peuvent reposer sur des connaissances différentes, dont certaines erronées ou inexactes ;
- des connaissances analogues peuvent conduire à des jugements évaluatifs différents, voire opposés.

⁴ Il s'agit de 18 enseignants en fin de formation initiale (IUFM), 12 professeurs des écoles et 6 professeurs de mathématiques (PLC).

⁵ S'ajoutent des connaissances curriculaires dont nous ne parlerons pas ici.

⁶ Les professeurs des écoles de formation « scientifique » possèdent au minimum un baccalauréat scientifique. Les PLC sont tous considérés de formation « scientifique ».

Nous allons illustrer cela brièvement à partir de l'évaluation d'une réponse graphique présentée lors d'une situation aménagée. Il s'agit d'un problème de location de cassettes vidéo, avec trois tarifs différents, dont deux avec abonnement. Les élèves devaient, entre autres, compléter un graphique (sur lequel sont déjà représentés deux des tarifs) en représentant la formule de tarification avec abonnement à 250 francs et 10 francs par cassette louée (cf. Annexe 3). Les élèves d'un binôme ne font pas débiter leur tracé au point (0, 250), mais à l'origine du repère, puis placent le point (1, 260). Sur le plan des jugements évaluatifs, dix-sept des dix-huit professeurs s'accordent pour reprocher aux élèves de faire débiter leur graphique à l'origine. Seul l'un d'entre eux hésite à émettre une critique sur ce point : il défend l'idée selon laquelle faire débiter le graphique à l'origine dénote d'une logique concrète (cf. infra enseignant « F »). Nous présentons ici uniquement quelques extraits d'échanges entre enseignants qui doivent s'accorder sur une évaluation commune. Pour l'analyse détaillée de la diversité des connaissances mobilisées par les évaluateurs nous renvoyons à Vantourout, 2004.

Le jugement des enseignants A et B laisse apparaître une divergence, avec aucune véritable argumentation. Finalement B suivra l'avis de A. Ce qui emporte la décision est la référence aux autres tracés (A2). L'intervention A2 donne aussi une indication sur la manière dont est construit le référent et sur sa solidité.

A1 : Donc, il y a une reprise du tracé graphique, mais je pense qu'il y a quand même l'erreur, car ils débutent de l'origine zéro

B1 : Ce n'est pas une erreur pour moi, ils débutent de l'origine zéro, c'est 250 francs en n'ayant pris aucune cassette, donc, ce n'est pas une erreur

A2 : Comme dans les autres tracés, on prend en compte directement l'abonnement

Les enseignants C et D sont confrontés à un dilemme que nous formulerons ainsi : comment gérer la dimension « réaliste » ou la référence au contexte qui se trouve en décalage avec le recours au modèle des fonctions affines et linéaires ?

C1 : Oui, mais tu ne veux pas louer de cassette tu ne payes pas

D1 : Si tu te casses une jambe et que tu ne peux plus aller à ton truc de vidéo, tu n'auras toujours pas de cassette de louée et puis tu auras quand même payé l'abonnement pour un an, tu ne crois pas ?

C2 : Oui, mais je pense que tu payes l'abonnement à partir du moment où tu loues une cassette ... Tu ne vas pas payer l'abonnement sans louer de cassette [Rires]

D2 : Attends, moi je commencerais à 1 ! Mais si attends, alors pourquoi ? Parce que ... Oui tu ne loues pas de cassette, tu ne payes pas d'abonnement ... Ouais, enfin si c'est vrai ce que tu dis

C3 (faisant référence à une autre formule et à son tracé sur le graphique) : La formule, c'est 100 plus 30 fois x quoi $[100 + 30x]$, x étant le nombre de cassettes ... donc pour x égal zéro, ça fait 100

Les enseignants E et F sont confrontés au même dilemme, mais l'un d'entre eux, F, parvient à prendre en considération la situation concrète évoquée dans l'énoncé et peut s'appuyer sur cette modélisation via les fonctions dans le cadre graphique.

E1 : Alors pour le graphique, il y a un problème pour le prix quand on prend zéro cassette, ils ont mis, ils ont pris l'origine aussi, ils n'ont pas décalé à 250 là

F1 : Ben qui est faux, moi je ne suis pas entièrement d'accord pour dire que c'est faux parce que ça se défend si t'as pas l'intention de louer de cassettes tu ne prends pas d'abonnement, donc finalement ça ne coûte rien

E2 : Ouais, mais enfin l'abonnement tu le payes après si tu veux zéro cassette tu as quand même payé tes 250 francs, si tu tombes malade ou je ne sais pas quoi

F2 : Cela dit souvent le jour où tu vas prendre ton abonnement tu prends une cassette, donc si tu n'as pas pris de cassette, c'est que tu n'avais pas ... Ouais enfin, moi je dis que ça se défend, zéro, zéro (0,0) [...]

E3 : Ouais ... Je ne sais pas si tu n'analyses pas un peu trop ?

F3 : C'est vrai qu'on part du schéma des fonctions affines, mais justement je n'ai pas fait attention [...] Donc moi, enfin je préfère privilégier le sens ... Pour moi, qu'ils aient mis le point là ou là, ce n'est pas très important, je ne

trouve pas ça insatisfaisant qu'ils aient mis le zéro, zéro, je ne trouve pas que ça soit un problème

Outre la question de la diversité connaissances mobilisées, se pose également, dans une perspective d'évaluation formative, celle de la diversité des feedbacks adressés aux élèves.

LA CORRECTION DE DISSERTATION EN SCIENCES ECONOMIQUE ET SOCIALE

La seconde illustration des options théoriques et méthodologiques que nous défendons pour l'étude de l'activité des évaluateurs vient d'une recherche menée sur la correction de dissertations en sciences économiques et sociales (Vantourout & Goasdoué, 2011). Bien que changeant de domaine disciplinaire, ce travail s'inscrit dans le prolongement des travaux menés en mathématiques, tout explorant des questions sur les jugements des enseignants moins saillantes dans la recherche précédente. Nous avons en effet choisi d'aller vers une épreuve plus ouverte, au sens où l'ensemble des solutions également valables est plus important, vers une matière où il est moins facile de trancher sur le caractère juste ou faux d'une réponse. Ces deux caractéristiques permettent par ailleurs de réinterroger la pertinence de la comparaison « référent/référé », puisqu'évidemment aucun enseignant ne s'engage dans la correction avec une « dissertation modèle » en tête. L'absence d'un référent identifiable ne signifie pas pour autant l'absence d'attentes très précises, mais nous les qualifierions alors plutôt de réseaux d'attentes que de référent.

Dans la continuité des travaux menés précédemment, cette enquête s'appuie sur une analyse détaillée des verbalisations d'enseignants au cours de leur correction de copies. Il s'agit à nouveau d'une situation aménagée, car ce travail d'explicitation de l'activité de correction est, dans une certaine mesure, artificiel tout en étant bien un témoignage du fonctionnement cognitif du correcteur. Après un temps de « préparation » du sujet de la dissertation, nous demandions aux enseignants de verbaliser tous les commentaires qu'ils jugeaient nécessaires pour que l'on comprenne leur activité de correction. Toujours dans l'idée de prolonger dans une autre perspective les travaux en docimologie, nous avons sélectionné trois copies dont une moins bonne et deux moyennes, car ces dernières sont identifiées de longue date (cf. Laugier, Weinberg 1936, cités par Piéron 1963) comme suscitant un maximum de divergences. Tous ces choix, une tâche ouverte, des réponses pas nécessairement tranchées, des copies moyennes, offrent une sorte de cas « extrême » pour mettre à l'épreuve les principaux résultats et hypothèses de la docimologie : les divergences entre évaluateurs et la comparaison « référent/référé ». Enfin pour évacuer l'effet possible du niveau de formation et de maîtrise des contenus évalués, nous nous sommes adressés à des enseignants expérimentés et presque tous agrégés en Sciences Economiques et Sociales.

Le premier constat quantitatif des divergences de notation est conforme à tout ce qui a été écrit à ce sujet depuis les travaux pionniers des docimologues des années 1930, la même copie a reçu la note de 8/20 ou 16/20 selon le correcteur. L'intérêt majeur de ce travail ne réside évidemment pas dans la répétition de ce constat, mais dans l'analyse du lien que l'enquête permet de faire entre notation et jugement. Si on compare les correcteurs par paires, on peut par exemple remarquer que des écarts plus faibles en nombre de points ne s'accompagnent pas nécessairement d'une convergence accrue de jugements simplement caractérisés par leur valence positive ou négative (cf. Vantourout & Goasdoué, 2010 ; Goasdoué, Vantourout & Bedoin (à paraître) pour plus de détail). Inversement, des enseignants ayant davantage de jugements communs à propos des mêmes éléments mettent des notes plus divergentes. Ce premier constat conduit évidemment à différencier divergences de jugement et divergences de notation. La nature de l'épreuve et les techniques de correction pourraient être invoquées pour expliquer ces décalages entre jugements et notation. L'adoption d'un barème commun pourrait

être vu comme une solution à ces divergences or il n'en est rien (Vantourout & Goasdoué, 2010). En effet, parmi les enseignants interrogés, plusieurs corrigeaient à l'aide d'une grille commune très détaillée à laquelle ils confiaient en quelque sorte le choix de la note qui n'était plus qu'une somme de l'ensemble des points accordés par items de la grille. Les divergences de notation entre ces correcteurs n'étaient pas inférieures à celle avec d'autres correcteurs corrigeant sans grille⁷. Une analyse plus approfondie, permet de constater que si de nombreux enseignants s'intéressaient aux mêmes passages de la dissertation (80% des verbalisations sont concentrées sur les mêmes phrases), ils n'en tirent pas toujours les mêmes conséquences. On retrouve ici les constats faits dans l'étude sur les représentations graphiques et la proportionnalité, la convocation d'une même connaissance ne conduit pas nécessairement au même jugement évaluatif comme en témoignent les extraits suivants :

A la lecture de ces deux phrases de la copie : « *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 adhérents.* »

- Un premier enseignant dit : « *Très bon mécanisme, très bonne argumentation. Chose 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* »
- Un second affirme au contraire que : « *Il faut définir le terme institution ; On parle plutôt d'institutionnalisation des syndicats* » et ajoute sur la copie : « *Ne sont plus des salariés présents dans l'entreprise. Donner un exemple pour construire la démonstration* »

Les deux correcteurs mobilisent la même notion d'institutionnalisation. Toutefois, seul le premier fait crédit à l'élève de la connaissance que les salariés ne sont plus présents dans l'entreprise, mais toujours rémunérés par elle. Le second correcteur pointe l'ambiguïté de la phrase, ce qui débouche évidemment sur des jugements différents.

Cet exemple montre à quel point la correction est une lecture singulière et approfondie de la copie où l'élève est présent à travers son texte. Le texte n'est pas pris comme une simple réponse, mais plutôt comme le symptôme d'un fonctionnement cognitif. L'intention ou la finalité de lecture de correction est implicitement diagnostique, c'est, comme l'affirmait Amigues (1996), une forme de dialogue différé. La correction et la notation d'une dissertation ne sont clairement pas une quantification des qualités de la copie. On perçoit ici le décalage profond entre la nature de cette activité et les critères d'évaluation de celle-ci imposés dans les travaux en docimologie. La fidélité, bien que posant des problèmes concrets très importants, notamment pour les examens, ne saurait être le seul critère de validité d'une évaluation. Sur l'ensemble des jugements produits par les enseignants presque aucun ne pourrait être qualifié d'erroné au sens décrit dans la première étude. Tous s'appuient sur des connaissances et des éléments bien clairs et légitimement relevés de la copie. En ce sens, on se démarque nettement des conclusions des docimologues qui tendent à disqualifier la validité d'une épreuve comme la dissertation.

Outre la question des divergences, cette étude contribue également à remettre en cause la place de la comparaison « référent/référé » dans la description de l'activité des évaluateurs. Remettre en cause la place accordée à ces processus ne conduit cependant pas à dénier toute cohérence à l'activité de correction et à supposer que l'enseignant procède par induction sans attente préalable. La technique de verbalisation concomitante à l'activité de correction permet au contraire de suivre l'évolution de la représentation que l'enseignant se fait de la copie. L'évocation fréquente de la cohérence de la copie, ainsi que la formation régulière d'anticipations sur la suite de la copie montre la richesse de cette activité qui ne saurait être

⁷ Toutefois, les correcteurs qui utilisent la grille auraient tendance à donner plus de points aux copies « moyennes ». L'utilisation de la grille pourrait servir de « garde-fou » en empêchant de sous-noter ce type de copies, comme le font certains correcteurs (Vantourout & Goasdoué, 2011).

réduite à une comparaison. A contre-pied des docimologues expérimentaux qui considéraient l'évaluation comme un cas particulier de perception (notamment en référence à la perception de grandeurs physiques), ce constat nous a conduit à considérer la lecture de correction plutôt comme un cas particulier de compréhension de texte. Le lecteur correcteur cherche en permanence construire une représentation cohérente du texte lu et, pour y parvenir, il produit de nombreuses inférences sur divers objets tels que les contenus, la structure du devoir ou les qualités « présumées » de l'élève.

POUR UNE APPROCHE PSYCHO-DIDACTIQUE DES EVALUATIONS

L'intention initiale des travaux présentés ici était en quelque sorte de poursuivre le projet déjà formulé, mais jamais vraiment réalisé dans les années 1930, d'étudier « la psychologie des examinateurs » (cf. In Piéron, 1963). Notre but était également de décrire les déterminants de l'activité évaluative notamment pour penser la formation à l'évaluation. Si nous partageons avec les docimologues ce postulat que mieux comprendre les examinateurs pourrait faire évoluer les pratiques d'évaluation, les chemins que nous empruntons sont nettement différents. Implicitement nous considérons que la portée diagnostique des évaluations est une dimension essentielle dans le champ scolaire. Si la psychométrie et ses développements récents ont permis d'améliorer la finesse discriminative des tests, la signification des différences constatées ne peut avoir de sens, selon nous, qu'au regard d'une analyse des épreuves et de l'activité de réponse des élèves.

Ce parti pris impose naturellement le recours aux didactiques disciplinaires et à la psychologie cognitive. Ainsi nous défendons, à l'instar de travaux plus centrés sur l'apprentissage (Maury, 2001), une approche psycho-didactique des évaluations (Vantourout & Goasdoué, 2014). Cette conjonction théorique et méthodologique permet d'une part de rendre compte de la spécificité des contenus évalués et d'autre part d'analyser les différents plans de l'activité évaluative : conception d'épreuve, réponse des évalués et correction. Bien que portant sur l'activité de correction, les travaux présentés ici permettent de penser plus généralement la question de la ou des validités des épreuves scolaires. S'assurer de la validité des épreuves ne saurait être résumé à des choix méthodologiques judicieux. De la même manière, la validité épreuves scolaires ne saurait se résumer à leurs qualités métriques de fidélité ou encore de finesse discriminative.

L'approche que nous défendons, bien que radicalement différente, n'est pas en conflit avec les conceptions psychométriques de la validité. À l'instar de notre travail sur la notation qui conduit à différencier validité de l'évaluation et validité des notes, nous défendons que la validité des épreuves ne peut être seulement abordée à travers une analyse des scores. Le constat d'un ensemble de réussites ou d'échecs à un ensemble d'items ne peut être signifiant qu'au travers d'une analyse a priori des items et complémentirement d'une analyse de l'activité de réponse des évalués ou des correcteurs. Ainsi de même que les connaissances didactiques sont essentielles pour garantir la validité des jugements lors de corrections, les choix méthodologiques dans la conception d'épreuves devraient être guidés par ces mêmes préoccupations. Pour compléter et clore l'éventail des questions qui nous semblent essentielles en évaluation, il reste à évoquer les considérations curriculaires qui ne sont pas traitées dans notre approche psycho-didactique qui a d'abord été conçue pour analyser l'activité des évaluateurs dans le cadre d'évaluations internes. La notion de validité épistémologique-didactique proposée et développée dans cette séance de séminaire (cf. Grugeon & Grapin) est ainsi un complément essentiel notamment pour l'analyse de la validité dans la conception d'épreuve.

REFERENCES BIBLIOGRAPHIQUES

- ALLAL L. (1979). Stratégies d'évaluation formative : conceptions psycho-pédagogiques et modalités d'application. In L. ALLAL, J. CARDINET ET P. PERRENOUD (Eds.) *L'évaluation formative dans un enseignement différencié*, 130-157. Berne : Peter Lang.
- AMIGUES R., BONNIOL J.-J., CAVERNI J.-P., FABRE J.-M., NOIZET G. (1975). Le comportement d'évaluation de productions scolaires : à la recherche d'un modèle explicatif. *Bulletin de psychologie*, 28, 793-799.
- AMIGUES R., & ZERBATO-POUDOU, M.-T. (1996). *Les pratiques scolaires d'apprentissage et d'évaluation*. Paris : Dunod.
- BARBIER J.-M. (1985). *L'évaluation en formation*. Paris : PUF.
- BODIN A. (2006). Les mathématiques face aux évaluations nationales et internationales. Repères IREM n° 65, 55-89. www.univ-irem.fr/commissions/reperes/consulter/65bodin.pdf
- CARDINET J. (1973/1988). L'adaptation des tests aux finalités de l'évaluation. In J. Cardinet *Évaluation scolaire et mesure* (pp. 24-59). Bruxelles : De Boeck.
- GOASDOUE R., VANTOUROUT, M., BEDOIN D. (à paraître). La correction de dissertations : nouveau regard sur la construction des jugements des évaluateurs. In L. MOTTIER LOPEZ & W. TESSARO, *Le jugement professionnel des enseignants*. Berne : Peter Lang.
- HADJI C. (1997). *L'évaluation démystifiée*. Paris : ESF.
- HADJI C. (1992). *L'évaluation des actions éducatives*. Paris : PUF.
- HADJI C. (1989). *L'évaluation, règle du jeu*. Paris : ESF.
- MAURY S. (2001). Didactique des mathématiques et psychologie cognitive : un regard comparatif sur trois approches psychologiques. *Revue française de pédagogie* n° 137, 85-93.
- MAURY S. (2002). A look at some studies on learning and processing graphic information, based on bertin's theory. In F. HITT (dir.), *Representations and Mathematics Visualization* (p. 297-309). Mexico : Cinestav - IPM.
- NABBOU M. (2006). Enseignement des probabilités en terminale au Liban : études des représentations et des pratiques dans des situations aménagées. Thèse non-publiée. Université René Descartes – Paris 5 Sorbonne.
- NOIZET, G., CAVERNI, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : PUF.
- PASTRE P. (2011). *La didactique professionnelle*. Paris : PUF.
- PIERON, H. 1963. *Docimologie et examens*. Paris : PUF.
- RODITI E., SALLES F. (2015). Nouvelles analyses de l'enquête PISA 2012 en mathématiques. *Éducation & Formations*, n° 86-87, 235-257.
- VANTOUROUT M., GOASDOUE R. (2014). Approches et validité psycho-didactique des évaluations. *Éducation & Formation – e-302*, 139-155.
- VANTOUROUT M., GOASDOUE R., MAURY S., NABBOU M. (2012). A la frontière entre l'écologique et l'expérimental : des situations aménagées pour l'étude de l'activité évaluative en mathématiques. In M. Altet, M. Bru, C. Blanchard-Laville (Eds.) *Observer les pratiques enseignantes* (pp.191-204). Paris : L'harmattan.
- VANTOUROUT M., GOASDOUE R. (2011). Correction de dissertations en SES. *Idées*, 163, 71-77.

VANTOUROUT M., GOASDOUÉ, R. (2010). Correction de dissertations : analyse de l'activité de professeurs engagés dans une approche pas compétences. In Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF), Université de Genève, 13-16 septembre. 12 pages (en ligne).

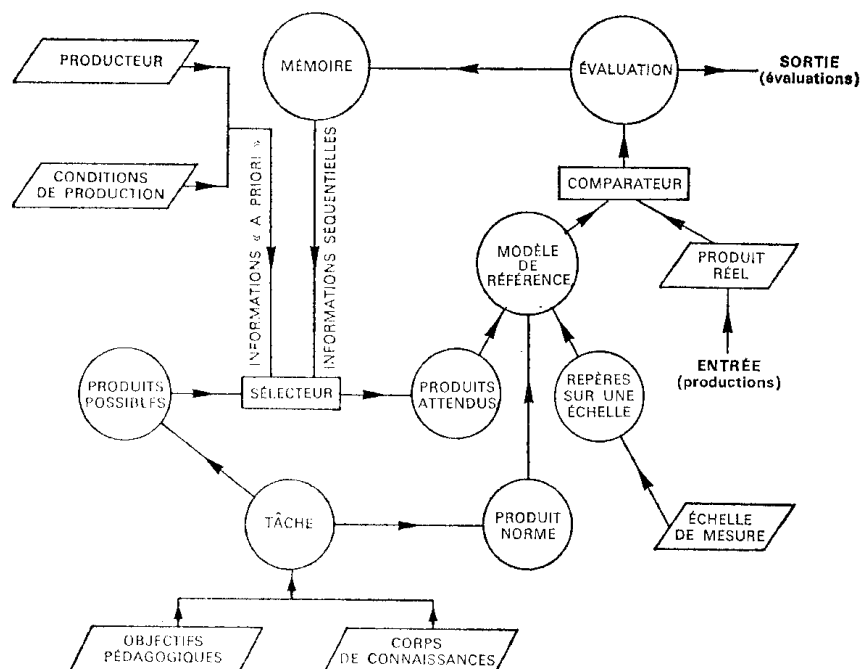
VANTOUROUT M. (2007). Étude de l'activité évaluative de professeurs stagiaires confrontés à des productions d'élèves en mathématiques : quel référent pour l'évaluateur ? *Mesure et évaluation en éducation*, vol. 30, n°3, 29-58.

VANTOUROUT M., MAURY S. (2006). Quelques résultats relatifs aux connaissances disciplinaires de professeurs stagiaires dans des situations simulées d'évaluation de productions d'élèves en mathématiques. *Revue des sciences de l'éducation*, vol. 32, n°3, 759-782.

VANTOUROUT M. (2004). Étude de l'activité et des compétences de professeurs des écoles et de professeurs de mathématiques dans des situations « simulées » d'évaluation à visée formative en mathématiques. Thèse non-publiée. Université René Descartes – Paris 5 Sorbonne.

ANNEXES

ANNEXE 1 : MODELE EXPLICATIF DES COMPORTEMENTS D'EVALUATION DE PRODUCTIONS SCOLAIRES – NOIZET & CAVERNI, 1978, p. 115



ANNEXE 2 : TYPES DE PAGES PRESENTEES LORS DES SITUATIONS AMENAGEES (PROBLEME : LE CYCLISTE)

Réponse écrite Alexis et David

3) Nous ne sommes pas d'accord
Méthode de David : si en 50 min il fait 10 km alors qu'il devrait faire 12,5 km car en 10 min il en fait 10 km en 40 min
Ensuite il revient à une allure régulière. En 120 min il doit en faire 30 car en 60 min il a parcouru 15 km donc en 30 min il parcourt 7,5 km. En 90 min il parcourt 22,5 km.

40	50	60	80	100	120
10	12,5	15	20	25	30

[Tableau construit par David.]

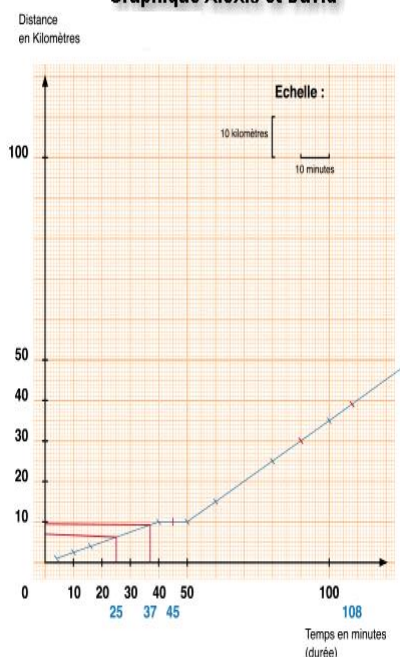
Distance parcourue en 90 min
 $120:80 = 90 / 30 = 7,5$ / $90 = 22,5$ km

Méthode d'Alexis

Distance parcourue en 90 min
 En 90 min il fera 30 km
 Car si de 80 min à 100 min il fait 10 km en 90 min il fait 30 km

5) En traçant avec la règle on découvre qu'il fait 7 km en 25 min de même pour 37/45/108
 37 min = 9,5 km
 45 min = 10 km
 108 min = 39 km

Graphique Alexis et David



Les évaluateurs ont également accès à la transcription des échanges entre Alexis et David alors qu'ils résolvent en binôme le problème.

ANNEXE 3 : PROBLEME « LOCATION DE CASSETTES VIDEO »

PRODUCTION GRAPHIQUE D'UN BINOME D'ELEVES

(Les tarifs « a » et « b », respectivement en rouge et bleu, sont déjà placés sur le graphique ; les élèves ne doivent représenter que le tarif « c », en noir et jaune).

