

24^e Colloque International de l'ADMEE-Europe

L'évaluation des compétences en milieu
scolaire et en milieu professionnel

11-13 janvier 2012, Luxembourg

Tables rondes

	Titre	Participants
Table ronde 2 : Comment concevoir une évaluation diagnostique des compétences scolaires ?	COMMENT ASSURER L'ÉVALUATION DIAGNOSTIQUE DES COMPÉTENCES SCOLAIRES ?	Nathalie Loye
	COMMENT ASSURER L'ÉVALUATION DIAGNOSTIQUE DES COMPÉTENCES SCOLAIRES ?	Annick Fagnant & Christophe Dierendonck
	L'ÉVALUATION DIAGNOSTIQUE DES COMPÉTENCES SCOLAIRES : QUELQUES ÉLÉMENTS À CLARIFIER	Géry Marcoux
Table ronde 3 : Que peut apporter la didactique professionnelle à l'évaluation des compétences ?	DIDACTIQUE PROFESSIONNELLE ET ÉVALUATION : LA VOIX DE L'ÉMANCIPATION	Marie-France Baroth
	QUE PEUT APPORTER LA DIDACTIQUE PROFESSIONNELLE À L'ÉVALUATION DES COMPÉTENCES ?	Florent Chenu

TABLE RONDE : COMMENT ASSURER L'  VALUATION DIAGNOSTIQUE DES
COMP  TENCES SCOLAIRES ?

Annick Fagnant * et Christophe Dierendonck**

* Universit   de Li  ge (Belgique) – afagnant@ulg.be

** Universit   du Luxembourg (Luxembourg) – christophe.dierendonck@uni.lu

Mots-cl  s : T  che complexe – Evaluation diagnostique - Math  matiques – Enqu  te    large   chelle

R  sum  . Le texte qui suit se propose d'apporter quelques   l  ments de r  flexions relatifs    la probl  matique de l'  valuation diagnostique des comp  tences scolaires. Il est structur   autour des trois questions guides propos  es par Pascal Ndinga, organisateur de la table ronde : « Qu'est-ce qu'une t  che ou une situation complexe ? » ; « Dans quelle mesure le d  coupage de la t  che complexe dans le mod  le de Rey et al. (2003) r  pond-il aux attentes en mati  re d'  valuation diagnostique des comp  tences ? » et « Comment assurer la fiabilit   et la validit   des informations recueillies au cours d'une   valuation de comp  tences ? ». En vue de concr  tiser ces questionnements, le domaine des math  matiques sert de fil rouge    notre r  flexion.

1. Qu'est-ce qu'une t  che ou une situation complexe ?

La notion de « t  che complexe » (ou de situation complexe) est intimement li  e au concept de comp  tence. Cependant, ce concept de comp  tence demeure un concept flou, r  guli  rement critiqu   pour sa polys  mie et ses fondements th  oriques apparemment fragiles. Il est d  s lors tr  s difficile de d  finir un concept par un autre qui s'av  re mal assur  .

Dans la multitude de d  finitions qui existent autour de la notion de comp  tence (Beckers, 2005 ; Carette, 2007 ; Crahay, 2006 ; De Ketele & G  rard, 2005 ; Rey, Carette, Defrance & Kahn, 2003 ; Scallon, 2007 ; Tardif, 2004), on parvient cependant    circonscrire certains   l  ments communs qui d  finissent comme comp  tents les individus capables de mobiliser, de fa  on int  gr  e, diff  rentes ressources pour accomplir une t  che. Cela dit, - et c'est peut-  tre une des raisons du flou conceptuel qui existe autour de la notion - rares sont les auteurs ayant clarifi   les nuances s  mantiques qui d  coulent du d  terminant d  fini ou ind  fini (la, les, une, des) associ   dans le discours au terme « comp  tence ». L'expression «   valuer la comp  tence » peut en effet   tre compris comme le fait d'  valuer une unit   curriculaire ou disciplinaire particuli  re (par exemple, la comp  tence « traiter de l'information » au sein d'une liste de comp  tences en lecture) ou comme le fait de placer un individu face    une t  che pr  cise et d'appr  cier ses comportements et la r  ponse qu'il donne face    la situation. Il y aurait donc d'un c  t   les comp  tences d  finies dans le curriculum et de l'autre c  t   la comp  tence d  montr  e effectivement par les individus dans les situations d'  valuation.

Dans la suite du texte, nous allons discuter bri  vement ce qui constitue le c  ur de la premi  re question guide,    savoir, la notion de « complexit   ». Dans un second temps, nous nous int  resserons sp  cifiquement aux comp  tences math  matiques et envisagerons la question de l'  valuation des comp  tences math  matiques dans l'  tude internationale PISA.

Actes du 24^e colloque de l'Adm  e-Europe
L'  valuation des comp  tences en milieu scolaire et en milieu professionnel

1.1. Mobiliser, de fa  on int  gr  e, diff  rentes ressources pour accomplir une t  che

Dans toutes les d  finitions propos  es, les « ressources » font r  f  rence aux diff  rents types de connaissances dont doit disposer le sujet : connaissances d  claratives (savoirs ou connaissances conceptuelles), connaissances proc  durales (savoir-faire ou habilit  s) et connaissances conatives (savoir-  tre ou attitudes). Certains (voir notamment Beckers, 2005 ; Tardif, 2004) ajoutent une dimension en mentionnant conjointement la mobilisation de ressources internes et l'utilisation ad  quate de ressources externes (documents divers, dictionnaires, outils de calcul, personnes ressources,...). L'utilisation de ressources externes pour d  montrer la comp  tence d'un individu est un   l  ment important sur lequel nous reviendrons    la fin du texte. Toutefois, pour r  pondre    la premi  re question guide, nous avons pris le parti de chercher    comprendre « o   commence la complexit   » lorsque l'on s'int  resse uniquement    la mobilisation des ressources internes.

Quelques d  finitions (voir notamment Scallon, 2007) ajoutent encore aux ressources internes les connaissances conditionnelles qui guident le choix des ressources    mobiliser (quand et pourquoi utiliser telle ou telle connaissance dans telle situation ?). De notre point de vue, l'int  gration des connaissances conditionnelles pr  sente une certaine redondance puisqu'il s'agit alors de mobiliser des connaissances dont la fonction est de guider la mobilisation... Cette notion de « mobilisation » est incontournable puisqu'elle permet de distinguer les t  ches permettant d'  valuer des comp  tences, de t  ches qui permettraient davantage d'  valuer l'application de proc  dures.

L'  l  ment « t  che » (ou « situation ») est   galement un   l  ment central dans la mesure o   la comp  tence semblerait devoir   tre d  finie par la ou les t  ches (cf. « familles de t  ches », Beckers, 2005) qui sont utilis  es pour l'  valuer. Si chacun semble s'accorder sur l'id  e que la t  che est une activit   finalis  e (qui a un but et qui doit d  boucher sur une action utile, Rey et al., 2003), on note n  anmoins des divergences importantes quant aux caract  ristiques que ces t  ches doivent pr  senter pour pouvoir parler r  ellement de comp  tence. A l'un extr  me, on peut situer la position de Rey et al. (2003) qui proposent de prendre le mot « t  che » aussi bien « dans le sens d'une activit   partielle (par exemple, faire une multiplication) que dans le sens d'une activit   globale (par exemple, enseigner) » (p. 15) ;    l'autre extr  me, on peut situer la position de nombreux p  dagogues qui con  oivent que la comp  tence ne peut s'exprimer que dans des t  ches (tr  s) complexes et in  dites. Crahay (2006) note ainsi une « d  rive » par rapport aux d  finitions originelles (notamment l  gislatives) qui impliquaient la mobilisation de ressources face    une « t  che » dont le terme semblait pouvoir   tre entendu au sens large.

Enfin, notons encore que dans la plupart des d  finitions, les termes de complexit   et d'int  gration sont   troitement li  s puisque c'est    la n  cessit   de devoir int  grer plusieurs ressources que renvoie le terme « complexit   ».

Les degr  s de comp  tences propos  s par Rey et al. (2003), tout comme les « typologies » propos  es par Scallon (2007) ou de De Ketele (2011), permettent d'  clairer la notion de « complexit   » dans le champ scolaire. Ces trois « cadrages th  oriques », qui pr  sentent d'ailleurs un parall  lisme important et ne sont pas sans rappeler les taxonomies d'objectifs, permettent en effet de cerner une certaine hi  rarchisation de la complexit   des t  ches ou situations d'  valuation.

Scallon (2007) distingue notamment les situations d'  valuation de connaissances, d'habilit  s et de comp  tences. Selon cet auteur, les connaissances s'  valueraient dans des situations de r  p  tition permettant de r  pondre de m  moire (il s'agit de savoirs ou de connaissances d  claratives) alors que les habilit  s s'  valueraient dans des situations de g  n  ralisation demandant    l'  l  ve de choisir quelle connaissance il doit utiliser pour r  pondre    la question pos  e (il s'agit de savoir-faire ou de connaissances proc  durales). L'auteur pr  cise alors plusieurs niveaux d'habilit  s, le premier s'apparentant    des situations d'application directe et le dernier n  cessitant que l'  l  ve analyse la situation pour rechercher    quel domaine de connaissances elle fait appel et quelle proc  dure il pourra utiliser. Au niveau de ces habilit  s de haut niveau, l'  l  ve peut faire appel    des « connaissances conditionnelles » dans la mesure o   il s'agit de d  terminer « quand » et

Actes du 24^e colloque de l'Adm  e-Europe
L'  valuation des comp  tences en milieu scolaire et en milieu professionnel

« pourquoi » utiliser telle ou telle proc  dure. Scallon (2007) parle    ce propos de « strat  gies » : l'individu doit choisir, entre plusieurs proc  dures, celle qui convient le mieux    une situation donn  e. Pour l'auteur, les « r  elles comp  tences » feraient appel    « quelque chose en plus » dans la mesure o   elles n  cessiteraient non seulement la mobilisation, mais aussi l'int  gration de diverses ressources face    une situation complexe. Autrement dit, les connaissances et les habilet  s seraient les « ressources »    mobiliser et    int  grer dans une situation ou face    une t  che.

Scallon (2007) insiste encore sur le lien entre la mobilisation et l'int  gration : « mobiliser ne signifie pas (...) se rappeler ou   voquer une    une des ressources pertinentes. Rien de comparable avec une addition d'  l  ments consid  r  s isol  ment : les savoirs et les savoir-faire mobilis  s sont transform  s » (p. 110). On y revient, ce qui distingue les « vraies » comp  tences des autres habilet  s, ce sont les notions de complexit   et d'int  gration. D'aucuns estiment d'ailleurs que la complexit   renvoie au nombre de ressources que l'individu doit mobiliser et int  grer...

Ainsi, complexit   et difficult   (ou « t  che complexe » et « t  che compliqu  e » pour reprendre la terminologie de Roegiers, 2007) ne sont pas synonymes : il est en effet possible qu'une t  che complexe (*au sens d'impliquant la mobilisation et l'int  gration de diverses ressources*) pr  sente un niveau de difficult   faible alors qu'une t  che   l  mentaire (*au sens de n  cessitant l'application ou la mobilisation d'une seule ressource*) s'av  re extr  mement difficile pour les   l  ves. De notre point de vue, d  finir la complexit   comme d  pendante du nombre de ressources    mobiliser et    int  grer pose probl  me d  s que l'on cherche    l'op  rationnaliser. C'est ce que nous allons tenter de faire dans le point suivant, dans le domaine des math  matiques.

1.2. Les t  ches complexes en math  matiques

Dans le domaine des math  matiques, on s'accordera sur l'id  e que l'on propose g  n  ralement un parall  lisme entre « comp  tences » et « r  solution de probl  mes ». Si l'on consid  re qu'en math  matique, tout probl  me implique *a minima* d'analyser la situation (de se la repr  senter) pour d  finir la proc  dure qu'il convient de *mobiliser*, on se trouve devant une situation int  grant minimalement une *connaissance conditionnelle* (qui va   valuer dans le choix la proc  dure    mobiliser face au probl  me), une *connaissance proc  durale* (la proc  dure    appliquer pour r  soudre le probl  me) et une *connaissance d  clarative* (la connaissance des r  gles qui r  gissent l'utilisation de cette proc  dure). D  s que la t  che est « non routini  re », l'  l  ve devra faire appel    des *strat  gies* pour s'aider    repr  senter le probl  me (ex. : faire un sch  ma),    organiser sa d  marche de r  solution (simplifier les nombres, raisonner par analogie, d  composer en sous-probl  mes,...) et    la r  guler (v  rifier sa d  marche et sa solution, s'assurer qu'elle r  pond    la question pos  e, estimer sa vraisemblance,...). Ses « croyances » (ou *connaissances conatives*, par exemple, « je ne dois pas me fier aux mots-cl  s pour choisir l'op  ration ») interviendront   galement, m  me de mani  re inconsciente.

Finalement, tout probl  me de math  matiques implique l'int  gration de diverses « ressources ». Dit autrement, tout probl  me de math  matique est une t  che complexe qui devrait permettre l'  valuation de la comp  tence math  matique.

A ce stade, on peut se demander si l'int  gration de plusieurs ressources constitue effectivement le propre de la « t  che complexe » et qui rend possible l'  valuation de « r  elles » comp  tences ? Mais peut-  tre se trompe-t-on de d  bat et que ce qui doit constituer une t  che complexe, ce n'est pas simplement qu'elle n  cessite la mobilisation int  gr  e de plusieurs « ressources » mais, plus pr  cis  ment, la mobilisation int  gr  e de plusieurs « proc  dures » ? C'est d'ailleurs bien ce terme que l'on retrouve dans la d  finition de Rey et al. (2003) lorsqu'ils pr  cisent que l'« authentique comp  tence » correspondrait    « la capacit      r  pondre    des situations complexes et in  dites par une combinaison nouvelle de proc  dures connues » (p. 26). En d  finitive, en corrigeant l  g  rement la d  finition que nous avons donn  e dans un texte pr  c  dent (i.e. *en parlant de « proc  dure » plut  t que de « ressource »*), nous pourrions alors avancer qu'« une t  che/une situation serait consid  r  e comme complexe    partir du moment o   elle n  cessite l'identification,

Actes du 24^e colloque de l'Adm  e-Europe
L'  valuation des comp  tences en milieu scolaire et en milieu professionnel

la mobilisation et l'int  gration de plus d'une *proc  dure apprise* et qu'elle n  cessite d  s lors une interpr  tation (ou un cadrage) de la situation et une organisation de la d  marche de r  solution » (Dierendonck & Fagnant, 2010a, p. 13, d  finition actualis  e). Autrement dit, dans le domaine des math  matiques, la r  solution d'une t  che complexe supposerait la mobilisation de plusieurs « proc  dures », voire plusieurs «   tapes » d  composables en sous-probl  mes imbriqu  s (cf. phase 2 du mod  le de Rey et al., 2003) ou ind  pendants (Crahay & Detheux, 2005 ; Dierendonck & Fagnant, 2010a).

Si cette d  finition nous para  t avoir un sens dans une perspective diagnostique, doit-on n  anmoins faire de ce nouveau crit  re de complexit   une condition *sine qua non* pour distinguer s'il convient ou non de parler d'une   preuve d'  valuation de comp  tences ? Cette question reste    d  battre... et ceci d'autant plus que rien ne prouve qu'il soit nettement plus ais   d'op  rationnaliser ce qui se cache derri  re les expressions « plusieurs proc  dures » et « plusieurs   tapes de r  solution » dans la mesure o   l'interpr  tation des demandes cognitives de la t  che d  pendra toujours non seulement des   l  ves auxquels elle sera soumise mais aussi de la fa  on dont ils vont r  ellement l'aborder (de ce point de vue, m  me la distinction entre mobilisation et application peut parfois   tre sujette    d  bat).

A l'heure actuelle, les   valuations externes    large   chelle ne font pas l'  conomie du d  bat sur les comp  tences : proposent-elles des t  ches complexes aux   l  ves et permettent-elles d'  valuer des comp  tences ? Les r  ponses    cette double interrogation varient fortement d'un auteur    l'autre, comme nous allons l'illustrer ici au d  part de l'  tude internationale PISA et de divers articles publi  s dans la revue « Mesure et Evaluation en Education ».

Si l'on s'appuie sur un article du num  ro de la revue Mesure et   valuation en   ducation dans lequel les auteurs s'int  ressent notamment    l'analyse du cadre d'  valuation et des questions propos  es dans PISA, on r  pondra positivement    cette double interrogation. En effet, Leduc, Riopel et Ra  che (2011) pr  cisent que « c'est la ma  trise fonctionnelle et l'acquisition de comp  tences d'une grande port  e, combinant les connaissances, les habilit  s ainsi que les attitudes (...) qui dominent » (p. 108). Plus loin, ils ajoutent : « PISA comporte des items qui   valuent la mise en   uvre cr  ative des comp  tences math  matiques dans des situations pr  sentant toutes sortes de probl  mes, des plus quotidiens et simples jusqu'aux plus inhabituels et complexes, et qui n'ont au d  part aucune structure math  matique apparente » (pp. 111-112). Proc  dant    une analyse de quelques questions, ils montrent alors non seulement que PISA   value bien de la r  solution de probl  mes math  matiques (*t  ches complexes en elles-m  mes, comme nous avons tent   de le d  montrer au point pr  c  dent*), mais l'on peut m  me penser que certains des exemples qu'ils retiennent impliquent plusieurs proc  dures ou tout au moins plusieurs   tapes de r  solution. Au final au travers de l'analyse des items accessibles, les auteurs concluent que « PISA vise principalement      valuer les comp  tences des   l  ves » (p. 127).

Pourtant, plusieurs tenants de l'approche par comp  tences s'opposent    cette id  e. Ainsi, dans un num  ro de 2005 de la m  me revue, De Ketele et G  rard (2005) critiquent les   preuves PISA en estimant qu'elles « ne sont pas pertinentes par rapport    l'approche par comp  tences » (p. 5), qu'elles portent essentiellement sur des ressources mais « ne permettent pas (ou peu) d'  valuer la facult   de mobiliser (...) celles qui sont pertinentes pour r  soudre des probl  mes ou effectuer des t  ches complexes » (p. 5). Dans le m  me ordre d'id  es, et toujours dans la m  me revue, Carette et Dupriez (2009) pointent les « discordances » entre les diff  rentes   preuves externes d'  valuation des acquis scolaires auxquels ont acc  s les enseignants en Communaut   fran  aise de Belgique. Ils avancent le m  me type d'arguments que De Ketele et G  rard (2005) pour conclure que seules les   preuves produites par les commissions des outils d'  valuation   valuent de r  elles comp  tences en proposant « une ou plusieurs t  ches complexes et in  dites qui demandent aux   l  ves de mobiliser leurs ressources » (p. 38) ; les autres   preuves (PISA, les   valuations externes certificatives et les   valuations externes non certificatives) conduisant selon eux    proposer des   preuves o   « l'  valuation des comp  tences ne semble pas fondamentalement se d  marquer de la p  dagogie par objectifs » (p. 38).

Le lecteur aura compris que nous ne partageons pas ce point de vue (voir notamment Dierendonck & Fagnant, 2010a). Selon nous, c'est essentiellement une course   la complexit  (relativement mal d finie par ailleurs) qui conduit   ce rejet syst matique des  preuves « classiques ». En d finitive, l'essentiel n'est pas de savoir dans quelle mesure chaque t che isol e d'une  preuve externe comme PISA est ou non suffisamment complexe pour  tre consid r e comme permettant ou non d' valuer quelque chose de l'ordre de la comp tence. L'int r t de ce type d'approche r side plut t dans la possibilit  de dire qu'un  l ve donn  a atteint un certain « niveau de comp tence » correspondant   sa capacit    ma triser un ensemble de t ches mobilisant des processus cognitifs similaires ou semblables. En ce sens, les  preuves externes   large  chelle d' valuation des acquis des  l ves comme PISA permettent d'ouvrir la possibilit  d'une autre conceptualisation de la complexit , non plus focalis e directement sur la t che, mais envisageant plut t la complexit  que traduit le niveau de comp tence atteint par les  l ves.

2. Dans quelle mesure le d coupage de la t che complexe dans le mod le de Rey et al. (2003) r pond-il aux attentes en mati re d' valuation diagnostique des comp tences ?

L'int r t du mod le de Rey et al. (2003) r side dans le potentiel diagnostique du phasage qu'ils proposent, en lien avec les trois niveaux de comp tence qu'ils ont identifi s : en phase 1, les  l ves sont confront s   une t che complexe exigeant le choix et la combinaison d'un nombre significatif de proc dures ; en phase 2, on leur propose la m me t che complexe, mais cette fois d coup e en t ches  l mentaires pr sent es dans l'ordre ou elles doivent  tre accomplies ; en phase 3, on leur soumet une s rie de t ches simples d contextualis es correspondant aux proc dures  l mentaires qui ont d   tre mobilis es pour accomplir la t che complexe (Carette, 2007). Les donn es pr sent es par Rey et al. (2003) fournissent une preuve empirique qu'il existe une certaine hi rarchie entre les trois phases de leur mod le d' valuation (la phase 3 est mieux r ussie que la phase 2 qui est elle-m me mieux r ussie que la phase 1) mais rien ne permet cependant d'affirmer qu'il s'agit d'une hi rarchie stricte (Dierendonck & Fagnant, 2010a).

Nous allons questionner ce mod le en nous axant sur le potentiel diagnostique de la phase 1 d'une part et sur celui de la transition entre les phases d'autre part.

2.1. Le potentiel diagnostique de la phase 1

Les donn es obtenues par Rey et al. (2003) montrent assur ment que les t ches complexes se r v lent tr s difficiles pour les  l ves. Dans ces  preuves, la t che complexe n'est pas corrig e de fa on dichotomique, mais   l'aide d'une s rie de crit res sp cifiques qu'il est possible de sommer pour obtenir une note globale (ainsi, on peut consid rer qu'un  l ve qui obtient une note de 80%   la t che complexe a rempli correctement une proportion de 8 crit res sur 10). Pour les deux t ches propos es en math matiques dans l'enseignement primaire, les r sultats montrent que seuls 3   5% des  l ves atteignent un seuil de r ussite de 80% en phase 1 (et ils ne sont encore que 21   46%   r ussir l'une ou l'autre de ces t ches avec un seuil de 50%).

Dans le cadre d'un m moire r alis  en 2011 en Sciences de l'Education   l'Universit  de Li ge, deux t ches complexes, propos es par la commission des outils d' valuation sur la base du mod le en phases et impliquant des  quations alg briques, ont  t  soumises   huit classes d' l ves de deuxi me ann e de l'enseignement secondaire (154  l ves) durant les mois de f vrier et de mars (Bouch , 2011)¹. En phase 1, une des t ches complexes est r ussie par 11% des  l ves (Cadeaux) et l'autre par 4% d'entre eux (Muriel 6). S'interrogeant sur le pouvoir diagnostique potentiel de cette phase, Bouch  (2011) note tout d'abord que 46% des  l ves rendent une feuille blanche pour

¹Il s'agit des t ches complexes « Cadeaux » et « Muriel 6 » disponibles sur le site <http://www.enseignement.be>

Actes du 24^e colloque de l'Adm  e-Europe
L'  valuation des comp  tences en milieu scolaire et en milieu professionnel

la premi  re t  che et 10% pour la seconde. Par ailleurs, les crit  res propos  s pour analyser cette premi  re phase donnent relativement peu d'informations pour la premi  re t  che (*si 37% des   l  ves identifient correctement l'inconnue    poser – dont les 11% qui ont pu r  soudre la t  che correctement - tous les autres crit  res indiquent que seuls les   l  ves qui ont r  ussi l'ensemble de la t  che ont pu les remplir*), mais qu'ils en fournissent davantage pour la seconde t  che (*64% des   l  ves identifient correctement l'inconnue, 48% parvient    mettre en relation les divers   l  ments de l'  nonc   et 16% pr  sentent correctement les   tapes de la r  solution du probl  me*). Finalement, dans la premi  re t  che, la conjugaison du taux important d'omissions (46%) et le peu d'informations fournies par les   l  ves qui n'ont pas pu r  soudre la t  che (*26% des   l  ves remplissent 1 crit  re sur 10 et 17% n'en remplissent aucun*) incite    penser que le pouvoir diagnostique de la premi  re phase d'  valuation est clairement limit  .

Dans le cadre de l'  valuation externe men  e au Luxembourg en octobre 2009, nous avons soumis quelques t  ches complexes aux   l  ves de 3^e ann  e secondaire (filibre g  n  rale). Ces t  ches peuvent   tre consid  r  es comme « complexes » au sens retenu au terme de la premi  re question guide. A titre illustratif, le probl  me intitul   « Judo » est un exemple de t  che complexe qui n  cessite de mobiliser deux types de proc  dures (*calculer la longueur d'un c  t   d'un carr      partir de son aire et d  duire la longueur d'un segment au d  part de deux longueurs*), de « cadrer » la situation pour d  terminer les proc  dures    mobiliser et de coordonner plusieurs   tapes de r  solution (*repr  sentation du probl  me, identification de l'inconnue, calcul de la longueur d'un c  t   de la salle, calcul de la longueur d'un c  t   du tatami et d  duction de la largeur du chemin au d  part de ces deux longueurs*). Cette t  che complexe, soumise    1769   l  ves, n'est r  ussie que par 14% des   l  ves (Dierendonck & Fagnant, 2010b). Bien que seule la r  ponse finale ait   t   prise en compte, l'analyse des r  ponses permet d'inf  rer les d  marches erron  es les plus courantes des   l  ves. A titre illustratif, 25% des   l  ves obtiennent une r  ponse t  moignant d'une erreur se situant    la toute derni  re   tape de r  solution (*la d  duction de la largeur du chemin n  cessite de soustraire les deux longueurs trouv  es pr  c  demment, puis de diviser le r  sultat par deux*), 11% des   l  ves obtiennent une r  ponse qui t  moigne d'une d  marche assez superficielle (*la soustraction des deux aires fournies conduit    trouver l'aire du chemin et non la largeur de celui-ci*) et seuls 6% des   l  ves ne proposent aucune r  ponse. Au final, cette t  che de complexit   mod  r  e, m  me corrig  e de fa  on binaire, semble offrir un certain pouvoir diagnostique d  s l'  tape 1.

En conclusion de ce premier point, nous retiendrons deux   l  ments : (a) les t  ches trop complexes perdent leur pouvoir diagnostique si elles d  stabilisent les   l  ves au point qu'ils rendent une feuille blanche ou ne proposent que des bribes de r  ponses; (b) des t  ches sollicitant une r  ponse num  rique pr  cise peuvent    la fois   tre des t  ches complexes permettant d'  tre corrig  es de fa  on binaire et offrir un certain pouvoir diagnostique d  s la phase 1.

2.2. Le potentiel diagnostique de la transition entre les phases

A priori, le passage entre les phases 1 et 2 du mod  le de Rey et al. (2003) pr  sente un potentiel diagnostique    un double niveau : (a) un   l  ve qui r  ussit    r  soudre un probl  me en phase 2 alors qu'il n'y parvenait pas en phase 1 ma  trise les ressources impliqu  es dans la t  che complexe, semble capable de les mobiliser dans des t  ches simples (d  compos  es) mais   prouve des difficult  s    organiser sa d  marche de r  solution et/ou    int  grer les diff  rentes ressources ; (b) pour les   l  ves qui ne parviennent pas    r  soudre le probl  me en phase 2, une analyse plus pr  cise des   tapes de cette phase 2 devrait permettre de cerner o   se situent les difficult  s. Tout en reconnaissant l'int  r  t potentiel de cette   tape de d  composition, nous pointerons ici quelques questionnements qu'elle soul  ve    nos yeux.

Tout d'abord, on peut craindre qu'un   l  ve qui   choue lors de la phase 1   choue aussi lors de la phase 2 parce qu'il est d  concert   par la situation propos  e ou, tout simplement, parce qu'il se sent d  courag  . A titre illustratif, dans l'  tude de Bouch   (2011), le passage de la phase 1    la phase 2 du probl  me « cadeaux » ne permet qu'une tr  s faible augmentation du taux de r  ussite (on passe

Actes du 24^e colloque de l'Admée-Europe
L'évaluation des compétences en milieu scolaire et en milieu professionnel

de 11 à 18%), ce qui pourrait laisser entendre que les élèves ont « décroché ». L'analyse plus fine des étapes de cette phase 2 permet toutefois de cerner un peu mieux les difficultés des élèves que lors de la phase 1 où, pour rappel, on notait 46% d'omissions.

Par ailleurs, on peut imaginer qu'un élève perturbé par le contexte de la tâche (ou découragé par la complexité de la phase 1) échoue lors de la phase de décomposition alors qu'il aurait peut-être pu répondre correctement à d'autres tâches évaluant des compétences de niveau 2 dans un autre contexte. A titre illustratif, dans l'épreuve externe soumise aux élèves de 3^e année du secondaire au Luxembourg, nous avons testé simultanément la décomposition de la tâche complexe et l'évaluation des mêmes procédures dans des problèmes « simples » (au sens d'impliquant la mobilisation d'une seule procédure). Les résultats montrent assez nettement que les tâches évaluant une ressource isolée dans un contexte différent sont mieux réussies que la tâche complexe décomposée qui nécessite la réussite partielle de toutes les étapes imbriquées. A nos yeux, la combinaison d'une décomposition de la tâche et d'une évaluation des ressources dans des contextes différents (Crahay & Detheux, 2005) semble pouvoir apporter des informations complémentaires intéressantes au niveau diagnostique (Dierendonck & Fagnant, 2010a, b).

D'autres questionnements portent sur l'impact potentiel des choix effectués en termes de structuration de la phase 2 ; nous en esquisserons trois.

Premièrement, on peut se demander si les étapes de la résolution peuvent s'organiser de différentes façons et, dans ce cas, si l'organisation proposée correspond ou non à la démarche intuitive des élèves. Autrement dit, ne risque-t-on pas de les déstabiliser en voulant les faire entrer dans une démarche qui n'était pas celle qu'ils avaient tenté de développer en phase 1 ? En résolution de problèmes mathématiques, il y a souvent plusieurs voies possibles pour résoudre un problème et force est de constater que la logique de l'adulte n'est pas toujours la même que celle de l'enfant ou de l'adolescent. Dans la mesure où, en terme méthodologique, on plaide généralement pour partir des démarches informelles des élèves, ce découpage « contraint » pose question : non seulement, il peut « perturber » la démarche spontanée de l'élève, mais il pourrait aussi laisser croire aux enseignants que cette façon de procéder (*analyser la tâche « à la place des élèves » - i.e. lors d'une phase collective par exemple - et organiser « leur » démarche de résolution*) pourrait se traduire efficacement en stratégie d'enseignement.

Deuxièmement, certaines tâches mathématiques peuvent donner lieu à des démarches de résolution différentes (au sens où les étapes peuvent s'organiser selon des ordres variables), mais elles peuvent parfois aussi être résolues en faisant appel à des procédures mathématiques différentes (*par exemple, certains problèmes peuvent être résolus en développant une démarche arithmétique – s'appuyant sur des opérations avec des nombres - ou une démarche algébrique – passant par la désignation d'une inconnue et la mise en équation du problème*). L'étude présentée par Demonty et Dupond dans le cadre d'un symposium du congrès 2012 de l'Admée-Europe permet d'illustrer très clairement ce type de comportements et les questionnements que cela soulève d'un point de vue diagnostique.

La troisième interrogation concerne le format même des questions proposées en phase 2. Par exemple, lorsqu'il s'agit de poser des équations dans des tâches complexes telles que celles testées par Bouché (2011), la présence de QCM en phase 2 dans l'une des tâches semble permettre à une plus grande proportion d'élèves d'entrer dans le problème que les questions ouvertes proposées dans l'autre tâche (*la proportion d'omissions est notamment nettement moindre*). L'impact des formats de questions à cette étape de décomposition (où l'on donne des indicateurs de démarches) joue sans doute une influence que l'on mesure assez difficilement et qui se distingue probablement du format de question proposé pour indiquer la réponse finale d'un problème².

² La réponse finale ne donne pas en elle-même d'indication sur la démarche à effectuer – même si les distracteurs peuvent être conçus pour mettre en lumière des démarches erronées – alors qu'une série d'équations posées en QCM lors de la phase de décomposition donnent des indications sur la démarche et « cadrent » la situation en indiquant quels sont les savoirs mathématiques à mettre en œuvre.

Enfin, pour terminer ce point tout en l'  tendant vers la phase 3 de l'  preuve, soulignons encore deux « lacunes diagnostiques potentielles » des   preuves en phases : (a) Quelle est la nature exacte des difficult  s des   l  ves qui parviennent    r  soudre le probl  me en phase 2 et non en phase 1 (pourquoi savent-ils appliquer, mobiliser, mais non « mobiliser et int  grer ») ? (b) Quelle est la nature exacte des difficult  s des   l  ves qui parviennent    r  soudre les items en phase 3 mais qui ne parviennent pas    r  soudre le probl  me en phase 2 (pourquoi savent-ils appliquer, mais non « mobiliser ») ? Autrement dit, si les informations issues de la phase 3 permettent de cibler certaines lacunes « techniques », relatives    la non-ma  trise de proc  dures ou de savoir-faire de base, elles ne permettent g  n  ralement pas    elles seules d'expliquer pourquoi autant d'  l  ves continuent      chouer en phase 2.

Qu'en est-il en outre de ces myst  rieux processus que sont le « savoir mobiliser » et le « savoir mobiliser et int  grer » qui sont point  s au c  ur de toutes les d  finitions de la notion de comp  tence ? Pour   clairer ce questionnement, il para  t essentiel de se d  tacher d'une analyse « d  contextualis  e » des r  sultats que l'on peut tirer d'une   valuation papier-crayon pour chercher    mieux cerner les processus cognitifs r  ellement    l'  uvre dans les t  ches complexes. Pour ce faire, il semblerait int  ressant de combiner des « laboratoires cognitifs » o   l'on observe un   l  ve face    une t  che complexe (tout en l'interrogeant sur ses d  marches), des observations centr  es sur l'analyse des interactions entre   l  ves lorsqu'ils r  solvent une t  che complexe en groupe et, enfin, des observations centr  es sur le r  le d'  tayage qu'un enseignant met en   uvre (ou pourrait mettre en   uvre) pour interagir efficacement avec les   l  ves.

3. Comment assurer la fiabilit   et la validit   des informations recueillies au cours d'une   valuation de comp  tences ?

En ce qui concerne les qualit  s psychom  triques des   preuves, il est courant de distinguer une validit   de « *construct* » (ou validit   th  orique) et une validit   de « contenu » (ou repr  sentativit  ) : l'  preuve mesure-t-elle bien ce qu'elle pr  tend mesurer et couvre-t-elle les aspects importants des apprentissages ? En termes de « validit   de contenu », on peut s'interroger sur les « limites » d'une   valuation de comp  tences qui serait compos  e uniquement de quelques t  ches complexes : d'une part, on est loin d'atteindre une repr  sentativit   des contenus enseign  s et, d'autre part, on peut s'interroger sur le respect d'une certaine concordance entre ce qui est enseign   en classe et ce qui est   valu   (ce qui serait une autre fa  on de rendre une   valuation « non valide » ; Beckers, 2005).

Concernant les liens entre enseignement, apprentissage et   valuation, les r  sultats particuli  rement faibles obtenus dans la plupart des   tudes   valuant des t  ches complexes (Bouch  , 2011, Crahay & Detheux, 2005 ; Fagnant & Dierendonck, 2010b ; Marcoux,    para  tre ; Rey et al., 2003) incitent    un double questionnement : d'une part, les enseignants proposent-ils des t  ches complexes    leurs   l  ves en situation d'apprentissage et d'autre part, s'ils proposent de telles t  ches, comment proc  dent-ils d'un point de vue m  thodologique ? S'interrogeant sur les faibles taux de r  ussite obtenus face aux t  ches complexes de math  matiques soumises    des   l  ves de 5^e ann  e primaire    Gen  ve, Marcoux (   para  tre) pointe l'  cart entre les intentions m  thodologiques des programmes et les situations d'  valuation : « Dans la lign  e d'une forme de socioconstructivisme, l'apprentissage des math  matiques    Gen  ve se con  oit comme une construction en interaction avec les autres (...) L'  l  ve genevois se voit principalement confront      la d  couverte d'apprentissages et    la r  solution de probl  mes quasi exclusivement en groupe. Il se peut que, dans ce cas, cette fa  on de faire soit plus productive qu'une transmission frontale du savoir, mais elle peut risquer aussi, par son caract  re de recours aux ressources externes (ici l'enseignant ou les pairs), de devenir un obstacle dans la r  solution de t  ches individuelles » (p. 15).

D  s les premi  res lignes de ce texte, nous avons   voqu   la question de la mobilisation conjointe de ressources internes et de ressources externes. Ces ressources externes, que nous avons dans un premier temps «   vacu  es » pour d  battre de la complexit   des t  ches, jouent pourtant un r  le

Actes du 24^e colloque de l'Admée-Europe
L'évaluation des compétences en milieu scolaire et en milieu professionnel

central chez tout expert dans un domaine. En se référant à des travaux déjà anciens, on rappellera que Perkins (1995, cité par Crahay, 1999) définit l'expert comme un « individu-plus » (par opposition à un individu-solo) qui « recourt continuellement à des outils divers » et qui « sollicite régulièrement l'une ou l'autre personne de son entourage pour obtenir une information ou une appréciation de son travail » (p. 340). Comment les évaluations de compétences (et, *a fortiori*, des évaluations externes à large échelle) pourraient-elle tenir compte de ce paramètre ? Par ailleurs, comment pourraient-elles être réellement valides sans permettre le recours à des ressources externes ? Le développement du testing informatisé ouvre sans doute une voie prometteuse en ce sens (*les technologies pourraient à la fois permettre le recours à des ressources matérielles et une communication entre plusieurs sujets qui résolvent une tâche conjointement*), mais il occasionne également de nouveaux questionnements (*notamment, comment tenir compte de l'apport des pairs pour déterminer la compétence d'un sujet donné*).

Le deuxième point (la question de la représentativité des contenus enseignés) semble pouvoir se lier à la question de la fiabilité de la mesure. La crainte principale face à des évaluations qui ne seraient constituées que de quelques situations complexes (telles que le proposent notamment De Ketele & Gérard, 2005) serait de passer à côté des compétences des sujets. En effet, un élève confronté à trois tâches complexes pourrait très bien rendre une feuille blanche pour chacune d'elle (*on a vu que le taux d'omission pouvait être très élevé face à ce type de tâches*), ce qui conduirait à conclure qu'il ne sait rien (qu'il n'a aucune compétence) alors qu'il dispose probablement *a minima* de certaines habiletés... Non seulement cela pose question d'un point de vue déontologique, mais cela n'aurait aucune portée diagnostique. En ce sens, le phasage proposé par Rey et al. (2003) semble nettement plus informatif puisqu'il envisage d'autres « dimensions » grâce aux étapes 2 et 3. Toutefois, si les tâches proposées sont trop complexes, on risque de ne pouvoir en proposer que deux ou trois (avec leur « phasage ») dans le cadre d'une évaluation d'une durée « raisonnable ». Du point de vue de la mesure, cela pose question : non seulement les différents critères qui pourraient présider à l'évaluation de la tâche complexe sont généralement liés (*on a vu qu'en phase 1, certaines situations conduisaient pratiquement à une dichotomie : remplir tous les critères ou n'en remplir aucun, Bouché, 2011*), mais nous avons également formulé l'hypothèse que les deux premières étapes du phasage risquaient de l'être également (*le contexte identique peut continuer à « perturber » les élèves ou conduire à les décourager*). Les modèles de réponse à l'item postulent l'indépendance des items : une fois le trait latent mesuré, aucune relation ne doit exister entre les réponses d'un sujet aux différents items. Cette condition d'application risque d'entrer en conflit direct avec les modèles d'évaluation qui s'appuient sur un découpage en phases.

Comme nous l'avons défendu dans un texte précédent (Dierendonck & Fagnant, 2010a), il nous semble que la conciliation entre les nouveaux outils d'évaluation des compétences et les épreuves classiques d'évaluation pourrait trouver un terrain d'entente moyennant deux conditions essentielles : s'accorder sur une définition moins ambitieuse des tâches complexes et s'assurer que les épreuves comportent suffisamment d'items (la première condition rendant la seconde possible).

Selon nous, les évaluations externes devraient donc contenir, en plus de quelques tâches complexes évaluant des compétences, un nombre suffisant de tâches élémentaires permettant d'évaluer la mobilisation de certaines ressources/procédures ainsi que leur application directe. Si certaines de ces tâches élémentaires peuvent être liées à la tâche complexe (selon le phasage proposé par Rey et al., 2003), il nous semble important que d'autres en soient totalement indépendantes, et ce pour deux raisons principales: d'une part, cela permet d'évaluer la mobilisation et l'application d'une plus grande diversité de ressources/procédures que celles directement impliquées dans la tâche complexe initiale et, d'autre part, on donne ainsi plus de chances aux élèves de démontrer leur maîtrise des compétences de deuxième degré (ou leurs habiletés) en leur proposant des contextes diversifiés. Si l'on s'accorde sur cette proposition (voir texte de Dierendonck & Fagnant dans le cadre du symposium 6 du congrès 2012 de l'Admée-Europe), les techniques classiques de validation des épreuves sembleraient pouvoir continuer à être utilisées...

Actes du 24^e colloque de l'Admée-Europe
L'évaluation des compétences en milieu scolaire et en milieu professionnel

Références

- Beckers, J. (2005). Est-il possible de faire de la pédagogie par compétences un allié de l'équité à l'école ? *Les cahiers du Service de Pédagogie expérimentale*, 21&22, 41-63.
- Bouché, B. (2011). *Comment différencier et développer des compétences en mathématiques au début de l'enseignement secondaire à partir des difficultés des élèves en utilisant les outils d'évaluation ?* Mémoire présenté en vue de l'obtention du master en Sciences de l'Education. Université de Liège, document non publié.
- Carette, V. (2007). L'évaluation au service de la gestion des paradoxes liés à la notion de compétence. *Mesure et Evaluation en Education*, 30(2), 49-71.
- Carette V., Dupriez V. (2009). La lente émergence d'une politique scolaire en matière d'évaluation des élèves. Quinze ans de transformations en Belgique francophone. *Mesure et Evaluation en Education*, 32 (3), p. 23-45
- Crahay, M. (1999). *Psychologie éducationnelle*. Paris : PUF.
- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de compétence. *Revue française de pédagogie*, 154, 97-110.
- Crahay, M. & Detheux, M. (2005). L'évaluation des compétences, une entreprise impossible ? (Résolution de problèmes complexes et maîtrise de procédures mathématiques). *Mesure et Evaluation en Education*, 28(1), 57-78.
- Demonty, I. & Dupont, V. *La mesure des compétences des élèves face à la résolution d'une tâche complexe en algèbre élémentaire : l'apport d'indicateurs issus des recherches centrées sur les difficultés spécifiques des élèves en algèbre*. Communication dans la cadre d'un symposium de l'Admée 2012, Luxembourg.
- De Ketele, J.-M. (2011). *Scores multidimensionnels : une nécessité pour évaluer les compétences*. Communication orale lors du deuxième congrès français de psychométrie. Sèvres, 16 mars 2011.
- De Ketele, J.-M. & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences, *Mesure et Evaluation en Education*, 28(3), 1-26.
- Dierendonck, C. & Fagnant, A. (2010a). Quelques réflexions autour des épreuves d'évaluation développées dans le cadre de l'approche par compétences. *Le Bulletin de l'ADMEE-EUROPE*, 2010/1, 5-20.
- Dierendonck, C. & Fagnant, A. (2010b). *Monitoring du système scolaire et évaluation des compétences en mathématiques: une étude exploratoire en vue de donner un feedback diagnostique aux enseignants*. Communication orale au 22^e colloque international de l'Admée-Europe. Evaluation des curriculums et des programmes d'éducation et de formation. Du 14 au 16 janvier 2010, Braga (Portugal).
- Dierendonck, C. & Fagnant, A. Evaluer des compétences en mathématiques dans le cadre d'une épreuve externe à large échelle au départ de tâches complexes, de tâches décomposées et de tâches élémentaires : quel pouvoir diagnostique ? Communication dans la cadre d'un symposium de l'Admée 2012, Luxembourg.
- Leduc, Riopel et Raïche (2011). L'influence des définitions des habiletés disciplinaires sur la création et le choix d'items dans le PISA et le TEIMS. *Mesure et Evaluation en Education*, 34(1), 97-130.
- Marcoux, G. (à paraître). *Quand la résolution de problèmes peut nous informer sur les inégalités : Quelle trame des échecs et des réussites ?*
- Rey, B., Carette, V., Defrance, A. & Kahn, S. (2003). *Les compétences à l'école — Apprentissage et évaluation*. Bruxelles : De Boeck.
- Roegiers, X. (2007). *Des situations pour intégrer les acquis scolaires*. Bruxelles: De Boeck.
- Scallon, G. (2007) (2^e édition). *L'évaluation des apprentissages dans une approche par les compétences*. Bruxelles : De Boeck Université.
- Tardif, J. (2004). Un passage obligé dans la planification de l'évaluation des compétences : la détermination des indicateurs progressifs et terminaux de développement *Pédagogie collégiale*, (1^{re} partie), 18(1), 21-26 ; (2^e partie), 18(2), 13-20.

TABLE RONDE
COMMENT ASSURER L'ÉVALUATION DIAGNOSTIQUE DES COMPÉTENCES SCOLAIRES?

Nathalie Loye

*Université de Montréal, nathalie.loye@umontreal.ca

Mots-clés : évaluation diagnostique, compétences

Résumé. *Le texte qui suit se propose d'apporter quelques éléments de réflexions relatifs à la problématique de l'évaluation diagnostique des compétences scolaires. Il est structuré autour des trois questions guides proposées par Pascal Ndinga, organisateur de la table ronde : « Qu'est-ce qu'une tâche ou une situation complexe ? » ; « Dans quelle mesure le découpage de la tâche complexe dans le modèle de Rey et al. (2003) répond-il aux attentes en matière d'évaluation diagnostique des compétences ? » et « Comment assurer la fiabilité et la validité des informations recueillies au cours d'une évaluation de compétences ? ».*

1. Introduction

Le titre de cette table ronde évoque à la fois le contexte dans lequel l'évaluation se fait dans plusieurs pays et la visée particulière de l'évaluation qu'est le diagnostic. Il est donc utile avant de répondre aux trois questions spécifiques qui structurent cette table ronde de prendre un peu de recul relativement à ces deux éléments.

Le premier concerne le contexte dans lequel plusieurs pays ont choisi d'intégrer leur système scolaire. L'approche par compétences est en effet entrée dans nos écoles depuis plusieurs années, sans toutefois obligatoirement être complètement entrée dans nos mœurs, peut-être à cause des divergences de compréhension de ce qu'elle implique pour l'évaluation. En effet, les compétences n'ont probablement pas la même saveur et le même sens pour chacun d'entre nous. L'un des éléments crucial est par exemple de définir le type de tâche qui permet de mettre en œuvre la compétence et par suite de l'évaluer. C'est l'objet de la première question de cette table ronde et un regard sur le contenu de la recherche empirique peut aider à y répondre.

Le deuxième élément est introduit par le mot *diagnostique*. Cet aspect semble quant à lui prendre très peu de place dans les questions de cette table ronde. En effet, il n'apparaît que dans la question 2 en posant l'hypothèse que le modèle de Rey et ses collègues (Rey, Carette, Defrance, & Kahn, 2003) est un modèle diagnostique. Ainsi, répondre à la question 2 implique nécessairement de définir l'évaluation diagnostique et de réfléchir à la portée diagnostique du modèle de Rey et al.

Quant à la dernière question, elle vise à examiner le processus d'assurance qualité de notre évaluation des compétences. Y répondre permettra de discuter des tensions entre nos conceptions classiques de la validité et de la fidélité et nos conceptions nouvelles des instruments nécessaires à l'évaluation des compétences, notamment dans une visée diagnostique.

2. Question 1 : Qu'est-ce qu'une tâche ou une situation complexe?

Nous sommes habituellement tous d'accord pour dire que la notion de compétence provient de la formation professionnelle, dans laquelle elle prend tout naturellement sa place. Par exemple, aucun d'entre nous ne

voudrait monter dans un avion dont le pilote ne serait pas compétent, ou se faire soigner par un médecin incompetent. Cela signifie entre autre qu'avant d'être diplômés, et ainsi jugés compétents, pilotes et médecins doivent avoir été évalués dans un contexte non seulement authentique, mais également réel. On peut ici penser au *test en vol du pilote* ou à l'externat et l'internat des étudiants en médecine.

Le milieu médical offre d'ailleurs un cadre théorique simple et utile pour comprendre les différents niveaux de compétence professionnelle, et par la suite pouvoir les évaluer. Miller (1990) propose en effet de décrire la compétence professionnelle dans une pyramide (voir figure 1) à quatre niveaux dont le niveau supérieur ne peut être évalué sans le recours à une mise en situation concrète et réelle (*fait, does*) et les trois autres niveaux à des situations que nous pourrions qualifier de plus scolaires (*montre comment, shows how; sait comment, knows how ; sait, knows*).

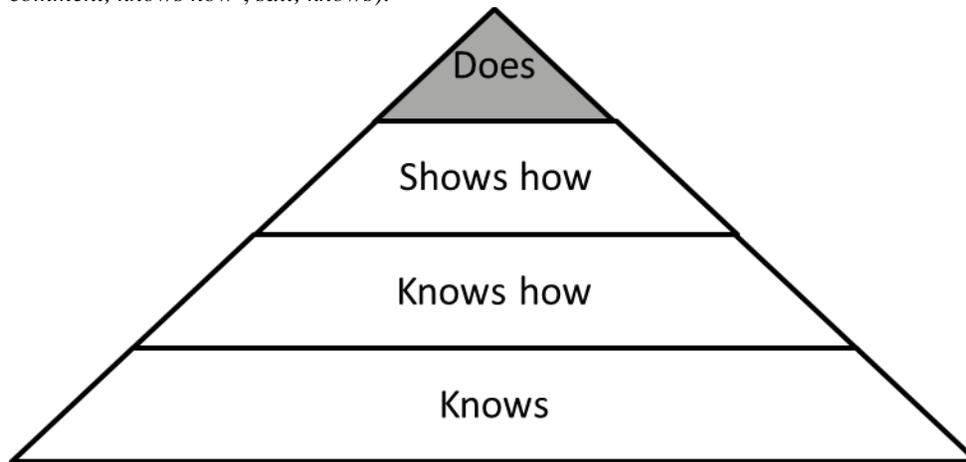


Figure 1 : La pyramide de Miller (1990)

À partir de la pyramide de Miller, nous pouvons d'abord apporter deux premiers éléments de réponse à la question. *Le premier est qu'on ne peut pas monter dans la pyramide sans passer par chacun des niveaux.* Il convient donc de proposer des tâches évaluatives qui correspondent à chacun d'eux. Le deuxième élément est une réflexion sur la possibilité de mettre les élèves dans des situations permettant d'évaluer le *Faire*. L'authenticité a un rôle essentiel à ce niveau. Cette authenticité doit absolument être ancrée dans la réalité lorsque les compétences à évaluer sont professionnelles. Elles nécessitent alors de recourir à des tâches réelles (ou au moins simulée de manière réaliste), et pas seulement ressemblant à la réalité (par exemple le *test en vol* du pilote). Pour l'élève à l'école, les tâches du *Faire* doivent le plus souvent prendre place en salle de classe ou d'examen. Dans certains cas, la technologie peut aider à améliorer le caractère authentique de la tâche. Toutefois, une grande quantité de tâches doivent être mises en œuvre et faire référence à divers contenus pour permettre de juger du niveau d'un élève. Ce dernier point semble contradictoire avec certaines tâches authentiques qui prennent place en salle de classe sur de très longues périodes compte tenu de leur complexité, limitant ainsi la quantité de situations proposées aux élèves. On peut également se questionner sur la part que doivent prendre les tâches du *Faire* dans le cas d'une formation scolaire et non professionnelle.

En ce qui concerne les trois niveaux inférieurs de la pyramide, et qui correspondent aux situations évaluatives que l'on souhaite le plus souvent mettre en place avec des élèves, il convient de définir les caractéristiques des tâches nécessaires à leur évaluation. Là encore, la littérature en médecine guide notre réflexion. Van der Vleuten et ses collègues (van der Vleuten, Schuwirth, Scheele, Driessen, & Currie, 2010) ont proposé une synthèse de recherches empiriques et posé les bases d'une théorie sur l'évaluation des compétences professionnelles. Leur revue de littérature les amène à plusieurs constats et nous en tirons quatre principes qui nous permettent de répondre à la question.

1) Le premier constat est que la mise en œuvre d'une compétence est fortement reliée aux contenus sous-jacents à la tâche. Ainsi, le fait de réussir une tâche prédit mal ou peu la réussite à une autre tâche. Autrement dit, une seule tâche ne mesure rien et n'est pas généralisable. On ne peut donc rien en déduire

sur le niveau de maîtrise du sujet. *Le premier principe est donc qu'il faut combiner des prises d'information à des moments différents, en variant les contenus.*

2) Le deuxième constat est relié au format de la tâche. Une épreuve évaluative quelle qu'elle soit est toujours formée d'un stimulus et d'un format de réponse. Ce format de réponse peut être oral ou écrit et se faire sur une page blanche dans laquelle produire une réponse longue (par exemple : phase 1 du modèle de Rey et al.), demander d'écrire une réponse courte, de choisir une réponse parmi une liste ou toute autre possibilité. Les études empiriques montrent que la validité, autrement dit la capacité à évaluer ce qu'on veut évaluer, dépend beaucoup plus du stimulus que du format de réponse. Ce stimulus doit être aussi authentique que possible, sans toutefois que la complexité de la description du contexte ne crée une surcharge cognitive pour l'élève et l'empêche de répondre correctement. *Le deuxième principe est donc qu'il faut mettre beaucoup d'emphase sur ce qui est demandé à l'élève et moins se préoccuper de la manière dont il doit fournir la réponse.*

3) Le troisième constat est que la fidélité des résultats dépend plus de la variété des tâches et des évaluateurs que de l'objectivité et de la standardisation des méthodes pour porter le jugement. Autrement dit, un jugement global porté à de multiples occasions et par plusieurs évaluateurs (par exemple : l'enseignant, l'élève, ses pairs) sera plus stable et permettra de mieux discriminer les différents niveaux de compétence qu'une évaluation plus analytique et plus objective qui risque de varier beaucoup d'une tâche à une autre. Ceci dit, il est évident que cela dépend de ce qui est demandé à l'élève. Certains types de questions ayant une seule bonne réponse sont naturellement reliés à une correction objective, l'objectivité n'est toutefois pas une condition nécessaire pour assurer la fidélité des jugements. De plus il n'y a pas vraiment de raison d'additionner les uns aux autres des scores obtenus à des questions portant sur des contenus différents, même si le format de ces questions est semblable. *Le troisième principe est donc que porter un jugement global est probablement plus approprié que d'imposer systématiquement une méthode analytique ou comptable.*

4) Enfin, le dernier constat est qu'une seule méthode ne pourra jamais suffire pour passer à travers les niveaux de la pyramide de Miller. *Le quatrième principe est donc qu'il faut se préoccuper de proposer des tâches multiples et variées aux élèves pour évaluer leurs compétences, mais également pour les aider à les développer, car c'est aussi le rôle de l'évaluation.*

Évaluer la compétence doit se faire à chacun des niveaux de la pyramide de Miller et donc avec *de multiples types de tâches*, en visant *une diversité de contenus* et à *de nombreuses occasions*. Pour chaque tâche, la *qualité du stimulus* proposé à l'élève est plus importante que le format de réponse et il n'est pas obligatoire d'imposer une approche analytique ou comptable pour assurer la fidélité du jugement tâche après tâche.

3. Question 2 : Dans quelle mesure le découpage de la tâche complexe dans le modèle d'évaluation des compétences en trois phases de Rey et al. (2003) répond-il aux attentes en matière d'évaluation diagnostique des compétences?

Le modèle d'évaluation de Rey et de ses collègues (Rey, et al., 2003) se caractérise par une même tâche déclinée en trois phases correspondant chacune à un niveau de compétence différent, allant du niveau le plus élevé (degré 1) au niveau le plus simple (degré 3). Selon les auteurs, la capacité diagnostique de ce modèle provient du fait « que les difficultés rencontrées éventuellement par les élèves dans une phase s'éclaircissent par leur performance à la phase suivante » (p. 47).

La question posée ici est double. D'abord, il convient de définir si ce modèle s'accommode de la réponse qui a été fournie à première question et qui visait à définir les caractéristiques de tâches pour évaluer des compétences. Il convient ensuite de définir si ce modèle répond aux exigences de la visée diagnostique de l'évaluation.

Le fait de réussir une tâche prédit très peu la réussite à une autre tâche. Nous en avons tiré le *principe 1* selon lequel il faut multiplier les contenus visés par les tâches pour évaluer efficacement les compétences d'un élève. Le fait de proposer la même tâche sous trois formats différents est lourd en termes de temps nécessaire à la mise en place du modèle, ce qui limite donc la possibilité de multiplier les tâches et les contenus, et par suite la qualité des inférences. De plus, le *principe 1* remet en cause la portée diagnostique du modèle puisqu'il est impossible de tirer des conclusions à partir d'une seule tâche. Au mieux, c'est la répétition des mêmes difficultés tâche après tâche qui permettrait de faire un diagnostic. Dans ce cas, le *principe 3* met en évidence le fait que porter un jugement global sur les diverses tâches sera efficace pour inférer le niveau de compétences de l'élève. Le modèle ne semble pas en contradiction avec ce principe et permet de dresser un portrait général de l'élève pour chaque tâche. Toutefois, le risque est que selon les contenus, ce portrait général varie beaucoup limitant ainsi les conclusions à en tirer et en particulier la possibilité d'établir un diagnostic.

Le *principe 2* établit que le stimulus est plus important que le format de réponse. Étant donné ce principe, on peut se questionner sur l'utilité de proposer la même tâche sous trois formats différents. Enfin, le *principe 4* pointe vers la nécessité de varier les types de tâches pour évaluer les compétences avec efficacité. Le modèle de Rey et de ses collègues peut donc être utilisé, mais ne devrait pas être généralisé. Ce qui précède met déjà en évidence le fait que le modèle de Rey et de ses collègues ne répond pas aux caractéristiques d'une évaluation diagnostique, puisqu'on ne peut pas tirer de conclusions à partir d'une seule tâche. Au-delà de ce premier point, il convient de définir les signes distinctifs d'une évaluation diagnostique afin de juger plus avant du potentiel du modèle à être utilisé dans cette visée.

Tout d'abord, le mot diagnostique vient des mots grecs *gnosis*, qui signifie connaissance, et *dia*, qui signifie à travers. De plus, dans le langage courant, un diagnostic est le point de départ à des actions concrètes. Par exemple, le médecin pose son diagnostic, puis propose un traitement spécifique pour vous soigner. Le mécanicien examine votre voiture et fait un diagnostic avant de définir quelles réparations vont régler le problème. En éducation, une évaluation diagnostique doit mettre en évidence des informations précises reliées à l'apprentissage des élèves. Ces informations sont souvent définies comme étant des forces et des faiblesses (Legendre, 2005). Le diagnostic doit fournir des indications suffisamment précises pour aboutir à des solutions individuelles et adaptées pour remédier aux difficultés identifiées.

Notre réflexion nous a amenée à considérer qu'il existe trois angles différents pour regarder l'évaluation : considérer (1) la position de l'habileté du sujet sur un continuum, (2) les erreurs commises et (3) la structure de l'épreuve évaluative. Ces trois angles permettent de classer les approches évaluatives existantes, d'en définir les avantages et inconvénients dans une perspective diagnostique et de chercher à situer le modèle de Rey et de ses collègues.

1) La position de l'habileté du sujet sur un continuum

La théorie classique des scores (TCS), la théorie de la généralisabilité (GEN) ou la théorie de réponse à l'item (TRI) ont été élaborées pour déterminer les items d'un test qui sont informatifs et qui discriminent bien les sujets. Les tests développés dans ces perspectives permettent de placer les sujets sur une échelle de mesure de leur habileté. Ces approches permettent donc de très bien évaluer l'habileté globale d'un élève, mais pas ses habiletés relatives aux différents éléments de son raisonnement. Elle nécessite en outre une variété d'items indépendants les uns des autres et est basée sur le concept d'unidimensionnalité qui décrit difficilement la complexité de la cognition humaine (Gitomer & Rock, 1993). Le modèle de Rey et ses collègues vise un contenu spécifique et propose quelques questions reliées les unes aux autres à travers trois phases. Il n'entre donc clairement pas dans cette catégorie, qui de toute manière est assez éloignée de la visée diagnostique de l'évaluation.

2) Les erreurs commises pour caractériser le sujet

Pour Scallon (1988), la détection systématique des erreurs par le correcteur offre la possibilité de déterminer les forces et faiblesses d'un élève. Toutefois, l'usage de stratégies différentes pour répondre à un item rend difficile le codage des erreurs. Le bénéfice est la mise en évidence précise des difficultés de chaque élève au prix toutefois d'un travail fastidieux de correction. Dans le cas de questions nécessitant la formulation d'une réponse par l'élève, il faut que les informations contenues dans la copie soient en quantité suffisante. Dans le cas d'un test à choix multiple (QCM), une possibilité consiste à coder les

leurres en fonction d'erreurs précises. Par contre, comme le format impose une bonne réponse et trois ou quatre leurres, il est souvent irréaliste de trouver des leurres similaires correspondant à des erreurs systématiques. Le modèle de Rey et de ses collègues pourrait éventuellement se prêter à l'exercice du codage des erreurs dans les trois phases de réalisation. Toutefois, dans sa forme actuelle, l'évaluation ne procède pas de cette approche, puisque c'est un regard global sur les trois phases qui permet de poser le diagnostic en définissant si la difficulté provient de la démarche, du choix ou de la mise en œuvre de la procédure. Étant donné que c'est la répétition du même type d'erreur commise, c'est-à-dire le côté systématique, qui est à la base du diagnostic, le modèle de Rey et de ses collègues n'entre donc pas non plus dans cette catégorie.

3) La structure du test pour caractériser le sujet

Dans une approche par objectifs, la structure est traditionnellement fournie par le tableau de spécification du test qui propose un découpage établi par des experts. Le diagnostic est alors constitué des scores partiels totalisés sur un ensemble d'items congruents à un même objectif (Scallon, 2000). Toutefois, « attitudes, processus cognitifs, stratégies métacognitives, stratégies de résolution de problèmes, pensée critique, capacité d'intégration [, de même que notion de] transfert [et de] compétence semblent [...] échapper à la façon traditionnelle d'explicitier les objectifs » (Scallon, 2000, p 153). La structure selon un tableau de spécification est donc difficile à envisager dans une approche par compétence.

Il faut donc penser différemment la structure d'un test diagnostique pour évaluer des habiletés sous-jacentes à des compétences. Le modèle de Rey et de ses collègues peut s'inscrire dans cette perspective avec les trois degrés de compétences correspondant aux trois phases du modèle. Toutefois, seule la mise en œuvre de plusieurs tâches selon ce modèle permettrait de collecter suffisamment d'informations pour tenter un diagnostic en portant un jugement global sur l'ensemble des tâches en fonction des performances observées aux différentes phases. Un tel diagnostic est toutefois, selon nous, insuffisamment précis pour guider efficacement la mise en place de mesures correctives adaptées aux besoins de chaque élève en difficulté.

Une autre possibilité fait l'objet d'un courant de recherche en pleine évolution. L'idée est d'identifier un ensemble de ressources (habiletés cognitives) nécessaires à la bonne résolution des tâches proposées, puis de définir comment chacune d'elle est reliée aux éléments de réponses attendus dans les différentes tâches que comporte une même épreuve évaluative (donc plus apparentée à un test). Selon cette approche, des processus cognitifs, des savoirs, des savoir-faire, des attitudes, des stratégies, etc. peuvent faire l'objet du diagnostic. Ce dernier est inféré par la modélisation des données obtenues suite à la passation de l'épreuve à l'aide de modèles psychométriques de classification diagnostique MCD (voir par exemple, Loye, 2010). Le diagnostic prend ensuite la forme de la liste des ressources avec la mention maîtrisée ou non maîtrisée pour chaque élève. Un diagnostic sous cette forme a le potentiel d'aider à mettre en place des mesures correctives parfaitement adaptées aux besoins de chaque élève. Le modèle de Rey et de ses collègues ne pourrait éventuellement se prêter à cet exercice qu'en étant utilisé à répétition, sur un laps de temps assez long, ce qui limite l'intérêt du diagnostic.

Ainsi Rey et ses collègues proposent un type de tâche permettant d'évaluer les élèves dans une approche par compétence. Toutefois, il est impossible de faire des inférences à partir d'une seule tâche et ce modèle ne permet donc pas de soutenir un diagnostic valide et réellement utile pour mettre en place des mesures de remédiations efficaces.

4. Question 3 : Comment assurer la fiabilité et la validité des informations recueillies au cours d'une évaluation des compétences?

Pour répondre à cette question, il convient de considérer deux cas. Le premier cas concerne la validité et la fiabilité des informations recueillies lorsque l'évaluateur est directement confronté à la performance de l'élève dans une situation unique (qui peut comporter plusieurs questions ou étapes de résolution). Selon le niveau auquel cette tâche se situe dans la pyramide de Miller, la validité peut alors être envisagée de manière différente.

Dans le cas des trois niveaux inférieurs, elle est très fortement reliée à la qualité des outils d'évaluation qui sont développés et utilisés. Van der Vleuten (2010) considère d'ailleurs que dans ce cas la validité est avant tout assurée par la qualité du stimulus et son authenticité, et que le format sous lequel la réponse doit être donnée est de moindre importance (*principe 2*). Il peut donc ici s'agir aussi bien de tests objectifs ou de tâches associées à des grilles d'évaluation critériées (analytiques ou holistiques), mais le stimulus et le contenu visé doivent être soigneusement définis. Tous les formats nécessitent de prendre le temps de bien les développer et de les soumettre à différentes expérimentations avant de les considérer comme acceptablement valides. Un point de départ à la démarche peut être un modèle basé sur le principe de l'argumentation de Toulmin (1958) et que l'on retrouve notamment dans les travaux de Mislevy (voir par exemple: Mislevy, 2007). Selon ce modèle, la validité de l'évaluation est assurée par l'argumentaire qui permet de proposer une *prétention* relative au sujet qui a réalisé la tâche. Les actions du sujet fournissent les données relatives à la performance. Des garanties doivent être présentes relativement à la manière de collecter les données. Évidemment, plus les enjeux associés à l'évaluation sont élevés, plus il faut être attentif à la validation et à définir les conditions de passation. Lorsque les données récoltées sont quantifiées, les approches psychométriques traditionnelles, telles que la théorie de réponse à l'item, peuvent être utilisées pour contribuer à les valider. Toutefois, la diversité des tâches à utiliser pour évaluer les compétences aboutissent souvent à des données beaucoup plus qualitatives que quantitatives. Les quantifier de force pour appliquer des modèles psychométriques risque de faire perdre la richesse de l'évaluation qui a été planifiée.

Traditionnellement, la qualité d'une évaluation dépend de trois critères qui sont la validité externe, la validité interne et la fiabilité ou l'objectivité. Pour Guba et Lincoln (1989), ces critères se déclinent toutefois de manière différente lorsque l'évaluation est qualitative et basée sur des critères. Ainsi, la validité externe qui consiste en la possibilité de généraliser les résultats peut devenir une partielle *transférabilité* des résultats grâce à la richesse des descriptions provenant des critères; la validité interne correspondant notamment à la mise en évidence de liens entre les questions d'une même épreuve peut devenir la *crédibilité* par la correspondance entre la réponse donnée à la question et la reconstruction de cette réponse issue de l'évaluation, et enfin la fiabilité ou l'objectivité peuvent être vus comme la relative *stabilité* des résultats dans le temps. Ce dernier point est d'ailleurs à mettre en lien avec le principe selon lequel un jugement global serait plus stable et utile lorsque le recours à une approche objective et quantifiable est peu approprié (*principe 3*).

Au niveau du *Faire* de la pyramide de Miller, les tâches proposées doivent être aussi ancrées dans la réalité que possible et donc, par nature, éloignées de procédures standardisées. Dans ce cas, l'expertise de l'évaluateur joue le rôle le plus important pour assurer la validité du jugement qui sera porté à partir des données récoltées. Cela aboutit à l'idée que l'instrument est ici moins important que l'expert qui doit juger, et donc être bien préparé à le faire. Pour van der Vleuten et ses collègues (2010), l'évaluation sera alors dépendante du travail que l'évaluateur sera capable de faire avec l'outil à sa disposition, plutôt que de la qualité de l'outil lui-même. Le levier de la validité est donc l'évaluateur et pas l'outil. Ici encore, plus les enjeux sont élevés, plus ce levier doit être utilisé, par exemple en entraînant et en accompagnant les évaluateurs de manière régulière.

Le deuxième cas concerne l'évaluateur qui doit compiler des informations récoltées sur une période de temps assez longue et à partir de plusieurs sources, par exemple en utilisant un portfolio.

Dans ce cas, la qualité de la mise en œuvre du portfolio est de première importance. Il devrait contenir des tâches de diverses natures et portant sur une variété de contenus. Selon van der Vleuten et ses collègues (2010), le recours à plusieurs tâches est nécessaire pour assurer la fiabilité des données recueillies. Ils proposent d'utiliser entre 8 et 10 tâches avant de pouvoir faire une inférence fiable sur la compétence. En outre, si le jugement a été porté de manière globale sur chaque tâche, il sera plus facile de réaliser les inférences. Il convient donc ici d'appliquer le *principe 1* et de combiner des prises d'information à des moments différents, en variant les contenus, ainsi que le *principe 3* visant à favoriser un jugement plus global.

En ce qui a trait au portfolio, la littérature fournit de nombreux exemples de méthodes de validation. Certains (voir par exemple: Gadbury-Amyot, et al., 2003) s'appuient sur des approches traditionnelles et utilisent des procédures statistiques (corrélations, analyses de variance) pour établir la validité selon la définition de Messick (1989). D'autres ont démontré comment les critères de la recherche qualitative

permettent d'établir la validité d'un portfolio. Par exemple, Driessen et ses collègues (Driessen, van der Vleuten, Schuwirth, van Tartwijk, & Vermunt, 2005) ont construit empiriquement un modèle basé sur trois stratégies qui sont la triangulation des informations, le temps investi et le nombre de personnes pour juger les données. Ce modèle vise à assurer la crédibilité d'un portfolio.

Enfin, dans le cas de l'évaluation diagnostique basée sur des tests et dont les données sont modélisées avec des modèles de classification diagnostique (MCD), de nombreux travaux restent à faire pour étudier la stabilité des diagnostics, ou la qualité des liens entre les questions posées et les ressources identifiées comme nécessaires à une bonne réponse. Les méthodes pour assurer la validité et la fidélité des données issues de ces approches sont d'ordre psychométrique, mais éloignées des approches basées sur l'unidimensionnalité sur laquelle reposent la TCT ou la TRI.

Pour conclure, assurer la validité et la fiabilité des données recueillies lors d'une évaluation des compétences dépend totalement des outils d'évaluation utilisés. D'un côté, les tests permettent de collecter en une seule fois, une variété d'informations relatives à des contenus différents. Dans ce cas, il est possible de faire des inférences à partir des données, qui sont souvent plutôt quantitatives, puis d'envisager l'assurance qualité en termes de procédure psychométrique (TRI ou MCD). De l'autre côté, des tâches complexes permettent de collecter des informations plutôt qualitatives sur le développement des compétences. Dans ce cas, plusieurs tâches sont nécessaires avant de pouvoir réaliser des inférences. L'assurance qualité est alors assurée au niveau de chaque tâche notamment en ce qui a trait au stimulus utilisé, puis au niveau de la synthèse des différentes tâches en s'inspirant des critères de qualité issus des approches qualitatives telles que définies par Guba et Lincoln (1989), comme la transférabilité, la crédibilité ou la stabilité.

5. Références

- Driessen, E. W., van der Vleuten, C. P. M., Schuwirth, L. W. T., van Tartwijk, J., & Vermunt, J. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation : a case study. *Medical education*, 39, 214-220.
- Gadbury-Amyot, C. C., Kim, J., Palm, R. L., Muills, G. E., Noble, E., & Overman, P. R. (2003). Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program. *Journal of dental education*, 67(9), 991-1002.
- Gitomer, D. H., & Rock, D. (1993). Addressing process variables in test analysis. In N. Fredericksen, R. J. Mislevy & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 125-150). Hillsdale, NJ: Erlbaum.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park CA: Sage Publication.
- Legendre, R. (2005). Dictionnaire actuel de l'éducation, 3ème édition. Montréal: Guérin.
- Loye, N. (2010). 2010, odyssée des modèles de classification diagnostique (MCD). *Mesure et évaluation en éducation*, 33(3).
- Messick, S. (1989). Validity In R. L. Linn (Ed.), *Educational measurement: Third edition* (pp. 13-104). New York: Macmillan Publishing Co.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad Med*, 65(9), S63-S67.
- Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, 36(8), 463-469.
- Rey, B., Carette, V., Defrance, A., & Kahn, S. (2003). *Les compétences à l'école*. Bruxelles: De Boeck.
- Scallon, G. (Ed.). (1988). *L'évaluation formative des apprentissages, 1 et 2*. Québec: Les presses de l'Université Laval.
- Scallon, G. (Ed.). (2000). *L'évaluation formative*. Québec: ERPI.
- Toulmin, S. E. (1958). *The uses of argument*. UK: Cambridge: University Press.
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Currie, E. (2010). The assessment of professional competence: building blocks for theory development. *Best practice & research clinical obstetrics and gynecology*, 24, 703-719.

**DIDACTIQUE PROFESSIONNELLE ET EVALUATION :
LA VOIX DE L'EMANCIPATION**

Marie-France Baroth*

* Université de Lyon 2, france.baroth@gmail.com

Mots-clés : didactique professionnelle, compétences comportementales, développement, acteur/sujet

Résumé. *A partir d'un point de vue diachronique sur l'exercice de l'évaluation annuelle au cours de ces vingt dernières années dans une organisation, ce texte propose un regard critique sur l'évolution des instruments qui en étayent les pratiques. Si le passage des qualités aux compétences – notamment comportementales – ne garantit pas l'objectivité ni même l'acceptabilité des jugements produits, il laisse dans l'ombre toute une part pourtant réelle de l'activité. A partir de ce constat, la didactique professionnelle est alors envisagée comme un moyen de mieux appréhender l'activité, mais aussi comme un instrument de professionnalisation qui offre des perspectives de développement des acteurs impliqués dans l'évaluation, aussi bien que des organisations.*

Avant d'être un objet pour les sciences sociales, l'évaluation a été une pratique, distribuée dans de nombreux champs tels que ceux régis par les droits comptable, fiscal et civil notamment. De fait, attribuer une valeur à un fonds de commerce, estimer un patrimoine ou la réparation d'un préjudice comme donner une valeur au travail se heurtent à la même difficulté : celle de produire un jugement socialement acceptable. Si l'adoption de règles ou l'application de formules peuvent servir d'étayages, l'évaluation de l'activité, au-delà de la performance, ne se résout jamais par un calcul. Car les parties engagées y engagent bien plus que leur jeu d'acteur de la production. Il y est - plus qu'implicitement parfois - question de position, de posture, de légitimité, de jugement et de justice, de reconnaissance ou de négation de soi comme sujet entièrement investi dans son rapport au travail. Jusqu'à tenir l'évaluation pour responsable de la dislocation du collectif et de l'affaiblissement de l'individu dans un isolement pouvant le conduire à souffrir voire à commettre des actes désespérés au sein même de l'entreprise (cf. notamment C. Dejours, 2003).

A travers le prisme d'un dispositif d'évaluation tel qu'il est proposé aujourd'hui mais aussi tel qu'il existait avant d'avoir été réformé deux fois lors des vingt dernières années, nous regarderons comment il a évolué, en nous focalisant sur la façon dont les qualités de la personne sont devenues des compétences comportementales ; sur les modes d'évaluation de ces attributs subjectifs et leurs dérivés possibles ; et enfin sur ce que la didactique professionnelle peut apporter selon nous à leur évaluation en termes de méthode, et porter comme projet de développement pour les sujets.

1. Focus sur le passé d'un dispositif d'évaluation annuelle

Le dispositif d'évaluation investigué est celui d'une entreprise française à statut particulier, investie de missions de service public, dont les modes de fonctionnement et d'avancement se rapprochent de ceux que l'on rencontre dans l'administration. L'évolution des carrières y est très encadrée et se fait sur la double base :

- de l'ancienneté (avec un minimum requis pour accéder à l'échelon supérieur et un minimum d'avancement garanti, échelon s'entendant ici au sens générique du terme) ;
- et de l'évaluation.

Le contexte de cette entreprise présente une spécificité notable et probablement impliquante au regard de la pratique de l'évaluation, en ce qu'elle subit une exposition récente et progressive à une logique concurrentielle. Les éléments retenus ont été les termes de l'évaluation, au double sens d'objets sur lesquels cette dernière portait, et de mots pour les désigner.

1.1. Le dispositif antérieur à 1990 : de « l'appréciation des qualités » à la « notation »

Les objets désignés par l'évaluation annuelle étaient les « qualités », catégorisées en trois types : qualités générales, professionnelles et humaines. Tout d'abord « appréciées », elles donnaient lieu ensuite à une « notation » sur un total de cent points, répartis sur vingt critères, parfois regroupés sous les termes de « qualités intellectuelles », de « comportement au travail », de « connaissances professionnelles » ou encore de « rapports avec les collègues ». Des regroupements hétérogènes donc, puisque certains portaient sur des ressources de l'action telles que les connaissances, d'autres sur des comportements comme les relations du sujet avec ses « collègues » et « hiérarchiques », d'autres encore qualifiant le travail en tant qu'activité plus directement productive, dans ses dimensions de « rapidité » ou de « sûreté » par exemple. Les critères retenus pour chaque catégorie de qualités étaient les suivants :

Qualités générales, 5 critères,

- Présentation ;
- Faculté d'assimilation ;
- Jugement ;
- Mémoire ;
- Faculté d'expression.

Qualités professionnelles, 10 critères,

- Assiduité ;
- Ponctualité ;
- Ordre et méthode ;
- Rapidité ;
- Activité ;
- Sûreté ;
- Soin apporté dans l'exécution du travail ;
- Acquis professionnel ;
- Souci du perfectionnement ;
- Esprit d'initiative et sens des responsabilités.

Qualité humaines, 5 critères,

- Sens du travail en commun ;
- Sociabilité ;
- Aptitude à guider les collègues ;
- Attitude devant l'autorité ;
- Personnalité.

A chacun de ces critères étaient assortis de trois à cinq *indicateurs – types*, parmi lesquels l'évaluateur était invité à choisir celui qui lui paraissait le mieux qualifier la personne et sa façon d'agir vis-à-vis des règles de production, de ses collègues ou de sa hiérarchie. Avant de nous prononcer sur le processus, revenons sur le terme *d'indicateur-type*. C'est ainsi que nous avons désigné les différents items à cocher par l'évaluateur. Ils indiquaient bien la valeur décernée à une caractéristique isolée de l'activité de l'évalué, qu'ils semblaient décrire à travers un comportement

observé et retracé de façon concrète : « brouillon », « paresseux », « insuffisant » pour les moins élogieux, mais aussi « esprit très vif et pénétrant », « actif et plein d'ardeur », « acquis professionnel ne se limitant pas à son seul emploi » etc. Derrière le concret apparent de ces évocations, aucun lien explicite n'était fait avec la réalité. Ce ne sont que des expressions qui conduisent de perceptions artificiellement radicalisées par des positions réduites en nombre et, à des jugements de valeurs caricaturaux que seule l'asymétrie du rapport hiérarchique semble légitimer. Plus que des indicateurs, ce sont des inducteurs d'appréciation, sans fondement objectif démontré, et qui pourtant produisent mécaniquement une notation sans aucune mesure préalable. Le caractère stigmatisant de certains items procède d'une incursion manifeste dans la subjectivité de l'évalué. Au-delà de l'activité, c'est bien la personne que l'on qualifie de lente, paresseuse ou routinière. L'on pourrait reprocher au processus l'absence de référent clair. Bien sûr, les items les plus cotés fournissent des informations précises sur les caractéristiques attendues. Ainsi, un évalué faisant preuve « d'une grande aisance » dans l'expression, d'un « jugement rarement pris en défaut », d'une « assiduité remarquable », « d'un sens rare de l'organisation », d'un « esprit curieux » et « d'une personnalité qui s'impose » pour ne retenir que quelques unes des qualités les plus appréciées, bénéficierait de la notation maximum et s'approcherait de l'idéal véhiculé par l'institution. Mais tout cela ne dit rien sur les spécificités du travail confié à chacun, ni sur les résultats visés. Ce sont les ressources et la façon dont la personne les met à disposition de l'organisation qui sont jugées. Plus que la performance c'est l'engagement qui semble être standardisé, requis, pesé. A travers cet engagement, les sujets incarnent l'institution plus qu'ils n'assurent un produit. Et leur conduite fait d'ailleurs l'objet de prescriptions qui débordent le temps professionnel en termes de moralité par exemple. L'environnement non concurrentiel n'est peut-être pas étranger à cette tendance.

La touche manichéenne de cette « appréciation » manifestée par le terme de « qualité », si elle paraît surannée dans sa déclinaison et irrecevable au regard des contraintes d'objectivité affichées aujourd'hui dans tout exercice d'évaluation, ne manque cependant pas de cohérence : la sécurité de l'emploi et l'engagement subjectif qu'elle réclame répondent aux valeurs de sécurité, de probité et de transparence dans les missions assurées pour le collectif. Elles se heurtent à d'autres limites lorsque l'on s'intéresse au développement des sujets. En effet, la formation des adultes tient pour acquis que les compétences se développent, que les connaissances s'acquièrent et que l'expérience se construit. Mais qu'en est-il des qualités ? Caractéristiques naturelles de l'être, rien ne dit qu'elles soient irrémédiablement figées. Et si elles sont susceptibles d'évolution, il n'en est pas fait mention dans ce dispositif qui n'évoque pas d'action en ce sens : aucune des formations que l'entreprise accordait pourtant volontiers à cette époque (chaque employé avait droit à une semaine de formation par an), par le biais d'une offre fournie dans des domaines professionnels très spécialisés aussi bien qu'en développement personnel par exemple n'est prescrite.

1.2. De 1990 à 2010 : « contribution », « qualités » et « compétences »

En 1990, le dispositif d'évaluation annuelle a été réformé, avec l'objectif affiché d'y introduire plus d'objectivité, de transparence et de sincérité dans les jugements émis. Peut-être induites par le caractère peu élogieux des commentaires précédents lorsqu'ils se proposaient de signaler une variable à faire évoluer, les pratiques en la matière étaient volontiers sur-valorisantes. Et il était parfois difficile dans ces conditions de dégager des marges de progrès pour la personne évaluée, aussi bien que d'arbitrer entre plusieurs agents en concurrence pour l'attribution d'un avancement. La nouvelle organisation de l'évaluation s'est structurée en deux niveaux d'évaluation distincts :

- un premier niveau assuré par le n+1 en référence à une fiche d'activité décrivant les missions et les activités du poste occupé, ainsi que les principales qualités attendues. La fiche d'évaluation se centrait sur la « contribution de l'année écoulée », mesurée autant que possible par des éléments factuels et dénombrables (par exemple le nombre de sessions gérées par une assistante de formation). Cette contribution était bien sûr censée correspondre, par nature, aux activités référence. Mais en volume, aucune attente n'était formulée. La référentialisation en la matière procédait de fait des normes qui pouvaient se dégager des tendances observées

- d'une année sur l'autre, et du panel d'employés occupant le même type de poste, c'est-à-dire en valeurs relatives. Les variations du contexte requérant un ajustement quantitatif ou qualitatif de l'activité étaient mentionnées le cas échéant dans la rubrique « modalités d'exercice des activités, et principales qualités mises en œuvre », en regard d'une évocation de conditions dégradées par un turnover inhabituel par exemple, ou encore par un sous-effectif. Les « qualités » invoquées étaient directement liées à une liste de six qualités maximum, choisies par l'évaluateur dans un glossaire constitué à cet effet et consignées dans la fiche d'activité servant de référent. Des qualités typologiquement proches des critères du dispositif précédent, en ce qu'elles pouvaient indifféremment qualifier les ressources mobilisées par le sujet dans l'action professionnelle (« expertise », « autonomie »...), ses relations avec ses collègues (« esprit d'équipe », « sociabilité »...) ou encore ses dispositions vis-à-vis de l'activité (« polyvalence », « puissance de travail », « fiabilité »...);
- le second niveau d'évaluation, assuré par le n+2 : reprenant les termes de l'évaluation de la contribution mais cette fois sur les trois années écoulées, il y ajoutait une rubrique sur les « compétences professionnelles ». Absentes du glossaire, rien n'était dit sur ce que visaient les compétences en question, ni de la façon de les repérer, de les décrire, de les analyser ou encore de les mesurer. L'adjectif « professionnel » laissait entendre que toutes ne l'étaient pas et que seules les professionnelles serviraient d'assise à l'évaluation. Ce qui n'en constituait pas pour autant une définition susceptible d'orienter l'exercice. Les pratiques étaient donc largement laissées à l'initiative des évaluateurs, et potentiellement hétérogènes. Ce qui posait - notamment du point de vue des évalués - la question du fondement des propositions d'avancement auxquelles elles donnaient lieu. Mais aussi celle de la validité de l'articulation entre les compétences évaluées et la formation éventuellement prescrite au terme de l'entretien.

1.3. A partir de 2011 : l'apparition des compétences comportementales

La réforme de 2011 a remis l'accent sur la sincérité des évaluations notamment, sans insister cette fois sur l'objectivité (bien que la mobilisation d'éléments factuels et concrets soit toujours requise pour statuer sur la contribution). En revanche le triptyque « compétences, techniques, managériales¹ et comportementales » a été présenté comme le nouveau fondement de cette évaluation, et les compétences comportementales comme une innovation majeure. On y retrouve les deux niveaux distincts n+1 et n+2.

Selon la note d'accompagnement de ce nouveau dispositif, le premier niveau donne lieu à un compte rendu d'entretien faisant le « bilan de l'année écoulée » en termes de « réalisations » ou « contributions » par rapport aux objectifs fixés ; puis à un jugement « d'adéquation entre les compétences de l'agent et le poste ». Formellement, la fiche d'évaluation prévoit pour l'année écoulée : un cadre invitant à préciser « la contribution et les réalisations des objectifs de l'année », juxtaposées aux « compétences techniques, managériales et/ou comportementales mises en œuvre ». La prescription n'est pas très précise et l'on ne sait si les compétences mises en œuvre doivent être évaluées au même titre que les contributions et réalisations, ou si ces deux dimensions de l'activité s'articulent par un lien d'inférence par exemple. Une seconde rubrique reprend les mêmes items mais sous l'angle des perspectives pour l'année à venir, en associant contributions et objectifs d'une part, et compétences et axes de progrès d'autre part. Se prononcer sur l'adéquation des compétences au poste à partir de cette évaluation semble signifier que tout écart constaté dans l'atteinte des objectifs vient nécessairement d'un écart de compétences ; ce qui présente la compétence comme une variable d'ajustement des ressources humaines aux résultats visés. Et donc à asservir le développement des compétences à une logique économique : conception réductrice si l'on se place du point de vue du sujet ; légitime dans une perspective strictement productiviste.

¹ Les compétences managériales ne concernent évidemment que les employés assumant des fonctions d'encadrement. Nous ne nous sommes intéressés qu'aux agents dits « non-cadres ».

Quant à l'évaluation de second niveau, la note d'accompagnement précise que l'évaluateur n+2 évalue « la contribution de l'évalué sur le poste au travers d'une grille d'évaluation qui porte sur les compétences techniques, ainsi que sur les compétences comportementales », et que « les compétences techniques et comportementales exigées pour tenir le poste sont évaluées au moyen de différents critères précisés dans les fiches d'activités et reprises dans les fiches d'évaluation ». Lorsqu'on se reporte à ces fiches d'activités, on constate que le choix des critères d'évaluation des compétences techniques est laissé à la libre initiative de l'évaluateur, sans aucune indication. Ce dernier est donc supposé savoir précisément ce qu'est une compétence technique, et comment il peut formuler des critères pour l'évaluer. Concernant les compétences comportementales, un certain nombre de propositions est fait parmi lesquelles l'évaluateur a le choix :

- Produire un travail fiable / *fiabilité* ;
- Faire preuve d'initiative ;
- Savoir travailler en équipe / *esprit d'équipe* ;
- Communiquer / *communication* ;
- Être autonome dans le cadre des orientations reçues / *autonomie* ;
- Savoir rendre compte / *rendre compte* ;
- Développer ses compétences et partager son savoir faire / *acquisition de compétences* ;
- S'adapter au changement.

Nous y avons associé (en italique) les qualités présentes dans l'ancienne formule de l'évaluation et qui paraissent correspondre. Seules deux compétences comportementales étaient absentes de la liste des qualités : faire « preuve d'initiative » et « s'adapter au changement », qui signent l'ouverture de l'organisation à une dynamique désormais concurrentielle. Pour les autres, la correspondance est évidente, même si la forme de l'énoncé change : elle est passée de noms à des verbes, avec l'ajout marginal de détails. Ainsi, « esprit d'équipe » est devenu « savoir travailler en équipe », où le verbe savoir appelle non seulement une disposition de l'individu mais une injonction de résultat (savoir suppose que le résultat sera à la hauteur des attentes) ; il en est de même pour « rendre compte » devenu « savoir rendre compte » ; alors que « autonomie » s'est muée en « être autonome dans le cadre des orientations reçues ». Si cette extension en termes de résultats semble dire qu'être bien disposé au regard de l'activité ne suffit plus mais que cela doit se vérifier dans les réalisations, nous sommes bien à un niveau d'analyse qui relève de la performance. Mais la compétence n'est pas explicitée par ce type de formulation : elle est espérée en tant que moyen de la performance. Et les « critères » annoncés ne sont que des énoncés visant des types d'actions caractérisés au regard des types de résultats attendus. Quant à dire que l'évaluateur n+2 évalue « la contribution de l'évalué sur le poste au travers d'une grille d'évaluation qui porte sur les compétences techniques, ainsi que sur les compétences comportementales », cela revient soit :

- à faire un lien mécaniste qui part des compétences (qu'elles soient dites techniques ou comportementales) pour en inférer la contribution, ce qui serait pour le moins original. Si l'inférence qui est habituellement faite des performances pour appréhender les compétences - à défaut de pouvoir les observer directement et encore moins les mesurer - est contestée du fait de la non proportionnalité établie entre performances et compétences, on peut se demander quelle pertinence et quel intérêt présente le fait de partir de l'intangible (la compétence) pour évaluer la performance qui présente bien plus d'aspects visibles, objectifs et mesurable que les moyens subjectifs mis en œuvre pour la produire... ;
- soit à poursuivre dans une conception privilégiant l'engagement subjectif à la valeur même des réalisations qui en découlent. Ce qui reviendrait à poser que l'effort, le mérite, la conformité des dispositions du sujet aux attentes institutionnelles sont de plus grande valeur que le versant concret de la performance.

Si l'on considère que les propositions listées ci-dessus sont les critères annoncés de l'évaluation des compétences, qu'en est-il de la valeur qui leur est attribuée ? Elle résulte d'un positionnement parmi quatre indicateurs-types, qui permettent de situer chacune au regard des attentes manifestées à son encontre et servant de référent : « en dessous des attentes », « proche des attentes »,

« conforme aux attentes », « au dessus des attentes ». Nous arrivons ainsi à un procès de conformité, exprimé à travers des niveaux. Plusieurs observations peuvent alors être faites :

- c'est un processus s'apparentant à du contrôle plus qu'à de l'évaluation ;
- il pourrait donner lieu à une notation qui ne ferait que traduire en chiffres des appréciations devenues quantifiables ;
- ce qui n'est pas sans rappeler le mécanisme en vigueur avant 1993, les jugements figuratifs en moins.

2. Le tour que jouent les « compétences » à l'évaluation du travail

L'introduction du terme de « compétences » dans l'évaluation donne des accents de technicité RH et devrait conférer à l'exercice et à ceux qui le pratiquent la légitimité que produit tout jargon professionnel. Or les usages observés questionnent l'objectivité et la neutralité apparemment induites par ce vocable.

2.1. Rives et dérives : l'annexion du sujet par l'acteur

Passer des qualités aux compétences donne de prime abord l'impression que l'on ne s'arroge plus le droit de qualifier les personnes, et qu'une ligne de démarcation sépare désormais les seuls moyens de la performance (objets du contrat passé - qu'elle qu'en soit la forme - avec l'organisation à laquelle on vend sa force de travail) des attributs naturels et subjectifs. Pourtant on l'a vu, le rapprochement des « qualités » initiales et des « compétences comportementales » montre à quel point les jugements émis sont toujours de même nature, et comment ils envahissent la sphère privée du sujet plutôt que de se limiter à l'acteur de la production. Cela est flagrant lorsqu'on relève une des compétences comportementales proposées pour les cadres, et qui s'exprime dans les termes suivants : « faire confiance ». Non seulement la confiance nous semble être bien éloignée d'une compétence en ce qu'il s'agit d'un énoncé décontextualisé, et pointant en cela une qualité de relation d'ordre intime entre deux personnes. Mais il y a selon nous une confusion entre la confiance considérée comme ressource cognitive possible d'une action donnée, et reposant à ce titre sur des critères d'analyse et des indicateurs prélevés dans une situation de travail (quand bien même ils resteraient implicites voire incorporés au sujet), confiance pouvant donner lieu à un acte de délégation par exemple ou influencer sur l'exercice d'un contrôle ; et la confiance donnée dans une relation affective et non finalisée. Il pourrait s'agir d'une forme d'invasion du sujet versus acteur, à l'instar de celle évoquée par C. Dejours (2003) lorsqu'il déclare que le temps de travail déborde désormais sur la vie privée des sujets qui emportent parfois les préoccupations professionnelles jusque dans leur sommeil. L'absence de véritables critères d'évaluation et d'indicateurs concrets, alors qu'ils sont annoncés, porte à penser que l'apparente neutralité scientifique du concept de compétence, ainsi que la polyphonie de ses usages servent d'autant mieux cette manipulation des sujets, annexés par l'évaluation, asservis au-delà des frontières auxquelles devraient légitimement s'arrêter les opérations de mesure et de contrôle destinées aux seules activités productives.

2.2. Évaluation et pouvoir d'agir : la face cachée de l'activité

En marge des éléments d'analyse présentés ici, nous pouvons aussi évoquer les versants cachés des engagements subjectifs au travail mis en évidence par P. Rabardel (2005), lorsqu'il distingue le « pouvoir d'agir » de la « capacité d'agir ». Il n'est nul besoin de leur associer de nouvelles données pour appréhender à quel point les empêchements inhérents aux environnements et aux contextes de travail, qu'ils émanent des dispositifs techniques, réglementaires ou des décisions humaines, hiérarchiques ou non, influent non seulement sur les compétences à l'œuvre dans leurs dimensions les plus palpables, mais aussi sur les performances². De même en ce qui concerne le

² Se reporter à l'extrait de « l'établi » de R. Linhart (2003) repris par P. Rabardel (2005) et les péripéties de Demarcy.

« réel de l'activité » tel qu'Y. Clot l'oppose à « l'activité réalisée », et qui désigne tout ce que le sujet retient ou ne peut concrétiser du fait d'une cause interne ou externe. Si une évaluation stricte des performances sort légitimement de son champ ces engagements invisibles et ces frustrations, (bien que la discussion soit toujours possible à leur propos), qu'en est-il de la compétence ? Sauf à l'analyser et à mobiliser pour ce faire des méthodes rigoureuses et exigeantes, mais surtout le sujet dont les compétences sont évaluées, elles sont condamnées à rester dans l'ombre et à alimenter le sentiment d'inachevé et d'injustice chez ceux que l'on prétend mesurer sans aucun étalon. Le recours aux compétences et notamment aux compétences comportementales revient alors à transformer un artefact en artifice, dans un détournement ou une catachrèse qui questionne les usages et les motifs des concepteurs de ces dispositifs : méconnaissance ou malhonnêteté ?

Enfin, pour mémoire et réflexion, et avant de se demander comment l'on peut faire des compétences un instrument au service de l'amélioration de la production et des acteurs qui en sont chargés, est-il utile de rappeler les origines du terme « comportemental » ? Au-delà de son côté suranné qui ne suffit pas à lui enlever toute utilité (cf. les rappels synthétiques sur la réflexologie, Pavlov, Bekhterev, puis Watson et le behaviorisme pour la version américaine proposés faits par J. Friedrich, 2010), il est certainement bienvenu de se souvenir que bien des progrès ont été faits depuis le temps où l'on considérait que seul le versant visible de l'activité humaine pouvait en nourrir la connaissance, et que le développement ne se réduit pas à un dressage destiné à faire produire et reproduire des processus et des résultats. La didactique professionnelle nous semble répondre à d'autres ambitions, et pouvoir apporter sa pierre à la problématique de l'évaluation.

3. Le *re-tour* de la didactique professionnelle : méthodologie et projet de développement

Re-tour désigne à la fois la réponse que la didactique professionnelle propose à l'usage subjectivement abusif de la compétence tel qu'il est décrit ci-dessus ; et le retour sur l'action auquel elle engage par ses méthodes et ses principes d'analyse.

3.1. Une pédagogie autant qu'une didactique ?

La didactique professionnelle mobilise des méthodologies d'élicitation du travail empruntées à l'ergonomie, à la clinique et aux analyses de l'activité. Elles ont en commun l'implication du sujet qui, sous des formes diverses, est invité à participer à la mise en intelligibilité de sa propre activité. Quel que soit le protocole retenu, (débriefing, commentaires en situation, instruction au sosie, auto-confrontation, entretien d'explicitation, traces de l'action etc.), toutes ces modalités exploratoires sont alors utilisées à des fins de formation, notamment par l'analyste du travail pour modéliser les situations apprenantes, et élaborer des contenus, des simulateurs et autres artefacts visant le développement des compétences. Mais elles deviennent elles-mêmes, chemin faisant, des artefacts au service de ce développement : en mettant le sujet en position réflexive par rapport à son activité de travail, elles permettent la prise de conscience de son organisation intrinsèque, et favorisent la conceptualisation.

3.2. Une arithmétique ou un instrument de professionnalisation ?

Comment cette conceptualisation procède-t-elle ? La dimension d'analyse du travail est évidemment inhérente à la didactique professionnelle. Mais elle ne consiste pas à conférer à l'analyste une position asymétrique de sachant, en regard d'un opérateur qui ne serait que pourvoyeur externe des éléments d'information voire d'explication requis sur sa propre activité, tout en restant ignorant de son organisation incorporée et implicite. Les outils d'analyse fournis par les théories de la conceptualisation dans l'action (G. Vergnaud, 1998), génèrent la conceptualisation dans l'interaction. C'est par la rencontre des activités du chercheur et du sujet, autour d'un objet commun (l'action professionnelle) et dans le cadre de l'analyse, que le développement advient. Mais c'est aussi par cette analyse que les ressources mobilisées et les configurations spécifiques de cette mobilisation, adaptée à une situation, peuvent être identifiées. Ce qui amène plusieurs observations. Au regard de l'évaluation, l'identification des ressources et

des dynamiques spécifiées de la compétence permettent de fonder un jugement de valeur voire de conformité à leur égard. Au regard du développement, l'analyse du travail telle qu'elle est conduite et interagie dans la tradition de la didactique professionnelle est non seulement :

- instrument de professionnalisation pour ses praticiens, qui enrichissent leur connaissance des situations apprenantes auxquelles ils s'intéressent, et donc le corps des savoirs susceptibles d'être transmis par la formation ;
- mais aussi instrument de professionnalisation pour les acteurs du travail objet de l'analyse, qui développent leurs compétences à son occasion par la conceptualisation qu'elle induit. Instrument se référant à la théorie instrumentale de P. Rabardel, chacun instrumentant les éléments d'analyse convoqués, tels que les buts, les invariants opératoires, règles d'action et autres inférences en situation constitutifs du « schème » selon G. Vergnaud, qui propose en cela une définition scientifique de la compétence.

Au regard des compétences dites « comportementales », l'on voit bien dans le descriptif du mode opératoire de l'interaction et de ce qu'elle produit, qu'elle peut être envisagée comme une activité productive à part entière. Où dans notre cas les effets de l'interaction consistent dans les apprentissages conjoints réalisés par le chercheur et par l'acteur du travail investigué.

3.3. La voix de l'émancipation ?

Par la consistance théorique et scientifique que la didactique professionnelle donne aux compétences et à leur évaluation, elle sert doublement les évalués. En redessinant les contours de la compétence, elle permet de dépasser les critiques légitimes formulées à l'égard des « compétences comportementales » qui n'apparaissent que comme une version pseudo-scientifique des « savoir-être » pour se constituer en instrument de domination symbolique (E. Sulzer, 1999). Elle fournit à cet égard un cadre d'analyse, objectivant ce qui ressort de l'acte professionnel et des qualités de la personne, rendant au sujet ce qui lui appartient, et ce qu'il lui appartient de prêter à l'entreprise dans le rôle d'acteur qu'il accepte de composer et d'interpréter en échange d'une rétribution. Échange sur la base duquel une évaluation est acceptable afin de s'assurer de la bonne exécution du contrat qui lie les deux parties, dans leur intérêt conjoint. Mais elle se propose aussi comme un instrument de professionnalisation à mettre au service des évaluateurs, qui, formés à l'analyse du travail, pourraient jouer pleinement leur rôle de garant du développement des compétences de leurs collaborateurs - rôle qui leur est de plus en plus souvent confié en adjoignant à la fonction de manager une fonction d'acteur RH, pour laquelle ils n'ont la plupart du temps que peu de ressources -. Instrument à mettre aussi au service des évalués, à qui il serait ainsi donné l'occasion, par le biais de l'analyse, de conceptualiser leurs compétences tout en leur permettant de prendre du recul par rapport à une activité professionnelle trop souvent source de souffrance à force d'implication subjective subie. Plutôt que de leur enjoindre de partager leurs « bonnes pratiques » dans un mouvement de « refordisation partielle » (P. Azkenazy, 2004) tout en renforçant la pression prescriptive, cette posture originale est aussi une alternative à la sclérose des activités et des métiers.

4. En conclusion

L'automatisation devait nous affranchir des tâches pénibles et répétitives. Mais la peur de l'intelligence, de la ruse et du zèle a lesté l'activité d'une inflation de contrôles qui se superposent et polluent la productivité aussi bien que la qualité du travail. Les méthodes d'audit prévoient d'ailleurs trois niveaux de contrôles, au centre desquels l'activité semble parfois se prendre comme entre trois miroirs et ne plus en finir de se regarder, jusqu'à en perdre de vue ce qu'elle produit vraiment... Ce qui fait le jeu au passage de ceux dont le confort intellectuel se nourrit mieux de comptages, de lignes et de colonnes que de réflexions et de questions sur la finalité, la pertinence, l'inventivité ou l'ingéniosité des solutions et résolutions des défis quotidiens du travail réel, sans considération de la beauté du geste professionnel. Tandis que les résistants qui ne peuvent se résoudre à la tâche inutile même si elle est valorisée risquent, au péril de leur santé mentale,

l'aliénation sociale (selon les termes de C. Dejours, 2003) pour se livrer encore, corps et âme, à la satisfaction du travail bien fait.

Dans ces univers de travail où les injonctions se sont déplacées de la production vers la conformité et l'implication des sujets, où l'étalonnage des êtres et des souffrances s'exprime sur des échelles de 1 à 4 ou de 1 à 10 dans les mains des évaluateurs ou des médecins du travail, où les maux sont tus et tuent parfois faute de mots pour les dire, « la petite voix de la didactique professionnelle » n'incite pas seulement comme y invite P. Pastré (2011) au développement des sujets, mais peut s'élever aussi en faveur de leur émancipation vis-à-vis d'organisations devenues anthropophages, qui paraissent ne pas comprendre elles-mêmes ce qu'elles font ni ce qu'elles défont.

5. Références et bibliographie

Askenazy, P. (2004). Les désordres du travail. Paris : Seuil.

Clot, Y. (2011). Le travail à cœur, pour en finir avec les risques psychosociaux. Paris : Éditions La Découverte.

Dejours, C. (2003). L'évaluation du travail à l'épreuve du réel, critique des fondements de l'évaluation. *Sciences en questions*. Paris : INRA Éditions.

Friedrich, J. (2010). Lev Vygotski : médiation, apprentissage et développement, une lecture philosophique et épistémologique. *Carnets des sciences de l'éducation*. Université de Genève.

Linhart, R. (2003). L'établi. Paris : Les Éditions de Minuit.

Millet, J. Caspar, P. Normand, C. Évaluation, trappe ou tremplin ?. Paris : Éditions d'organisation.

Pastré, P. (2011). La didactique professionnelle, approche anthropologique du développement des adultes. Paris : PUF.

Sulzer, E. (1999). Objectiver les compétences d'interaction, critique sociale du savoir-être. *Éducation Permanente*, 140, 51-59.

Rabardel, P. (2005). Instrument subjectif et développement du pouvoir d'agir. Rabardel, P. et Pastré, P. (dir.), *Modèles du sujet pour la conception* (pp.11-29). Toulouse : Octarès.

Roelens, N. (2000). Intoxication productiviste et déshumanisation des rapports humains, une psychologue du travail analyse les causes de son épuisement professionnel. *Travailler*, 4, 93-122.

Vergnaud, G. (1998). Au fond de l'action, la conceptualisation. Barbier J-M (dir.), *Savoirs théoriques et savoirs d'action* (275-292). Paris : PUF.

QUE PEUT APPORTER LA DIDACTIQUE PROFESSIONNELLE À L'ÉVALUATION DES COMPÉTENCES ?

Florent Chenu*

* Université de Liège, florent.chenu@ulg.ac.be

Mots-clés : évaluation des compétences, didactique professionnelle, expérimentation

Résumé. Ce texte présente les réponses aux trois questions posées par Philippe Astier dans la cadre de la table ronde. Ces questions sont : Y a-t-il un modèle de l'évaluation des compétences en didactique professionnelle ? Quelles opérations (ou quelle activité) recouvre le terme d'évaluation pour la didactique professionnelle ? Quelles précautions, démarches, instances, règles, déontologie ou pratiques paraissent utiles pour que l'univers de l'évaluation en didactique professionnelle demeure celui du développement des sujets et non pas de l'asservissement, de la persécution, de la souffrance et de la destruction des personnes et des collectifs ? Les réponses sont basées sur une expérimentation réalisées dans le domaine de l'utilisation professionnelle d'un traitement de texte.

1. Y a-t-il un modèle de l'évaluation des compétences en didactique professionnelle ?

Ce n'est pas la première fois qu'il est question d'évaluation de compétences et de didactique professionnelle à l'ADMEE. Lors du colloque organisé en 2003 à l'Université de Liège, un atelier a été consacré à ce thème. En particulier, dans une communication intitulée *Comment la didactique professionnelle peut contribuer à l'évaluation des compétences professionnelles*, Patrick Mayen recommandait d'évaluer les compétences à l'aide de critères relatifs aux performances et aux modalités d'action mais aussi de prendre en compte dans cette évaluation la dimension cognitive des compétences : les conceptualisations construites pour raisonner et pour agir. Ceci pouvait se faire, selon lui, à travers un dialogue évaluatif entre évaluateur et évalué à propos d'une classe de situations la plus concrétisée possible.

Si je rapporte les propos développés par Mayen lors de cette communication, c'est parce que j'y ai assisté alors que je réalisais mes premiers travaux de recherche sur l'évaluation des compétences. Cette communication a orienté partiellement les projets de recherche que j'ai développés par la suite et m'a amené à m'intéresser à la didactique professionnelle.

La théorie de la conceptualisation développée par Gérard Vergnaud constitue un pilier fort de la didactique professionnelle. Elle-même s'inscrit dans la suite des apports du constructivisme de Piaget. Dans *La prise de conscience* (1974a) puis dans *Réussir et comprendre* (1974b), Piaget s'était déjà intéressé au lien entre performance et conceptualisation (et donc aux dimensions que Mayen (2003) propose de prendre en compte pour évaluer des compétences). Dans le deuxième ouvrage, il explique que dans le cas d'épreuves simples, « l'action précède la prise de conscience descriptive et la compréhension » (1974b, pp. 111-112). Cependant, il poursuit en indiquant qu'« au fur et à mesure de la complexité croissante des actions, celles-ci comportent une conceptualisation qui peut les devancer pour les guider, les accompagner ou les suivre de plus ou moins près » (Ibid.). Dans ses conclusions (pp. 241-242), il écrit :

Comprendre consiste à dégager la raison des choses tandis que réussir ne revient qu'à les utiliser avec succès, ce qui est certes une condition préalable de la compréhension, mais que celle-ci dépasse puisqu'elle en arrive à un savoir qui précède l'action et peut se passer d'elle.

Approfondissant la célèbre formule de Vergnaud « *Au fond de l'action, la conceptualisation* », la didactique professionnelle postule que l'action efficace, la compétence, est organisée et que cette organisation se présente sous la forme de ce que Pastré (2001, p. 6) appelle la *structure conceptuelle d'une situation* :

La compétence d'un opérateur ne se réduit évidemment pas à la maîtrise de la structure conceptuelle de la situation. Mais celle-ci constitue la clé de voûte de la compétence autour de laquelle vont s'organiser de façon hiérarchisée des habiletés, des procédures, la gestion de

ressources, etc. L'identification de la structure conceptuelle d'une classe de situations professionnelles correspond à la dimension cognitive des compétences.

Une structure conceptuelle peut être vue comme un système construit autour de concepts organisateurs (en nombre restreint) organisant globalement l'activité sur base de variables. Ces variables renvoient à des indicateurs, c'est-à-dire des observables permettant d'évaluer la valeur prise par ces variables dans une situation déterminée. Ce système permet de faire un diagnostic sur la situation en cours d'activité et d'orienter cette dernière en conséquence.

Selon Pastré, la compétence d'un opérateur est fonction de la proximité entre son modèle opératif et la structure conceptuelle de la situation à laquelle il est confronté (2008a, p. 60) :

[Avec la structure conceptuelle,] on se place du point de vue de la tâche, c'est-à-dire des éléments objectifs à prendre en compte pour que l'action soit efficace. Si on se place maintenant du côté de l'activité des sujets, on ne parlera plus de structure conceptuelle de la situation mais de modèle opératif. On voit tout de suite que le modèle opératif d'un acteur a tout intérêt à être fidèle à la structure conceptuelle de la situation pour que l'action soit efficace. Mais ce n'est pas garanti. On peut même penser que ce qui distingue les experts des novices est que ceux-ci ont un modèle opératif très fidèle à la structure conceptuelle de la situation, alors que ceux-là ont un modèle opératif approximatif dans la mesure où il n'est pas encore complètement construit.

Dans une perspective d'évaluation des compétences, la voie que l'on peut tracer à partir des apports de la didactique professionnelle consisterait à examiner dans quelle mesure une performance observée chez un individu constitue en fait le résultat observable du fonctionnement d'un modèle opératif fidèle à la structure conceptuelle de la situation dans laquelle on a observé cette performance. De cette manière, on rencontrerait ce que Mayen appelle le *défi* posé par l'évaluation des compétences, à savoir « trouver dans les situations et les expériences particulières ce qui ne s'y réduit pas mais comporte une dimension générique propre à assurer que le candidat peut agir dans une situation déterminée, mais pourrait aussi s'adapter à d'autres situations de la même classe » (2003, p. 303).

Comment dès lors appréhender, rendre visible, les modèles opératifs des évalués ? Si l'observation des performances est importante, le recours au langage et à la verbalisation apparaît comme une nécessité. Même si on peut se heurter au *laconisme des experts* (Pastré, 2001), examiner les analyses que les évalués peuvent mener après coup de leur action permettrait d'appréhender les conceptualisations sous-jacentes à celle-ci. Particulièrement, il conviendrait, selon Pastré (2008b), de dégager de ces discours des *jugements pragmatiques*, c'est-à-dire des énoncés généraux *apodictiques*, exprimés sous la forme *il faut que...*, *il doit...*. Ces jugements pragmatiques permettraient de comprendre le fondement et l'organisation de l'activité telle qu'elle existe chez un acteur à un moment donné. En les extrayant de ses verbalisations après coup, il serait possible de reconstituer son modèle opératif. Or, pour Pastré, seul le langage permet de repérer ces jugements pragmatiques : seul le langage permet de passer de l'*assertorique* (*les choses sont ceci, les choses sont cela*) à l'*apodictique* (*il est nécessaire que..., il faut que...*).

Existe-t-il un modèle de l'évaluation des compétences en didactique professionnelle ? Si la réponse est affirmative, je dirais que ce modèle doit consister en l'observation d'une performance en situation, éclairée par le discours que l'évalué est capable de tenir sur son action. Si le discours permet, à travers les énoncés apodictiques, de rendre compte d'un modèle opératif fidèle à la structure conceptuelle de la situation ayant servi à l'observation de la performance, on peut faire le pari que la performance est la manifestation observable d'une compétence plus générale qui conduirait à d'autres performances dans d'autres situations.

J'appartiens à une unité de recherche en éducation où la mise à l'épreuve des faits et la validation expérimentale des modèles constituent une préoccupation, une exigence et même une tradition. C'est dans la perspective d'apporter des éléments de validation du modèle qui vient d'être présenté que j'ai entrepris de mener des travaux de recherche. D'une manière générale, les questions que je me suis posées sont les suivantes :

- Est-ce que les sujets les plus performants sont ceux qui, dans leurs discours, font preuve d'une plus « grande » ou d'une « meilleure » conceptualisation ?
- Comment peut-on caractériser cette « meilleure » conceptualisation ?

- Les discours des sujets qui parlent « le mieux » de leur activité révèlent-ils des concepts pragmatiques et/ou organisateurs, des modèles opératifs renvoyant à une structure conceptuelle ? D'une manière générale, comment les actions sont-elles justifiées et quel est le contenu des explicitations ?

Pour y répondre à, j'ai réalisée une expérimentation portant sur 31 étudiants suivant un cursus professionnalisant dans le domaine de la bureautique. Je l'ai également répliquée auprès de quatre secrétaires chevronnées en vue de soumettre à une première validation les résultats et observations dégagés auprès des étudiants. Le plan expérimental avait pour but de mettre en relation trois types de données :

- les performances des sujets dans le cadre de deux tâches telles que celles traitées par des professionnels en entreprise (réaliser un tableau présentant les menus d'une semaine dans un internat, réaliser un tableau présentant les activités extrascolaires de trois ateliers organisés par un centre culturel) ;
- les différents types de justifications verbalisées dans le cadre d'un entretien d'explicitation invitant à revenir sur l'activité pendant la réalisation des tâches (la conceptualisation étant définie ici de manière très générale comme la compréhension de ce qu'on fait et pourquoi (Beckers, 2009, p. 10)) ;
- les scores à un test mesurant la maîtrise de procédures indépendamment de leur mobilisation en situation professionnelle (ex : mettez ceci en italique, insérez un tableau de 3 colonnes et 5 lignes, modifiez le styles de ces bordures, etc.) Dans l'esprit de la didactique professionnelle, on pourrait parler de concepts techniques non pragmatisés, non inscrits dans une sémantique de l'action.

Les résultats mettent en évidence que les sujets les plus performants (ceux qui réalisent les meilleures productions aux deux tâches) sont aussi ceux qui justifient le plus d'actions de manière rationnelle. Ces résultats quantitatifs sont développés ailleurs (Chenu, 2012 ; Chenu, soumis) et j'insisterai ici uniquement sur la conclusion à laquelle ils aboutissent : oui, performance et conceptualisation (au sens de justification de l'action) sont liées. Ceux qui parlent le mieux de leur action (le plus rationnellement) sont ceux qui produisent les meilleurs documents.

A côté des résultats quantitatifs, une observation qualitative me semble essentielle à rapporter. L'analyse des discours des stagiaires qui « parlent le mieux » de leur action a permis de mettre en évidence un grand nombre de *règles d'action* qui peuvent être, comme le dit Vergnaud (2001), énoncées sous la forme *si... alors...* Toutefois, ces règles ne semblent pas renvoyer à un fonctionnement selon un modèle opératif et donc à une structure conceptuelle tels que décrits par la didactique professionnelle.

Alors que pour Vergnaud (2001, p. 6), les raisons qui « *relient les différentes conditions possibles et les différentes activités qui leur sont associées* » sont d'ordre conceptuel (les concepts-en-acte et les théorèmes-en-acte), les justifications recueillies ici semblent plutôt se rapporter à des *préoccupations*. Celles-ci sont de trois types :

- La lisibilité : Comment le destinataire va-t-il lire le document ? Qu'est-ce qui est important pour lui ? Comment faciliter cette lecture ?
- La convenance : Quelles sont les attentes explicites et implicites du commanditaire de la tâche ? Quel but veut-il atteindre avec le document qui doit être élaboré ?
- L'économie : Comment est-il possible de travailler plus vite ? De gagner du temps ? D'éviter des opérations inutiles ? De réduire les risques d'erreurs ? De rendre le document facilement réutilisable à un autre moment et/ou par quelqu'un d'autre ?

Dans l'analyse des résultats de notre expérimentation, la première préoccupation apparaît comme une clé importante de l'action efficace. Elle fait écho à des résultats de Kunégel (2007) et de Mayen (2008) notamment, qui considèrent que la centration de l'opérateur sur le bénéficiaire, dans le domaine de la mécanique et de la carrosserie automobile pour le premier et dans le domaine des soins aux personnes pour le second, constitue un indicateur important du développement professionnel.

Dans une étude sur la conduite d'entretiens par des formateurs, Michel & Allemand (2008) parlent d'un élément organisateur au sujet de la centration sur l'interviewé. Toutefois, cet organisateur n'est pas d'ordre conceptuel, mais plutôt de l'ordre de l'attitude et de la posture. Ils étayent leur analyse par des considérations similaires de Pastré sur l'activité enseignante : « à côté d'organisateur plus classiques de l'activité enseignante, on peut penser qu'il existe des organisateurs de plus grande profondeur, portant sur le développement » (2007, p. 91). Pastré commente d'ailleurs les analyses de Michel et Allemand dans le même sens qu'eux :

Aussitôt qu'on a affaire à l'activité d'un humain sur et avec d'autres humains, il se pose la question d'attribution réciproque : qu'est-ce que ego attend d'alter ? Qu'est-ce que alter attend d'ego ? Qu'est-ce qu'ego comprend de l'attente d'alter ? etc. Autrement dit, la posture des acteurs fait partie intégrante de l'organisation de leur activité. Dans les apprentissages à dominante technique et à dominante gestuelle on pourrait aussi trouver cette question de posture, mais on peut penser qu'elle ne fait pas partie du noyau central des concepts organisateurs, alors qu'elle est centrale dans la conduite d'un entretien. (2008c, p. 132)

Je pense que cette analyse conjointe de Michel & Allemand et de Pastré s'applique également très bien aux observations relatives au rôle des préoccupations dans l'action : elles en constituent un élément organisateur important tout en étant de l'ordre de l'attitude et de la posture, pas du concept. En conclusion, évaluer des compétences selon un modèle construit avec le cadre théorique de la didactique professionnelle reviendrait à observer une performance en situation et à analyser le discours que l'évalué est capable de tenir sur son action, en vue d'en repérer les éléments organisateurs, qu'ils soient d'ordre conceptuel et/ou de l'ordre de la posture.

2. Quelles opérations (ou quelle activité) recouvre le terme d'évaluation pour la didactique professionnelle ?

J'ai pris beaucoup de temps pour répondre à la 1^{re} question. Ceci était d'autant plus nécessaire que le modèle d'évaluation que j'ai présenté conditionne, dans une certaine mesure, l'activité qui est attendue de l'évaluateur qui l'adopte.

Selon ce modèle, le processus d'évaluation se déroule en deux phases : une phase d'observation de performance(s) suivie d'une phase d'explicitation en lien avec la production de cette performance. Dans ce cadre, l'évaluateur est amené à remplir plusieurs missions :

1. Sélectionner la ou les situations qui serviront à observer les performances. Celles-ci doivent être emblématiques du champ professionnel concerné par l'évaluation et les plus authentiques possible ;
2. Juger les performances des évalués à l'aide de critères de performances et de critères de modalités d'action. Ceci nécessite une observation fine des productions et des démarches mises en œuvre ;
3. Mener un entretien d'explicitation en vue de mettre en évidence les éléments organisateurs de l'action de l'évalué ;
4. Si l'évaluation est réalisée dans un but certificatif, intégrer les résultats relatifs à la performance et l'analyse du discours après coup pour émettre un jugement global sur la compétence de la personne évaluée.

Ces différentes tâches sont exigeantes pour l'évaluateur. Personnellement, c'est la réalisation de l'entretien d'explicitation et la formulation d'un jugement global qui me semblent les plus problématiques. C'est pourquoi c'est celles-là que j'ai choisi de développer dans mon intervention.

Vermersh (1994) a développé plusieurs recommandations intéressantes à considérer pour mener un entretien d'explicitation. Selon lui, obtenir une réponse non tronquée à la question du *pourquoi* ne constitue pas une évidence. Or, cette sans doute la question la plus riche en enseignement lorsqu'on évalue des compétences. Dans cette optique, il est souhaitable de passer par la description (*comment*) avant d'aller vers la justification et l'abstraction (*pourquoi*) :

Si je suis [directement] à un niveau plus abstrait, la verbalisation va porter non plus sur l'action, mais sur la structure de cette action, les invariants propres à la réalisation de cette classe de tâches. Je risque de sauter une étape dans la collecte des informations relatives au faire. Pour aller vers la généralisation de manière valide, il est nécessaire d'assurer une collecte correcte au niveau de l'action telle qu'elle est incarnée dans une action spécifiée. (p. 53)

Systématiquement ramener le sujet interrogé à ce qu'il a concrètement fait en situation et limiter au les discours généraux sans lien direct avec l'activité m'apparaît être une nécessité absolue dans la conduite d'un entretien d'explicitation. En effet, Saint-Arnaud souligne la tendance des professionnels à développer des discours sans lien avec leur activité (2001, pp. 20-21) :

Dans les situations difficiles, il y a un écart systématique entre la théorie professée par l'acteur (epoused theory) et la théorie d'usage (theory-in-use) qu'on peut inférer à partir du

comportement de l'acteur [...] Lorsque les praticiens réfléchissent sur leur action, ils s'évaluent en fonction de ce qu'on peut appeler un surmoi professionnel. Dans chaque discipline, on doit respecter des valeurs, des principes et des normes dont l'efficacité professionnelle est reconnue : il y a des choses qui se font et d'autres qui ne se font pas.

Poser la question du pourquoi sans pouvoir la poser directement et explicitement et en partant systématiquement de ce qui a été fait exige une activité complexe de la part de l'évaluateur qui conduit l'entretien après coup. Cela nécessite de bien connaître le domaine professionnel évalué et, dans une certaine mesure, de pouvoir se mettre « à la place » de l'évalué pour comprendre ses réponses. C'est à ce prix que des les raisonnements ou/et des éléments organisateurs de l'action, conceptuels ou attitudeux, peuvent être appréhendés

Pour les entretiens que j'ai réalisés, certaines techniques m'ont permis de susciter plus facilement ce type d'explicitation. En voici quelques exemples :

- Exploiter les erreurs, c'est-à-dire les retours en arrière, quand le bouton Annuler a été utilisé, quand des parties du travail ont été supprimées et/ou recommencées, quand des mots déjà tapés ont été modifiés,... Faire décrire ces situations et demander, par exemple, « qu'est-ce qui s'est passé dans ta tête à ce moment-là ? », « qu'est-ce qui te dérangeait ? » fournit de l'information de qualité en quantité ;
- Revenir sur les hésitations, les soupirs, les verbalisations spontanées en cours d'activité. Là-aussi, les mêmes questions peuvent être posées ;
- Comparer des actions par rapport à des situations « identiques en apparence ». Jouer l'avocat du diable. Demander pourquoi une règle ou un principe que la personne dit avoir suivi dans un cas ne l'est pas dans un autre ;
- Demander d'adopter une autre position : « si tu étais prof de bureautique et que tu devais enseigner le principe que tu respectes en faisant ceci ou cela, tu dirais quoi ? » ;
- Demander au sujet les avantages de telle ou telle manière fonctionner, de faire tel ou tel choix. Demander quels sont les inconvénients qui sont évités.

Une autre source de complexité dans l'activité de l'évaluateur qui adopterait le modèle d'évaluation construit avec la didactique professionnelle pour cadre théorique réside, selon moi, dans la nécessaire synthèse à établir à partir des informations que le processus a permis d'engendrer. C'est lorsqu'on poursuit une finalité de certification que le problème se pose.

Si on imagine relativement bien pouvoir mettre au point des grilles de critères permettant d'apprécier la performance et, sur cette base, déterminer qui « réussit » et qui « échoue », les choses se corsent lorsqu'une démarche similaire doit être adoptée pour « apprécier la qualité de l'analyse après coup ». En effet, tout comme il apparaît nécessaire pour considérer qu'une performance est atteinte, que l'évalué rencontre un certain nombre de critères (dont certains décisifs), l'entretien d'explicitation devrait permettre de s'assurer que tels ou tels raisonnements décisifs sont effectivement menés. Ceci ramène à une question générale que je posais au début de mes travaux (Chenu, 2004a) : à partir de quel « seuil » de justification, de conceptualisation, de réflexivité peut-on considérer que le sujet est suffisamment « réflexif » pour être efficace ? Qui « réussit » l'entretien d'explicitation et qui « échoue » ? Y a-t-il des justifications rationnelles qu'il faut absolument être capable de formuler pour que le discours soit jugé satisfaisant ? Inversement, y a-t-il des (types de) justifications non rationnelles qui ne peuvent apparaître dans un discours acceptable ? Et comment juger le discours d'un candidat qui justifie par des éléments non rationnels (préférences personnelles, absence de justification, centration sur soi, indices de surface, présupposés, ...) tout en prenant conscience du caractère inapproprié de ces justifications au moment de l'entretien d'explicitation ?

Ces questions sont d'autant plus importantes qu'une des observations claires qui se dégagent de l'expérimentation est que très peu de sujets se caractérisent par des justifications d'un seul type. Parmi l'ensemble des verbalisations relatives aux actions d'un même sujet, on en observe certaines reposant sur des éléments rationnels, d'autres reposant sur des éléments superficiels. Le degré de rationalité varie selon l'action concernée. Chez certains sujets, il est possible de dégager un type de justifications dominant. Chez d'autres, les choses sont beaucoup moins claires : les justifications de différents types se côtoient de manière relativement équilibrée. Certains sujets font même intervenir différents modes de justification à l'intérieur d'une même explicitation d'action. Dès lors, dans le cadre d'une évaluation certificative, comment peut-on formuler une appréciation globale sur base d'une telle mixité ? Comment faire la part des choses entre le superficiel et le rationnel ? Bien sûr, on peut espérer être capable d'évaluer à leur juste valeur, sans trop se tromper, les discours des sujets qui justifient très superficiellement ou très rationnellement. Mais comment dire, pour les sujets justifiant différemment selon les actions concernées, si leur discours est « suffisamment argumenté » ou pas ? La

praticabilité de la prise en compte dans une évaluation certificative des propos recueillis dans le cadre d'un entretien d'explicitation se limite-t-elle aux cas « clairement catégorisables » ?

3. Quelles précautions, démarches, instances, règles, déontologie ou pratiques paraissent utiles pour que l'univers de l'évaluation en didactique professionnelle demeure celui du développement des sujets et non pas de l'asservissement, de la persécution, de la souffrance et de la destruction des personnes et des collectifs ?

Avant de répondre à cette dernière question, je voudrais préciser que, pour moi, l'évaluation et ses méthodologies ne doivent pas être considérées comme néfastes en tant que telles. C'est surtout l'objectif poursuivi quand on y recourt qu'il importe d'interroger. Au-delà du questionnement des intentions, concevoir et utiliser des outils et des méthodologies d'évaluation fiables, c'est-à-dire principalement fidèles et valides, c'est aussi lutter contre l'injustice des jugements subjectifs voire arbitraires. Aujourd'hui, l'évaluation fait partie intégrante de la vie scolaire et professionnelle. Son rôle est crucial à certains moments : examens de fin d'études, concours pour accéder à un cursus, épreuves de sélection pour un emploi, procédures de VAE... Des outils non fiables, peu valides, inéquitables et/ou affichant une caution scientifique douteuse ne devraient pas avoir leur place dans les procédures d'évaluation qui mettent en jeu l'avenir des évalués. C'est avec cette préoccupation de justice que je voudrais répondre à la dernière question. J'évoquerai deux problèmes posés par l'évaluation des compétences telles que je l'ai modélisée à partir de la didactique professionnelle.

Premièrement, la capacité d'un évalué, lors d'un entretien d'explicitation, à tenir un discours « approprié » sur son activité doit être relativisée. En effet, je pense que ce que la personne est capable d'explicitier dépend non seulement du degré de conceptualisation et/ou d'organisation de son activité, mais aussi de la qualité du questionnement qui l'incite à verbaliser. Or, cette qualité est susceptible de varier d'un évaluateur à l'autre, voire chez un même évaluateur ! C'est sur base de l'expérience que j'ai vécue dans la conduite des entretiens que je m'autorise cette affirmation. Lors de ceux-ci, les actions observées en situation étaient questionnées après coup avec une des deux intentions suivantes :

- soit vérifier, dans les propos du sujet, la présence de justifications que j'avais déjà repérées chez d'autres sujets rencontrés préalablement ;
- soit découvrir un nouveau type de justification, encore non envisagé ou non observé chez quelqu'un d'autre. Dans ce cas, je ne faisais pas d'hypothèse a priori sur le pourquoi des actions : le but était vraiment de comprendre.

Ceci fait qu'au fur et à mesure de la réalisation des entretiens, la qualité de mon questionnement s'est améliorée : il est plus facile de faire verbaliser quelque chose quand on connaît déjà différentes réponses possibles auxquelles on peut s'attendre. Et, plus on réalise d'entretiens, plus on connaît de réponses possibles. Par ailleurs, les techniques de questionnement s'améliorent. On sait beaucoup mieux comment poser les questions au 30^e entretien qu'au 1^{er}. Les questions « qui fonctionnent bien », qui facilitent la verbalisation, sont découvertes au fur à mesure des entretiens et reprises systématiquement dans les entretiens ultérieurs.

Dans le cadre de l'expérimentation, cette variation ne semble pas avoir constitué un biais trop important dans la mesure où les liens entre performance et conceptualisations apparaissent malgré tout. Cependant, c'est cette observation qui m'incite à penser que la qualité des verbalisations et des justifications d'un sujet, sa capacité à expliciter, est en partie tributaire de la qualité du questionnement. Dès lors, la réussite dépend en partie des personnes qui conduisent l'entretien et de leur expérience de cette conduite au moment de l'évaluation. Dans une perspective de certification, ce problème de fidélité, et donc de justice, est préoccupant.

Le deuxième problème que je voudrais soulever concerne le regroupement de situations en familles. J'y ai également consacré une recherche dont les résultats ont été présentés au 17^e colloque de l'ADMEE à Lisbonne (Chenu 2004b) puis publiés (Chenu, 2005, 2006).

La notion de famille de situations est intimement liée à celle de compétence ou, comme le dit Crahay (2002), « le concept de compétence subsume celui de familles de situations » (p. 199). Pour construire des référentiels de compétences, qui serviront éventuellement à évaluer, on délimite des familles de situations à partir desquelles on infère des compétences censées y faire écho. C'est donc la famille de situation qui détermine la compétence, et non l'inverse. L'étendue d'une famille de situations – les situations qu'on jugera comme appartenant à la même famille – délimite la zone de transférabilité de la compétence qu'on y associe, cette zone impactant le degré de

généralité de la compétence qu'on inscrit dans le référentiel : Plus une famille regroupe des situations diversifiées, plus la formulation de l'intitulé de la compétence du référentiel est générale. Dès lors, si on se soucie d'évaluer des compétences de manière valide, il faut pouvoir définir, de manière tout aussi valide, l'étendue des familles de situations dans lesquelles ces compétences sont mobilisées. Questionner l'objectivité du regroupement de situations en familles est donc important.

Le problème posé ici renvoie à des préoccupations d'équité entre les différents partenaires impliqués dans l'élaboration de référentiels. Comme le souligne le CEREQ (1999, p. 2), « construire un référentiel en vue d'une évaluation consiste [...] à rechercher les dénominateurs communs entre des situations diverses, exercice dont la difficulté s'accroît lorsque les certifications concernent plusieurs secteurs d'activités [...] On entrevoit ici la tension entre la recherche de critères communs élargissant la validité de la certification et la spécificité des compétences que l'on tente d'évaluer. Un référentiel de certification est donc toujours le fruit d'un consensus où l'on tente de réduire cette tension entre le général et le spécifique. ». Pour Alpin & Shackleton (1997, p.8), cette tension entre général et spécifique et le consensus qu'elle nécessite peuvent générer des tensions entre employeurs et employés : « un des problèmes est le conflit d'intérêt qui existe entre les salariés qui désirent que les compétences acquises au travail soient reconnues et certifiées afin d'être transférables, et les employeurs qui craignent de perdre un personnel formé à grands frais du fait de la mobilité accrue qui en découle pour les salariés ». Lorsque j'ai présenté mes travaux sur les familles de situations au colloque de l'ADMEE, Françoise Ropé¹ a insisté sur l'importance de l'égalité des points de vue des acteurs concernés par la construction de référentiels. Selon elle, si aucune démarche objective ne permet de regrouper des situations en familles (et de construire ainsi les référentiels), tant les employés que les employeurs doivent pouvoir donner leur avis sur ce regroupement, l'avis des uns n'étant pas plus « légitime » que celui des autres. Or, plusieurs auteurs (Crahay, (2002, 2006), Rey (1996) et Perrenoud (2010) notamment) se montrent assez sceptiques quant à la possibilité de regrouper objectivement, scientifiquement, des situations en familles.

C'est le partage de ce scepticisme qui m'a conduit à consacrer une étude exploratoire à propos de la notion de famille de situations. Cette recherche a consisté à soumettre 5 paires de descriptions de situations de type professionnel en bureautique (situations présentées deux par deux) à 40 personnes (formateurs, professionnels, étudiants, employeurs, tous dans le domaine de la bureautique) et à leur demander de juger globalement, sur une échelle de type Likert, dans quelle mesure les situations de chaque paire étaient similaires ou non. Les personnes argumentaient ensuite leurs jugements.

Le tableau suivant présente comment ont été jugées les situations de chaque paire par les 40 personnes.

Tabelau 1 : Jugements sur la similarité des situations

	Différentes	Légèrement identiques	Assez identiques	Fortement identiques	Totalement identiques
Paire 1	13	8	7	9	3
Paire 2	10	5	16	7	2
Paire 3	20	9	5	6	0
Paire 4	18	12	5	3	2
Paire 5	18	5	9	5	3

Dans quatre cas sur cinq, tous les échelons de l'échelle ont été choisis et on ne retrouve jamais plus de 50% de la population sur un seul échelon. Les résultats montrent que les jugements sont extrêmement variables selon les sujets, certains considérant deux situations comme totalement identiques, d'autres considérant les deux mêmes situations comme différentes. C'est sur base de ces résultats que je soutiens l'idée que le fait de juger de la similarité de situations est subjectif, c'est-à-dire lié aux sujets. Chaque sujet, sans doute en fonction de son expertise et de son habitus socioprofessionnel, adopte des critères et un grain d'observation qui lui sont propres pour émettre son avis.

Face à cette subjectivité, la didactique professionnelle m'a semblé, à une période, proposer une alternative intéressante pour regrouper des situations en famille. Plus particulièrement, un article de Pierre Pastré avait suscité mon intérêt. Il y émettait l'idée que le fait de considérer deux situations comme similaires ou non dépend de l'individu qui les examine. Voici ce qu'il écrivait (2001, p. 6) :

¹ Cette sociologue française s'est notamment intéressée aux compétences dans Ropé (1999) et Ropé & Tanguy (1994).

Plus le sujet a réussi à trouver de l'invariance à un haut niveau d'abstraction, plus sa capacité d'adaptation aux variations des situations va s'accroître.

Selon ces propos, c'est le sujet qui trouve de l'invariance entre les situations. En cela, ce que je lisais rejoignait mon point de vue sur le caractère subjectif du regroupement de situations en familles. Toutefois, un peu plus loin dans le même texte, Pastré semblait également parler d'une invariance absolue, intrinsèquement liée à la situation :

Pour chaque classe de situations professionnelles, il existe une structure conceptuelle, qu'on peut dégager par l'analyse cognitive de la tâche et qui est en quelque sorte le socle invariant qui va servir à organiser l'action efficace du sujet [...] L'identification de la structure conceptuelle d'une classe de situations professionnelles correspond à la dimension cognitive des compétences » (Pastré, 2001, p. 6).

En complétant ses propos, Pastré défendait l'idée que, même si c'est l'individu qui « trouve » de l'invariance entre situations, il existe toutefois un socle invariant qui constitue un attribut de ces situations - la structure conceptuelle - et qu'il semble nécessaire que le sujet « fasse sien » pour agir efficacement. Autrement dit, si les individus peuvent trouver de l'invariance à différents niveaux entre situations, ceux qui sont efficaces dans ces situations ont comme particularité commune le fait d'atteindre un niveau minimum (*un socle*) d'invariance. Dès lors, c'est le fait, pour des situations, de partager ce même socle, cette même structure conceptuelle, qui permet qu'on les considère, dans l'absolu, comme appartenant à la même famille. Pour moi, à travers ces quelques lignes, Pastré réussissait, à la fois, à conforter l'idée d'un regroupement subjectif des situations en familles, tout en laissant ouverte la possibilité de constituer de manière objective, « désindividualisée », des familles de situations, ou, pour reprendre ses termes, des *classes de situations*.

Malheureusement, comme je l'ai expliqué plus haut, je n'ai pas pu mettre en évidence de modèle opératif renvoyant à une telle structure conceptuelle dans le domaine de la bureautique. Ceci m'incite à exprimer quelques réserves sur l'intérêt de cette voie pour construire des familles de situations dans différents champs professionnels. Le problème ne me semble pas prêt d'être solutionné. Dès lors, parler de la notion de famille de situations avec prudence et être conscient de sa relativité et de son manque de stabilité scientifique me paraît constituer une précaution essentielle que doit prendre toute personne qui se propose d'évaluer des compétences professionnelles.

Références bibliographiques

- Alpin, C., & Shackelton, J.R. (1997). Les tendances du marché de l'emploi et les besoins d'information : leur impact sur les politiques du personnel. *Revue Européenne de Formation Professionnelle*, 12, 7-14.
- Beckers, J. (2009). Contribuer à la formation de « praticiens réflexifs ». *Puzzle*, 26, 4-14.
- Centre d'Etude et de Recherche sur les Qualifications (CEREQ). (1999). Evaluer les acquis de l'expérience. Entre normes de certification et singularité des parcours professionnels. *Bref*, 159.
- Chenu, F. (2004). Réflexion sur les obstacles et les paradoxes d'une évaluation des compétences professionnelles. In *Actes du 3e Congrès des Chercheurs en Education* (pp. 119-122). Bruxelles : Ministère de la Communauté française.
- Chenu, F. (2004b). La dimension "famille de situations" de la notion de compétence: Quelle objectivité? Quelle validité? In *Actes du 17e Colloque de l'Association pour le Développement des Méthodologies d'Evaluation en Education - Reconnaître et valoriser les acquis de l'expérience*. Lisbonne.
- Chenu, F. (2005b). Les compétences et les familles de situations. Etude exploratoire de la complexité d'un jugement. *Les Cahiers du Service de Pédagogie expérimentale*, 21-22, 103-125.

- Chenu, F. (2006). La dimension "famille de situations" de la notion de compétence: Quelle objectivité? Quelle validité ? In Alves, P., Figari, G., Rodrigues, P. & Valois, P. (Ed.), *Evaluation des compétences et apprentissages expérientiels : Savoirs, modèles et méthodes*. Educa : Lisboa.
- Chenu, F. (2012). *L'évaluation des compétences. De la méthodologie de la réflexivité garante de la transférabilité aux perspectives ouvertes par la didactique professionnelle*. Thèse de doctorat en sciences psychologiques et de l'éducation. Université de Liège.
- Chenu, F. (soumis). Eclairer les performances par un entretien d'explicitation pour évaluer des compétences professionnelles : intérêt et limites. In R. Goasdoué, L. Mottier Lopez & M. Vantourout, *L'évaluation des compétences à l'épreuve du réel*, Bruxelles : De Boeck.
- Crahay, M. (2002). Rapport de l'atelier sur le thème « savoirs, compétences disciplinaires, compétences transversales : évolution ou révolution. In *Actes du 2e congrès des Chercheurs en Education*. Bruxelles : Ministère de la Communauté Française.
- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue Française de Pédagogie*, 154, 97-110.
- Kunégel, P. (2007). Que font les tuteurs ? Une exploration de la partie énigmatique de l'alternance. *Education permanente*, 173, 109-119.
- Mayen, P. (2003). Comment la didactique professionnelle peut contribuer à l'évaluation des compétences professionnelles. *Cahiers du Service de Pédagogie expérimentale*, 15-16, 301-310.
- Mayen, P. (2008). Incompétences et compétences : peut-on définir des repères de professionnalité ? In L. Mottier Lopez, Y.-E. Dizerens, G. Marcoux. & A. Perréard-Vité (Éd.). *Entre la régulation des apprentissages et le pilotage des systèmes: évaluations en tension*. Actes du 20e colloque de l'ADMEE-Europe, Université de Genève. Accès : <https://plone.unige.ch/sites/admee08/tables-rondes/v-tr2/v-tr2-3>
- Michel, G. & Allemand, D. (2008). L'apprentissage de la conduite d'entretien par les formateurs. *Travail et Apprentissages*, 2, 111-130.
- Pastré, P. (2001). Les compétences professionnelles et leur développement. *Revue Cfdt*, 39, 3-10.
- Pastré, P. (2007). Quelques réflexions sur l'organisation de l'activité enseignante. *Recherche et formation*, 56, 91-93.
- Pastré, P. (2008a) Apprentissage et activité. In Y. Lenoir & P. Pastré (Ed.), *Didactique professionnelle et didactiques disciplinaires en débat* (pp.53-79). Toulouse : Octarès.
- Pastré, P. (2008b). Quelles questions à propos de l'analyse de l'activité ? Conférence lors de la rencontre annuelle de l'association Recherche et Pratiques en Didactique Professionnelle, Paris, le 31 janvier. Accès : <http://didactiqueprofessionnelle.ning.com/page/reunion-annuelle-2008>
- Pastré, P. (2008c). Etude comparative des trois chantiers de l'Afpa. *Travail et apprentissages*, 2, 131-139.
- Perrenoud, P. (2010, octobre). Famille de situations et référentiel de compétences. Diaporama de la conférence à l'Institut de Formation et de Recherche en Enseignement Supérieur, 6 octobre, Université de Liège.

Piaget, J. (1974a). *La prise de conscience*. Paris : Presses Universitaires de France.

Piaget, J. (1974b). *Réussir comprendre*. Paris : Presses Universitaires de France.

Rey, B. (1996) *Les compétences transversales en question*. Paris : ESF.

Ropé, F. & Tanguy (1994). *Savoirs et compétences. De l'usage de ces notions dans l'école et l'entreprise*. Paris : L'Harmattan.

Ropé, F. (1999). La validation des acquis professionnels : entre expérience, compétences et diplômes. In J. Dolz, & E. Ollagnier (Ed.), *L'énigme de la compétence en éducation* (pp. 203-226). Bruxelles : Boeck.

Vergnaud, G. (2001). Forme opératoire et forme prédicative de la connaissance. In J. Portugais (Ed.), *La notion de compétence en enseignement des mathématiques, analyse didactique des effets de son introduction sur les pratiques et sur la formation. Actes du colloque GDM*. Accès : <http://smf4.emath.fr/Enseignement/TribuneLibre/EnseignementPrimaire/ConfMontrealmai2001.pdf>

Vermersch, P. (1994). *L'entretien d'explicitation*. Paris : ESF