

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

Annick Fagnant * et Christophe Dierendonck**

* Université de Liège (Belgique) – afagnant@ulg.be

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

Mots-clés : *Tâche complexe – Evaluation diagnostique - Mathématiques – Enquête à large échelle*

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 ? ». En vue de concrétiser ces questionnements, le domaine des mathématiques sert de fil rouge à notre réflexion.*

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

La notion de « tâche complexe » (ou de situation complexe) est intimement liée au concept de compétence. Cependant, ce concept de compétence demeure un concept flou, régulièrement critiqué pour sa polysémie et ses fondements théoriques apparemment fragiles. Il est dès lors très difficile de définir un concept par un autre qui s'avère mal assuré.

Dans la multitude de définitions qui existent autour de la notion de compétence (Beckers, 2005 ; Carette, 2007 ; Crahay, 2006 ; De Ketele & Gérard, 2005 ; Rey, Carette, Defrance & Kahn, 2003 ; Scallon, 2007 ; Tardif, 2004), on parvient cependant à circonscrire certains éléments communs qui définissent comme compétents les individus capables de mobiliser, de façon intégrée, différentes ressources pour accomplir une tâche. Cela dit, - et c'est peut-être une des raisons du flou conceptuel qui existe autour de la notion - rares sont les auteurs ayant clarifié les nuances sémantiques qui découlent du déterminant défini ou indéfini (la, les, une, des) associé dans le discours au terme « compétence ». L'expression « évaluer la compétence » peut en effet être comprise comme le fait d'évaluer une unité curriculaire ou disciplinaire particulière (par exemple, la compétence « traiter de l'information » au sein d'une liste de compétences en lecture) ou comme le fait de placer un individu face à une tâche précise et d'apprécier ses comportements et la réponse qu'il donne face à la situation. Il y aurait donc d'un côté les compétences définies dans le curriculum et de l'autre côté la compétence démontrée effectivement par les individus dans les situations d'évaluation.

Dans la suite du texte, nous allons discuter brièvement ce qui constitue le cœur de la première question guide, à savoir, la notion de « complexité ». Dans un second temps, nous nous intéresserons spécifiquement aux compétences mathématiques et envisagerons la question de l'évaluation des compétences mathématiques dans l'étude 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   valuation 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 habilet  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>

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'étendant vers la phase 3 de l'épreuve, soulignons encore deux « lacunes diagnostiques potentielles » des épreuves en phases : (a) Quelle est la nature exacte des difficultés des élèves qui parviennent à résoudre le problème en phase 2 et non en phase 1 (pourquoi savent-ils appliquer, mobiliser, mais non « mobiliser et intégrer ») ? (b) Quelle est la nature exacte des difficultés des élèves qui parviennent à résoudre les items en phase 3 mais qui ne parviennent pas à résoudre le problème 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 à la non-maîtrise de procédures ou de savoir-faire de base, elles ne permettent généralement pas à elles seules d'expliquer pourquoi autant d'élèves continuent à échouer en phase 2.

Qu'en est-il en outre de ces mystérieux processus que sont le « savoir mobiliser » et le « savoir mobiliser et intégrer » qui sont pointés au cœur de toutes les définitions de la notion de compétence ? Pour éclairer ce questionnement, il paraît essentiel de se détacher d'une analyse « décontextualisée » des résultats que l'on peut tirer d'une évaluation papier-crayon pour chercher à mieux cerner les processus cognitifs réellement à l'œuvre dans les tâches complexes. Pour ce faire, il semblerait intéressant de combiner des « laboratoires cognitifs » où l'on observe un élève face à une tâche complexe (tout en l'interrogeant sur ses démarches), des observations centrées sur l'analyse des interactions entre élèves lorsqu'ils résolvent une tâche complexe en groupe et, enfin, des observations centrées sur le rôle d'étayage qu'un enseignant met en œuvre (ou pourrait mettre en œuvre) pour interagir efficacement avec les élèves.

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

En ce qui concerne les qualités psychométriques des épreuves, il est courant de distinguer une validité de « *construct* » (ou validité théorique) et une validité de « contenu » (ou représentativité) : l'épreuve mesure-t-elle bien ce qu'elle prétend mesurer et couvre-t-elle les aspects importants des apprentissages ? En termes de « validité de contenu », on peut s'interroger sur les « limites » d'une évaluation de compétences qui serait composée uniquement de quelques tâches complexes : d'une part, on est loin d'atteindre une représentativité des contenus enseignés et, d'autre part, on peut s'interroger sur le respect d'une certaine concordance entre ce qui est enseigné en classe et ce qui est évalué (ce qui serait une autre façon de rendre une évaluation « non valide » ; Beckers, 2005).

Concernant les liens entre enseignement, apprentissage et évaluation, les résultats particulièrement faibles obtenus dans la plupart des études évaluant des tâches complexes (Bouché, 2011, Crahay & Detheux, 2005 ; Fagnant & Dierendonck, 2010b ; Marcoux, à paraître ; Rey et al., 2003) incitent à un double questionnement : d'une part, les enseignants proposent-ils des tâches complexes à leurs élèves en situation d'apprentissage et d'autre part, s'ils proposent de telles tâches, comment procèdent-ils d'un point de vue méthodologique ? S'interrogeant sur les faibles taux de réussite obtenus face aux tâches complexes de mathématiques soumises à des élèves de 5^e année primaire à Genève, Marcoux (à paraître) pointe l'écart entre les intentions méthodologiques des programmes et les situations d'évaluation : « Dans la lignée d'une forme de socioconstructivisme, l'apprentissage des mathématiques à Genève se conçoit comme une construction en interaction avec les autres (...) L'élève genevois se voit principalement confronté à la découverte d'apprentissages et à la résolution de problèmes quasi exclusivement en groupe. Il se peut que, dans ce cas, cette façon de faire soit plus productive qu'une transmission frontale du savoir, mais elle peut risquer aussi, par son caractère de recours aux ressources externes (ici l'enseignant ou les pairs), de devenir un obstacle dans la résolution de tâches individuelles » (p. 15).

Dès les premières lignes de ce texte, nous avons évoqué la question de la mobilisation conjointe de ressources internes et de ressources externes. Ces ressources externes, que nous avons dans un premier temps « évacuées » pour débattre de la complexité des tâches, jouent pourtant un rôle

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.