

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 à la complexité (relativement mal définie par ailleurs) qui conduit à ce rejet systématique des épreuves « classiques ». En définitive, l'essentiel n'est pas de savoir dans quelle mesure chaque tâche isolée d'une épreuve externe comme PISA est ou non suffisamment complexe pour être considérée comme permettant ou non d'évaluer quelque chose de l'ordre de la compétence. L'intérêt de ce type d'approche réside plutôt dans la possibilité de dire qu'un élève donné a atteint un certain « niveau de compétence » correspondant à sa capacité à maîtriser un ensemble de tâches mobilisant des processus cognitifs similaires ou semblables. En ce sens, les épreuves externes à large échelle d'évaluation des acquis des élèves comme PISA permettent d'ouvrir la possibilité d'une autre conceptualisation de la complexité, non plus focalisée directement sur la tâche, mais envisageant plutôt la complexité que traduit le niveau de compétence atteint par les élèves.

2. 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 ?

L'intérêt du modèle de Rey et al. (2003) réside dans le potentiel diagnostique du phasage qu'ils proposent, en lien avec les trois niveaux de compétence qu'ils ont identifiés : en phase 1, les élèves sont confrontés à une tâche complexe exigeant le choix et la combinaison d'un nombre significatif de procédures ; en phase 2, on leur propose la même tâche complexe, mais cette fois découpée en tâches élémentaires présentées dans l'ordre ou elles doivent être accomplies ; en phase 3, on leur soumet une série de tâches simples décontextualisées correspondant aux procédures élémentaires qui ont dû être mobilisées pour accomplir la tâche complexe (Carette, 2007). Les données présentées par Rey et al. (2003) fournissent une preuve empirique qu'il existe une certaine hiérarchie entre les trois phases de leur modèle d'évaluation (la phase 3 est mieux réussie que la phase 2 qui est elle-même mieux réussie que la phase 1) mais rien ne permet cependant d'affirmer qu'il s'agit d'une hiérarchie stricte (Dierendonck & Fagnant, 2010a).

Nous allons questionner ce modèle 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 données obtenues par Rey et al. (2003) montrent assurément que les tâches complexes se révèlent très difficiles pour les élèves. Dans ces épreuves, la tâche complexe n'est pas corrigée de façon dichotomique, mais à l'aide d'une série de critères spécifiques qu'il est possible de sommer pour obtenir une note globale (ainsi, on peut considérer qu'un élève qui obtient une note de 80% à la tâche complexe a rempli correctement une proportion de 8 critères sur 10). Pour les deux tâches proposées en mathématiques dans l'enseignement primaire, les résultats montrent que seuls 3 à 5% des élèves atteignent un seuil de réussite de 80% en phase 1 (et ils ne sont encore que 21 à 46% à réussir l'une ou l'autre de ces tâches avec un seuil de 50%).

Dans le cadre d'un mémoire réalisé en 2011 en Sciences de l'Éducation à l'Université de Liège, deux tâches complexes, proposées par la commission des outils d'évaluation sur la base du modèle en phases et impliquant des équations algébriques, ont été soumises à huit classes d'élèves de deuxième année de l'enseignement secondaire (154 élèves) durant les mois de février et de mars (Bouché, 2011)¹. En phase 1, une des tâches complexes est réussie par 11% des élèves (Cadeaux) et l'autre par 4% d'entre eux (Muriel 6). S'interrogeant sur le pouvoir diagnostique potentiel de cette phase, Bouché (2011) note tout d'abord que 46% des élèves rendent une feuille blanche pour

¹Il s'agit des tâches complexes « Cadeaux » et « Muriel 6 » disponibles sur le site <http://www.enseignement.be>

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

la première tâche et 10% pour la seconde. Par ailleurs, les critères proposés pour analyser cette première phase donnent relativement peu d'informations pour la première tâche (*si 37% des élèves identifient correctement l'inconnue à poser – dont les 11% qui ont pu résoudre la tâche correctement - tous les autres critères indiquent que seuls les élèves qui ont réussi l'ensemble de la tâche ont pu les remplir*), mais qu'ils en fournissent davantage pour la seconde tâche (*64% des élèves identifient correctement l'inconnue, 48% parvient à mettre en relation les divers éléments de l'énoncé et 16% présentent correctement les étapes de la résolution du problème*). Finalement, dans la première tâche, la conjugaison du taux important d'omissions (46%) et le peu d'informations fournies par les élèves qui n'ont pas pu résoudre la tâche (*26% des élèves remplissent 1 critère sur 10 et 17% n'en remplissent aucun*) incite à penser que le pouvoir diagnostique de la première phase d'évaluation est clairement limité.

Dans le cadre de l'évaluation externe menée au Luxembourg en octobre 2009, nous avons soumis quelques tâches complexes aux élèves de 3^e année secondaire (filiale générale). Ces tâches peuvent être considérées comme « complexes » au sens retenu au terme de la première question guide. A titre illustratif, le problème intitulé « Judo » est un exemple de tâche complexe qui nécessite de mobiliser deux types de procédures (*calculer la longueur d'un côté d'un carré à partir de son aire et déduire la longueur d'un segment au départ de deux longueurs*), de « cadrer » la situation pour déterminer les procédures à mobiliser et de coordonner plusieurs étapes de résolution (*représentation du problème, identification de l'inconnue, calcul de la longueur d'un côté de la salle, calcul de la longueur d'un côté du tatami et déduction de la largeur du chemin au départ de ces deux longueurs*). Cette tâche complexe, soumise à 1769 élèves, n'est réussie que par 14% des élèves (Dierendonck & Fagnant, 2010b). Bien que seule la réponse finale ait été prise en compte, l'analyse des réponses permet d'inférer les démarches erronées les plus courantes des élèves. A titre illustratif, 25% des élèves obtiennent une réponse témoignant d'une erreur se situant à la toute dernière étape de résolution (*la déduction de la largeur du chemin nécessite de soustraire les deux longueurs trouvées précédemment, puis de diviser le résultat par deux*), 11% des élèves obtiennent une réponse qui témoigne d'une démarche assez superficielle (*la soustraction des deux aires fournies conduit à trouver l'aire du chemin et non la largeur de celui-ci*) et seuls 6% des élèves ne proposent aucune réponse. Au final, cette tâche de complexité modérée, même corrigée de façon binaire, semble offrir un certain pouvoir diagnostique dès l'étape 1.

En conclusion de ce premier point, nous retiendrons deux éléments : (a) les tâches trop complexes perdent leur pouvoir diagnostique si elles déstabilisent les élèves au point qu'ils rendent une feuille blanche ou ne proposent que des bribes de réponses; (b) des tâches sollicitant une réponse numérique précise peuvent à la fois être des tâches complexes permettant d'être corrigées de façon binaire et offrir un certain pouvoir diagnostique dès 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 modèle de Rey et al. (2003) présente un potentiel diagnostique à un double niveau : (a) un élève qui réussit à résoudre un problème en phase 2 alors qu'il n'y parvenait pas en phase 1 maîtrise les ressources impliquées dans la tâche complexe, semble capable de les mobiliser dans des tâches simples (décomposées) mais éprouve des difficultés à organiser sa démarche de résolution et/ou à intégrer les différentes ressources ; (b) pour les élèves qui ne parviennent pas à résoudre le problème en phase 2, une analyse plus précise des étapes de cette phase 2 devrait permettre de cerner où se situent les difficultés. Tout en reconnaissant l'intérêt potentiel de cette étape de décomposition, nous pointerons ici quelques questionnements qu'elle soulève à nos yeux.

Tout d'abord, on peut craindre qu'un élève qui échoue lors de la phase 1 échoue aussi lors de la phase 2 parce qu'il est déconcerté par la situation proposée ou, tout simplement, parce qu'il se sent découragé. A titre illustratif, dans l'étude de Bouché (2011), le passage de la phase 1 à la phase 2 du problème « cadeaux » ne permet qu'une très faible augmentation du taux de réussite (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.