

L'évaluation des compétences à travers des situations complexes¹

François-Marie GERARD, BIEF

1. Nécessité et difficultés des épreuves par situations complexes

L'approche par les compétences, aujourd'hui largement promue dans de nombreux systèmes éducatifs, vise à rendre les élèves capables de mobiliser leurs savoirs et savoir-faire pour résoudre des situations-problèmes (LE BOTERF, 1994 ; REY, 1996 ; CRAHAY, 1997 ; PERRENOUD, 1997 ; FOUREZ, 1999 ; DE KETELE, 2000, 2001 b ; ROEGIER, 2000, 2003 ; TILMAN, 2000 ; LEGENDRE, 2001 ; DOLZ & OLLAGNIER, 2002 ; JONNAERT, 2002).

Dans cette perspective, l'évaluation des acquis des élèves prend une nouvelle dimension : l'élaboration des outils d'évaluation ne peut plus se limiter à prélever un échantillon de contenus et/ou d'objectifs spécifiques et opérationnels représentatif de l'univers de référence en termes de contenus ou d'objectifs, mais propose une ou des situations complexes, appartenant à la famille de situations définie par la compétence, qui nécessitera (ont) de la part de l'élève une production elle-même complexe pour résoudre la situation. (DE KETELE, 2001 a ; ROEGIER, 2004 ; REY, CARETTE, DEFRANCE & KAHN, 2003 ; SCALLON, 2004 ; DE KETELE & GERARD, 2004 ; GERARD, 2005).

L'élaboration, l'utilisation et la validation de tels outils ne sont pas sans difficulté (DE KETELE & GERARD, 2004) :

- la théorie classique des scores ou les théories de réponse à l'item sont fondées en grande partie sur le postulat d'une distribution normale des résultats et sur l'unidimensionnalité du trait mesuré, ce qui est peu compatible avec une évaluation par situations complexes qui inévitablement exigeront de la part des élèves une production elle-même complexe et flexible, non réductible à une seule et unique démarche ni à un seul et unique produit ;
- l'équivalence des situations, leur appartenance à une même famille de situations, n'est évidente ni au moment de leur construction ni pour leur validation. Encore faudrait-il que les auteurs s'entendent sur le concept de « famille de situations », ce qui est loin d'être le cas. CRAHAY et DETHEUX (2005) relèvent qu'une conception serait de considérer « qu'une famille de situations correspond à un ensemble de tâches ou de problèmes qui partagent en commun le fait d'être résolus par une (ou un ensemble spécifié de) procédure(s) spécifique(s) et ceci, quelles que soient les caractéristiques de surface ». Cette conception se fonderait sur la relation *famille de*

¹ Référence : GERARD, F.-M. (2005), L'évaluation des compétences par des situations complexes, *Actes du Colloque de l'Admee-Europe*, IUFM Champagne-Ardenne, Reims, 24-26 octobre 2005.

situations X → *procédure de type X*, mais – comme les auteurs le soulignent – « face à certains problèmes, il n'existe pas d'algorithmes de résolution ». C'est justement pour cela que la situation est complexe et nécessite une véritable compétence pour pouvoir la résoudre, celle d'analyser la situation, d'identifier et de mobiliser des ressources pertinentes – qui peuvent être différentes selon les élèves – et de les articuler entre elles de manière inédite, ponctuelle et éphémère et souvent propre à chaque individu ;

- les exigences de validité et de fiabilité des épreuves d'évaluation nécessitent de prendre en compte la représentativité de l'échantillon des questions par rapport à l'univers de référence : représentativité en termes de nombres d'items, mais aussi en termes de recouvrement qualitatif des différentes strates contenues dans l'univers de référence. Dans le cas de situations complexes, cette représentativité est quasi impossible à respecter dans la mesure où on ne peut décemment proposer aux élèves qu'une ou deux situations complexes à résoudre. C'est pourtant sur la base de ces quelques situations qu'il faudra inférer la maîtrise de la compétence, c'est-à-dire considérer si l'élève est capable ou non de résoudre toutes les situations de la famille de situations ;
- enfin, si l'évaluation par situations complexes est la seule pertinente dans le cadre d'un système éducatif s'étant inscrit dans la perspective de l'approche par les compétences de base (ROEGIERS, 2000), elle présente certaines difficultés en termes d'acceptabilité sociale, tant pour les élèves et les enseignants que pour les parents. Il n'est pas facile de passer d'une culture du *par cœur* ou de l'application mécanique à celle de la résolution de problèmes, d'autant plus que celle-ci entraîne d'autres manières de corriger et d'autres manières de communiquer l'information, et qu'elle n'offre pas le même caractère de légitimité que les épreuves classiques où la réponse pseudo-objective conduit trop souvent à la seule sélection.

Ces difficultés réelles ne doivent cependant pas empêcher la réalisation et l'utilisation de tels outils d'évaluation des acquis des élèves. En effet, la qualité première d'un outil d'évaluation doit être sa pertinence par rapport au système dans lequel il s'insère. À quoi servirait-il de déclarer vouloir développer des compétences si l'évaluation des acquis, qu'elle soit formative ou certificative, ne se fonde que sur la maîtrise de savoir-reproduire et/ou de savoir-faire isolés, aussi performants soient les outils utilisés ?

Inversement, il serait absurde de vouloir évaluer les acquis scolaires par l'intermédiaire de situations complexes si la résolution de problèmes n'a pas fait préalablement l'objet d'un apprentissage systématique, que celui-ci soit permanent ou ponctuel. Cette remarque est particulièrement importante : tout dispositif d'évaluation de compétences par des situations complexes devrait d'abord s'assurer que les élèves ont réellement appris à résoudre des situations appartenant à la famille de situations visée. On peut légitimement penser que la plupart des enseignants ont toujours souhaité que leurs élèves soient compétents en termes de résolution de problèmes. Cependant, peut-on affirmer que ces mêmes enseignants ont mis en œuvre un processus d'enseignement-apprentissage tourné spécifiquement vers la résolution de problèmes ? Il semblerait plutôt que les enseignements se concentrent sur les savoirs et les savoir-faire, sur les connaissances déclaratives et procédurales.

Trop rares sans doute sont les moments où les élèves sont confrontés – dans le cadre d'un apprentissage, avec toute la régulation que celui-ci nécessite – à des situations complexes à propos desquelles est réalisé en classe un travail systématique d'analyse de la situation, d'émission d'hypothèses de résolution, d'identification des ressources pertinentes, de mobilisation et de mise en œuvre de celles-ci pour résoudre la situation. Tout se passe comme si les enseignants considéraient qu'à partir du seul apprentissage des ressources, les élèves pouvaient apprendre d'eux-mêmes à identifier celles qui sont pertinentes dans une situation, à les mobiliser et à les intégrer dans une démarche complexe, alors qu'on sait que c'est bien ce fait-là qui est le principal agent de la « reproduction » : certains savent hors école comment faire et l'École ne l'explique pas aux autres (BOURDIEU & PASSERON, 1970).

L'utilisation d'outils d'évaluation fondés sur des situations complexes ne conduit alors qu'à constater que les élèves ne savent pas résoudre celles-ci. Loin de pouvoir en inférer que les élèves ne sont pas compétents, qu'il leur est impossible de résoudre des situations complexes, il faudrait surtout se contenter de constater que les élèves ne savent pas réaliser ce qu'ils n'ont pas appris ! L'évaluation des compétences par des situations complexes ne peut avoir lieu que si un apprentissage systématique de ce type de résolution est préalablement réalisé. En bonne logique, il faudrait d'ailleurs que les situations proposées lors de l'évaluation soient plus simples, plus faciles que les situations proposées lors de l'apprentissage.

2. Une méthodologie pour l'élaboration des épreuves

Sollicité par les grands bailleurs de fonds en charge de l'éducation, qu'ils soient multilatéraux (Banque Mondiale, UNESCO, UNICEF...) et/ou nationaux (USAID, ADF, CTB...), le BIEF accompagne en ce sens des projets de réécriture des curriculums et d'élaboration de systèmes d'évaluation scolaire dans une quinzaine de pays, dont notamment la Mauritanie, Djibouti, Madagascar, le Liban, l'Algérie, le Bénin..., depuis la première année de l'enseignement de base jusqu'à la dernière année du secondaire, mais aussi au niveau de la formation initiale et continue des enseignants, des directeurs et des inspecteurs.

Si la contextualisation des outils est un souci premier, la démarche méthodologique proposée par le BIEF est globalement la même quel que soit le terrain de réalisation :

- des groupes d'enseignants et/ou d'inspecteurs, désignés par le Ministère de l'Éducation, définissent les compétences de base qui sont visées à la fin d'un cycle ou d'une année, à raison de 3 ou 4 compétences par discipline et d'un objectif terminal d'intégration (DE KETELE, 1984) ;

Exemple de compétence de base, en Chimie, pour la 9^e année de l'enseignement de base – Liban

Face à une situation problème, nouvelle et significative pour l'élève, liée à l'utilisation des réactifs et des produits chimiques et leur impact sur l'environnement, l'élève sera capable d'identifier le problème et de proposer une solution appropriée à ce problème tout en faisant appel à ses connaissances concernant : l'atome, la classification, les liaisons chimiques, l'électrochimie, la

chimie organique et chimie et environnement.

Exemple de compétence de base, en Français oral (2^e langue d'enseignement), pour la 6^e année de l'enseignement de base – Mauritanie

Produire dans une situation significative pour l'élève un récit oral cohérent, respectant la suite des actions, intégrant la description de certains éléments du contexte et contenant l'explication d'au moins un fait.

- ils associent à chaque compétence les ressources (savoirs et savoir-faire) qui peuvent être mobilisées lors de sa mise en œuvre. Ces ressources correspondent le plus souvent aux objectifs qui figurent dans les programmes fondés sur la pédagogie par objectifs, mais ils sont réorganisés en fonction des compétences. C'est donc l'occasion de réaliser un *élagage* important : tous les objectifs qui ne sont pas directement reliés à une compétence disparaissent du curriculum ;
- ils élaborent 2 ou 3 situations qui illustrent la famille de situations susceptibles d'être résolues par l'élève maîtrisant la compétence. C'est une démarche relativement pragmatique : les concepteurs se demandent quelles situations concrètes permettraient de certifier la maîtrise des compétences visées. Ce n'est que sur la base de ces situations concrètes qu'ils dégagent les paramètres de la famille des situations, c'est-à-dire les éléments communs à toutes les situations de la famille. Souvent même, ils reviennent sur la formulation des compétences qui s'expriment alors seulement en termes de résolution de problèmes. Ces paramètres concernent l'univers de référence en termes de ressources, le type de situations, le type et le nombre de supports, le type de tâche attendue, les conditions de résolution, le type de critères utilisés pour évaluer la production... Les paramètres de la famille de situations ne se limitent donc pas aux connaissances ou aux procédures à mobiliser, mais reprennent toute une série d'éléments qui devraient permettre à d'autres concepteurs d'élaborer de nouvelles situations, considérées comme équivalentes ;

Exemple (incomplet) de paramètres de la famille de situations de la compétence de base, en Chimie, pour la 9^e année de l'enseignement de base – Liban

Univers de référence en termes de ressources à mobiliser (savoirs, savoir-faire, savoir-être) :

Savoirs : constituants de l'atome, classification des éléments, réactions chimiques, formules et équations, solutions, électrochimie, hydrocarbures, polymérisation et polymères, pollution ...

Savoir-faire : Exploiter les différents types de supports, trier les informations, les traiter ; analyser ; déduire, communiquer par écrit (et/ou oral), tracer des courbes, des schémas et des diagrammes..., réaliser des activités expérimentales en utilisant un protocole

Savoir-être : se comporter positivement envers l'environnement, avoir un sens critique, être autonome.

Type de situations :

Situation-problème liée à la vie quotidienne, environnementale (urbaine et/ou rurale) technologique ou industrielle et au travail au laboratoire, pollution, dégradation des sols, alimentation, polymères, etc., faisant intervenir l'exploitation d'un ou de plusieurs supports nécessitant d'identifier un problème et de le résoudre.

Type et nombre de supports :

Texte, photo, schéma et tableau. Nombre : entre 2 et 4 ...

Type de tâche attendue :

Identifier un problème, exploiter un document, un graphe..., traiter des résultats expérimentaux, et résoudre le problème.

Conditions de résolution :

Temps : 30 minutes

Travail individuel en classe ou au laboratoire

Critères utilisés pour évaluer la production :

I- Critères minimaux² :

1. *Adéquation de la production de l'élève à la situation*
2. *Utilisation correcte des outils de la discipline*
3. *Cohérence de la production de l'élève.*

II- Critère de perfectionnement :

4. *Qualité de la présentation de la production de l'élève.*

- sur la base de ces paramètres, les épreuves d'évaluation sont élaborées en veillant d'une part à ce que les situations respectent les paramètres de la famille de situations et puissent donc être considérées comme équivalentes les unes aux autres et d'autre part à ce que la règle dite des 2/3 soit respectée (DE KETELE, 1996 ; ROEGIERS, 2000). Cette règle consiste – au moment de l'élaboration de l'épreuve – à s'assurer que l'élève aura au moins 3 occasions indépendantes de vérifier chaque critère, et – au moment de la correction – à considérer qu'un critère est maîtrisé si au moins 2 occasions sur 3 sont positives. Les deux dimensions de cette règle – au moins 3 occasions présentes et au moins 2 occasions réussies – contribuent à éviter les échecs abusifs, car elles permettent d'éviter de mettre en échec un élève sur la base d'une seule erreur ;
- les critères d'évaluation sont détaillés pour chaque situation en indicateurs sur la base desquels un barème de notation est élaboré consistant avant tout à déterminer quand le critère sera ou non considéré comme maîtrisé. Cette phase est particulièrement importante et présente plusieurs difficultés auxquelles nous reviendrons plus loin ;
- chaque situation d'évaluation est préexpérimentée auprès de quelques élèves, essentiellement à des fins de régulation, et ensuite expérimentée dans un échantillon représentatif afin de permettre sa validation empirique ;
- à côté de ces épreuves « compétences » fondées sur des situations complexes, des épreuves « ressources » semblables aux épreuves classiques sont également élaborées non seulement pour accroître l'acceptabilité de l'évaluation, mais surtout pour en accroître le sens : si un élève ne parvient pas à résoudre une situation complexe, il est important de savoir si cela est lié au fait qu'il ne maîtrise pas les ressources à mobiliser ou qu'il ne parvient pas à les mobiliser ni à les intégrer pour résoudre la situation-problème³.

² Les *critères minimaux* sont ceux qui doivent absolument être maîtrisés pour certifier la maîtrise de la compétence alors que les *critères de perfectionnement* concernent des qualités dont la présence est préférable, mais non indispensable. En termes de certification, ce n'est que sur la base des critères minimaux que la réussite est décidée ou non.

³ Ces épreuves « ressources » contribuent ainsi au diagnostic de l'évaluation des compétences. Il y a plusieurs approches en ce qui concerne l'évaluation des compétences : certains la limitent à des épreuves par situations complexes, d'autres présentent d'abord des situations complexes et ensuite des épreuves visant à évaluer la maîtrise des procédures particulières ((REY, CARETTE, DEFRANCE & KAHN, 2003 ; CRAHAY & DETHEUX, 2004), alors que d'autres évaluent d'abord la maîtrise des ressources avant de s'occuper des compétences par situations complexes (ALLAL, ROUILLER, SAADA-ROBERT & WEGMULLER, 1999). La position du BIEF est qu'il est nécessaire d'évaluer les compétences et les ressources, sans se prononcer sur l'opportunité d'évaluer les unes avant les autres. Il est d'autre part évident que les ressources peuvent et doivent également être évaluées en tant que

Durant toute la mise en œuvre de ce dispositif, de fréquents allers-retours entre les équipes de concepteurs et le terrain sont indispensables, tout en prenant en compte que les compétences ne sont souvent pas encore travaillées en tant que telles dans les classes. L'évaluation porte donc – durant la phase d'élaboration des outils – sur des compétences qui n'ont pas toujours été développées : si les ressources ont, la plupart du temps, été travaillées par les enseignants des classes concernées, ce n'est pas le cas de leur mobilisation ni de leur intégration dans la résolution de problèmes. C'est donc avec prudence qu'il faut traiter et analyser les résultats et procédures de validation.

3. Quelques étapes-clés du processus d'élaboration et de validation

3.1. La définition des paramètres de la famille de situations

Ce travail est essentiel parce que c'est lui qui permet de délimiter de manière concrète et précise quelle est la famille de situations concernée par la compétence. De sa qualité dépendra fortement la validité de l'inférence qui sera réalisée à partir d'une ou deux situations d'évaluation sur la maîtrise de la compétence. Si la famille de situations est délimitée de manière très vague, elle ne permettra pas d'élaborer des situations suffisamment équivalentes et la seule conclusion acceptable au terme de l'évaluation serait que les élèves sont capables de résoudre ou non la situation particulière sur laquelle ils ont travaillé, sans pouvoir élargir le champ d'inférence.

Le *paramétrage* de la famille des situations n'est pas à confondre avec l'*habillage* des situations. Le paramétrage permet de définir les invariants entre toutes les situations qui devraient pouvoir être résolues dans le champ de la compétence, alors que l'*habillage* d'une situation est ce qui caractérise chaque situation en particulier. Normalement, l'*habillage* ne devrait pas avoir d'influence sur le pouvoir de résolution d'une situation, de la même manière qu'une personne reste elle-même quels que soient ses habits du jour. Il est parfois difficile de savoir si un élément fait partie du paramétrage ou de l'*habillage*. Par exemple, dans les petites classes, on peut légitimement émettre l'hypothèse que les éléments affectifs liés à une situation particulière ont une grande influence sur la capacité de l'élève de s'investir dans la résolution de cette situation et pourraient donc être inclus dans le paramétrage.

La qualité de l'inférence sur la maîtrise de la compétence dépend donc non seulement de la précision des paramètres de la famille de situations qui lui est liée, mais aussi du moment de la phase de formation. Moins on est avancé dans le processus de formation, plus la précision des paramètres devra être grande et détaillée. En 1^{re} année de l'enseignement fondamental, on précisera, en mathématiques, que « les situations problèmes nécessitent une addition de nombres compris entre 1 et 10, sous la forme $a + b = x$ », alors qu'en 6^e année, on se contentera de dire que « les situations problèmes

telles, essentiellement dans une perspective formative, pour permettre leur régulation lors des apprentissages ponctuels.

feront appel à des additions, des soustractions, des multiplications et/ou des divisions ». De la même manière, en formation professionnelle, les paramètres de la famille de situations seront plus ou moins précis selon que l'on vise à évaluer un « compétent novice » ou un « compétent expert ». Pour ce dernier, et pour autant qu'il faille l'évaluer, on se contentera d'une définition très vague de la famille de situations qu'il est appelé à résoudre – du style « situations relatives à l'ingénierie de la formation » – parce que sa compétence d'expert est justement de pouvoir résoudre toutes les situations relatives à son champ d'expertise.

3.2. La définition des critères d'évaluation

Les critères d'évaluation sont un des paramètres essentiels de la famille de situations puisque c'est à travers eux que toutes les situations relatives à la compétence vont être analysées. Un critère est un *point de vue* auquel on se place pour évaluer un objet d'évaluation, une *qualité* attendue de cet objet. Les critères doivent permettre d'évaluer la production de l'élève, c'est-à-dire « d'en faire sortir la valeur » (*e-valuer*). Il est aussi indispensable que ces critères soient indépendants l'un de l'autre afin de ne pas évaluer deux fois la même chose (et donc de ne pas pénaliser deux fois l'élève pour la même erreur).

Par exemple, un élève qui est « hors sujet » ne peut être pénalisé qu'une fois, à travers un critère de « pertinence » (adéquation de la production à la situation). Il faut absolument qu'on puisse évaluer positivement sa production pour d'autres critères – par exemple sa correction orthographique –, même s'il est « hors sujet ». Au bout du compte, cet élève sera vraisemblablement en échec, parce qu'il ne satisfait pas à un critère minimal, mais au moins ce qu'il aura réussi sera reconnu.

Il est sans doute impossible et inutile de proposer une liste de critères qui seraient pertinents et exhaustifs pour toutes les disciplines et tous les niveaux. Néanmoins, il est possible d'identifier — de manière non exhaustive — quelques critères auxquels on peut se référer (voir aussi ROEGIERS, 2004).

Deux critères – ou plutôt « ordres de critères » – semblent devoir s'imposer quels que soient les disciplines, les compétences ou les niveaux évalués :

- la **pertinence**, ou adéquation de la production à la situation (et notamment à la consigne), qui consiste à répondre aux questions « est-ce que l'élève fait bien ce qu'il doit faire ? n'est-il pas *hors sujet* ? » ;
- la **correction** ou utilisation correcte des concepts et des outils de la discipline, qui concerne la question « est-ce que l'élève fait correctement ce qu'il fait ? », même si ce n'est pas cela qu'il doit faire. Ce critère portera ainsi sur l'orthographe correcte (d'usage et grammaticale), l'utilisation correcte des techniques de calculs (même si l'opération concernée ne devait pas être effectuée), l'exactitude scientifique d'une justification, etc.

À côté de ces deux critères de base, on peut identifier deux critères importants, mais dont la pertinence varie en fonction de la discipline concernée, du niveau des élèves, etc., tout en étant le plus souvent considérés comme minimaux s'ils sont retenus :

- la **cohérence**, c'est-à-dire l'utilisation d'une démarche logique, qui ne présente pas de contradictions internes (même si elle n'est pas pertinente), le choix cohérent des outils, l'enchaînement logique de ceux-ci, l'unité de sens de la production, etc.⁴ ;
- la **complétude**, c'est-à-dire le caractère complet de la réponse, pour autant bien entendu qu'il soit possible de déterminer ce qu'est une réponse complète.

Enfin, on peut identifier un certain nombre de critères dont la pertinence variera en fonction des objectifs que l'on poursuit, des valeurs qu'on souhaite privilégier, d'exigences spécifiques, etc. La plupart du temps, ces critères seront considérés comme critères de perfectionnement, c'est-à-dire qui ne doivent pas nécessairement être maîtrisés pour certifier la compétence, mais dont la maîtrise est néanmoins préférable et attendue. Parmi ces critères, on peut identifier, de manière non exhaustive :

- la **qualité de la langue**, pour les disciplines non littéraires ;
- la **qualité de la présentation** du travail, de la copie de l'élève ;
- l'**utilité sociale** ou la **profondeur** des propositions émises, par exemple dans l'analyse de problèmes sociaux, historiques, géopolitiques, d'environnement... ;
- l'**originalité** de la production, spécialement pour les disciplines littéraires ou sociales ;
- ...

Ces critères n'ont bien sûr pas tous toujours le même poids ni la même pertinence. Par exemple, selon les disciplines et/ou selon les niveaux, le critère de « qualité de la présentation » sera

- parfois retenu comme un critère minimal, c'est-à-dire qui doit absolument être maîtrisé pour considérer que la compétence est maîtrisée (par exemple dans la formation d'une secrétaire, lorsqu'elle doit écrire une lettre) ;
- parfois considéré comme un critère de perfectionnement, lorsqu'il s'agit d'une qualité attendue de la part de l'élève, mais sans que cela ne détermine sa réussite ;
- parfois non retenu comme critère du tout, par exemple lorsqu'on privilégie la quantité de production ou la spontanéité.

Le mot qu'on utilise pour désigner le critère n'est pas toujours uniforme : il peut varier notamment en fonction des disciplines. Par exemple, le critère de « **pertinence** » pourra signifier et/ou être appelé :

- en langues, « adéquation de la production avec l'énoncé » (ou encore une autre appellation),
- en mathématiques, « interprétation correcte du problème » ou « choix des bons outils »,
- en sciences sociales, « correspondance avec ce qui est demandé », etc.

Le critère de **correction** (ou utilisation correcte des concepts et des outils de la discipline) est proposé au sens de « qualité de ce qui est correct, conforme aux règles et aux convenances ». Il n'est pas à confondre avec « l'exactitude de la réponse ». Dans l'évaluation traditionnelle ou l'évaluation des ressources, ce critère d'exactitude de la

⁴ Il peut aussi correspondre aux questions : est-ce que les idées sont convenablement organisées ? est-ce que la production est de qualité quant au fond ? est-ce que le travail est bien organisé ? est-ce que la réponse est de bon sens ?...

réponse est quasiment le seul utilisé. Dans l'évaluation des compétences, ce critère – qui ne fonctionne que dans le cas d'une production fermée à réponse unique – est à éviter, car il s'agit d'un critère englobant. Il n'est réellement utilisable que d'une seule manière : pour effectuer un premier tri de copies. Toutes les copies qui ont la réponse exacte au bout d'une production à réponse unique satisfont inévitablement aux critères de pertinence, de cohérence, de correction et de complétude. Dans ce cas, il n'est pas nécessaire de se demander si ces autres critères sont maîtrisés ou non : ils le sont inévitablement et de manière parfaite (pour autant bien sûr que l'élève n'a pas « copié », mais toute évaluation ne fonctionne évidemment que si l'élève n'a pas copié...). Si par contre, l'élève n'a pas au bout du compte la réponse exacte, alors il faut absolument passer en revue tous les autres critères. Et il est possible, selon le barème de correction, que l'élève ait au bout du compte une note maximale, même en n'ayant pas la « réponse exacte »... Le critère « correction » n'est donc pas à comprendre comme « exactitude de la réponse », mais il indique qu'il faut regarder si l'élève utilise correctement les outils de la discipline, des concepts, des techniques, etc.

La **complétude** est un critère qui – s'il est retenu – s'ajoute aux autres et est donc bien indépendant : l'élève peut produire quelque chose de pertinent, de cohérent et de correct, mais ne pas être complet... C'est un peu le cas du mécanicien qui face à un pneu crevé, fait tout ce qu'il faut dans les règles de l'art, mais qui par manque de temps « oublie de resserrer les 4 boulons ». Sa réparation est tout à fait pertinente, cohérente et correcte, sauf qu'il n'a pas été complet... avec des conséquences dramatiques ! Ce critère n'est à retenir que lorsqu'il se justifie vraiment, notamment parce qu'il se situe sur un autre plan que les autres critères : on aborde ici un aspect plutôt quantitatif, alors que les autres critères se situent plutôt sur un plan qualitatif. À nouveau, ce critère de complétude se déclinera différemment selon les disciplines :

- en langues, on se demandera si l'élève a répondu à toutes les questions de la situation de communication, s'il a présenté tous les arguments à envisager...,
- en sciences (sociales ou exactes), s'il a étudié toutes les facettes du phénomène en question...,
- en maths, s'il est arrivé au bout de sa démarche et a répondu à toutes les questions, etc.

La liste des critères découle toujours d'un choix des concepteurs-évaluateurs. Ce choix sera effectué en fonction de leurs objectifs (et de ceux de leur système éducatif), des accents qu'ils souhaitent mettre, en tenant compte bien sûr de la spécificité de chaque discipline et de chaque niveau. En d'autres termes, il n'y a **pas de liste idéale de critères, pas de modèle** qui s'imposerait de manière absolue. La réalité est toujours complexe. Une modélisation vise à introduire un peu de simplicité, mais elle ne peut jamais être au détriment de la complexité de la réalité. D'autres critères que ceux proposés ici peuvent être utilisés, avec d'autres formulations, mais il faut alors toujours s'assurer que le critère est pertinent par rapport à la compétence et vérifier l'indépendance intercritères.

Cette affirmation d'absence de modèle absolu au niveau des critères ne va pas à l'encontre des efforts légitimes d'harmonisation nationale ou internationale des critères. Par exemple, les partenaires libanais du BIEF qui élaborent un nouveau système

d'évaluation fondé sur l'approche par les compétences ont insisté sur leur souhait d'avoir le plus de cohérence possible entre les disciplines et au sein de celles-ci dans le choix des critères, alors que ce projet porte sur toutes les disciplines tout au long des 12 premières années de scolarisation. Dans un contexte international, l'équivalence des diplômes nécessiterait également une harmonisation des critères d'évaluation.

3.3. La définition et la gestion des indicateurs

L'évaluation critériée est complexe, et il ne faut pas la rendre plus complexe qu'elle ne l'est. Tout le monde en est conscient, et tout tourne finalement autour des indicateurs et de ce que l'on décide de faire avec eux.

Si le critère est général et abstrait, l'indicateur est contextualisé et concret. C'est lui qui – dans la production de l'élève – apporte de l'information sur sa maîtrise de la compétence. Il apporte de l'information et, à ce titre, ne donne jamais qu'une indication. Un des dangers, fréquemment observé, est d'accorder à l'indicateur plus de valeur qu'il n'en a.

Dans l'évaluation d'une situation complexe en vue d'inférer la maîtrise d'une compétence, ce que l'on évalue, c'est la maîtrise du critère, mais pas la maîtrise de l'indicateur. Celui-ci n'est qu'un élément parmi d'autres qui permet à l'évaluateur d'estimer, d'apprécier (par un jugement de valeur) le degré de maîtrise du critère.

Dans ce cadre, un indicateur ne devrait pas être un « indicateur minimal », c'est-à-dire quelque chose qui devrait absolument être présent pour décider de la maîtrise du critère. C'est le critère qui est minimal. Pas l'indicateur. Celui-ci ne fait que donner une indication. Certains indicateurs sont bien sûr plus importants que d'autres, mais cela ne devrait jamais rendre un indicateur indispensable pour attester de la réussite d'un critère (et donc de la réussite de la compétence, et donc de la réussite de la discipline, et donc de la réussite du cycle...).

D'autre part, la gestion des indicateurs ne doit pas transformer l'évaluation critériée en évaluation sommative. En d'autres termes, il ne faut pas accorder 1 point à tel indicateur, 0,5 point à tel autre, etc. Non seulement cela risque fort de déboucher sur un système impraticable, impossible à gérer pour les correcteurs, mais plus fondamentalement, cela revient à nouveau à évaluer les indicateurs, et non pas le critère.

La question évidemment est de savoir comment passer de l'identification et de l'analyse des indicateurs à l'appréciation du critère... Il n'y a sans doute pas de bonne et unique solution. Mais des solutions. Il faut néanmoins **éviter de déboucher sur un système infaisable**, parce que trop difficile à gérer par les correcteurs. S'ils doivent tout le temps se référer à leur grille de correction, à leur barème, on peut être sûr que très rapidement, ils ne le feront plus... et finiront même par revenir à une évaluation classique.

L'expérience nous a montré que l'acte de correction d'épreuves constituées de situations complexes est une opération difficile et reste fortement influencé par la personnalité, les valeurs et l'histoire des correcteurs. Lors de la correction d'épreuves classiques constituées d'items, il suffit – dans la plupart des cas – de constater si la réponse est exacte ou non. Au bout du compte, on fait la somme du nombre de points attribués à chaque item réussi pour obtenir la note finale. Le seul critère utilisé, de

manière plus ou moins implicite, est « l'exactitude de la réponse » et la réponse à chaque item peut être évaluée de manière séparée. L'évaluation critériée d'une production complexe nécessite de la part des correcteurs une sorte de révolution culturelle dans la manière de considérer les copies des élèves.

Prenons l'exemple d'une épreuve de français dans laquelle il est demandé aux élèves, après avoir lu un fait divers, de raconter par écrit un événement tragique auquel ils ont assisté afin d'en faire une publication dans le journal du collège. Les productions devraient être évaluées à travers 3 critères : l'adéquation de la production à la situation, la cohérence et la correction orthographique, chacun de ces critères étant noté sur 3 points, ainsi qu'un critère de perfectionnement : la qualité de la présentation. Voici deux copies :

Copie 1

Un accident rackettait à decallare le jeudi 1 mai 03 dernier à 8h30. A l'établissement Abdi.

Une fille (qui) ~~traversait la rue~~ foillée traverse la route, un grand voiture qui veneée derrière elle a cassée. Et elle venée de chez sa soeure et après elle voillée partir chez s'elle.

La fille est sour palasément mort, ses corp est devenu comme la sable et tous la rue est devenu de sens. Après tous les gens sont ramassée le corp de la fille.

Après, quelque seconde, ~~tous~~ ses parents et ses famielle en venus eux qui sont inqueur et ont pleurent. Ils s'étaient ont été tursétée.

Copie 2

Mon père, il est policier et quand il voit des voleurs il les met en prison. Il a un revolver à sa ceinture et parfois il doit s'en servir.

Il fait des patrouilles dans les rues de la ville, deux fois par jour. Il aide les personnes qui ont besoin de lui, mais il donne aussi des contraventions.

Le dimanche, il ne se repose pas toujours. Souvent il doit travailler, et la famille doit aller chez les cousins sans lui. Heureusement, ma mère ne travaille pas et elle peut s'occuper de nous.

Lors de corrections de ces copies, un correcteur a accordé 2/10 aux deux copies. Selon lui, la première copie était quasi illisible et bourrée de fautes d'orthographe. La seconde était « hors sujet ». Même après discussion, il a été impossible de faire admettre à ce correcteur que certes les élèves ne maîtrisaient pas un critère minimal – pour la copie 1, la correction orthographique, et pour la copie 2, l'adéquation de la production à la situation – mais que chacun maîtrisait parfaitement l'autre critère. Au bout du compte, la maîtrise de la compétence n'est effectivement atteinte par aucun de ces deux élèves, puisque tous les deux ne maîtrisent pas un critère minimal, mais l'évaluation critériée devrait pouvoir mettre en avant les qualités de chacun de ces copies. Cet exemple illustre en tous les cas combien il est difficile pour un correcteur du terrain de faire abstraction de ses habitudes de correcteur *classique*. Notamment, on voit bien que le correcteur utilise consciemment deux critères – le *fond* et la *forme*, qui sont connus depuis longtemps par la plupart des correcteurs, spécialement dans les disciplines littéraires –, mais que cette utilisation n'est pas faite de manière cohérente dans la mesure où ce n'est pas toujours le même critère qui prime. La « logique » de ce correcteur est de privilégier ce qui pénalise l'élève, alors que la logique critériée de l'évaluation, surtout en termes de certification, devrait être au contraire de privilégier les réussites des élèves (GERARD, 2004).

Ce travail de définition et de gestion des indicateurs est fondamental, car il permet de réguler les exigences attendues dans le cadre de l'approche par les compétences. Il faut en effet avoir conscience que, si l'approche par les compétences de base poursuit explicitement des objectifs d'équité, elle place la barre très haut en termes pédagogiques :

- d'une part, elle se situe aux niveaux taxonomiques les plus élevés : la mobilisation et la résolution de problèmes sont les deux niveaux supérieurs de la taxonomie des opérations cognitives proposées par D'HAINAUT (1983) ;
- d'autre part, elle établit un seuil de réussite élevé puisque – dans son application stricte – elle conduit à ne considérer une compétence comme maîtrisée que lorsque tous les critères minimaux sont maîtrisés. En bonne logique, la réussite dans une discipline devrait être liée à la maîtrise de toutes les compétences de base⁵ et la réussite d'une année ou d'un cycle scolaire devrait se fonder sur la réussite dans toutes les disciplines...
- enfin, l'expérience nous montre que les concepteurs de situations ont souvent tendance à élaborer des situations plus complexes, voire plus compliquées, que les paramètres de la famille de situations ne le requièrent. L'idée que les situations utilisées dans les évaluations devraient être d'un niveau plus faible que les situations travaillées en classe a parfois du mal à s'imposer !

Il faut donc introduire un peu de souplesse quelque part. Cela peut se faire principalement dans la définition des indicateurs et dans leur gestion définie au sein d'un barème de notation et concrètement utilisée au moment de la correction. Il ne s'agit évidemment pas d'être laxiste, mais de définir des seuils de réussite suffisamment souples – ce n'est pas la perfection qui est recherchée – et d'orienter le processus de correction – surtout au moment de la certification – vers la valorisation de ce qui a été réussi plutôt que vers la sanction de ce qui a été raté.

4. Résultats d'études évaluatives de la maîtrise de compétences par des situations complexes

Dans la plupart des pays engagés dans une réforme d'évaluation des acquis scolaires par des situations complexes, cette innovation s'accompagne ou est précédée par une réécriture des curriculums selon l'approche par les compétences, par une formation des inspecteurs et des enseignants et/ou par une élaboration de manuels scolaires et autres outils didactiques conformes à l'approche. Avant généralisation des nouveaux outils, une phase expérimentale permet de les tester dans une perspective de régulation sur un échantillon représentatif.

Des séries de données ont ainsi pu être récoltées à Djibouti, en Mauritanie et à Madagascar auprès d'un échantillon expérimental d'élèves ayant bénéficié de

⁵ Rappelons que dans le cadre de l'approche par les compétences de base, il n'y a que 3 ou 4 compétences de base qui sont définies pour chaque discipline et chaque niveau.

l'approche par les compétences de base et d'un échantillon témoin équivalent constitué d'élèves poursuivant leur parcours avec les anciens curriculums.

Ce qui différencie le plus les deux échantillons, c'est l'organisation de « modules d'intégration » dans les classes expérimentales. Les apprentissages des ressources, c'est-à-dire les objectifs pédagogiques classiques, sont réalisés de la même manière dans toutes les classes, mais après 5 ou 6 semaines, les élèves des classes expérimentales sont confrontés pendant une semaine à des situations complexes appartenant à la famille de situations de la compétence travaillée. Dans le meilleur des cas, les élèves sont confrontés à 3 situations équivalentes : la première de manière collective et/ou par des travaux de groupes, la deuxième de manière individuelle avec accompagnement de l'enseignant suivi d'une résolution collective, et la troisième de manière individuelle afin de permettre une évaluation formative débouchant sur des remédiations. Ensuite, les apprentissages ponctuels reprennent de manière classique pendant 5 ou 6 nouvelles semaines avant un nouveau travail d'intégration pendant une semaine.

Il n'est ici ni possible ni pertinent de présenter de manière détaillée les résultats issus de ces travaux (ADEN & ROEGIERS, 2003 ; O/ DIDIYE, EL HADJ AMAR, GERARD & ROEGIERS, 2005 ; RAJONHSON, RAMILJAONA, RANDRIANIRINA, RAZAFINDRALAMBO, RAZAFINDRANOVONA, RANOROVOLOLONA & GERARD, 2005) réalisés dans des contextes et des niveaux très différents. Nous nous contentons donc de faire une présentation sur des données conglomérées, en ayant conscience des limites d'une telle présentation. Notamment, les seules comparaisons possibles sont celles au sein d'un même pays à l'exclusion de comparaisons inter pays.

Tableau 1 - Moyennes sur 10 aux épreuves par situations complexes
(APC = Échantillon expérimental, bénéficiant de l'approche par les compétences,
(N-APC = Échantillon témoin, ne bénéficiant pas de l'approche par les compétences)

Pays	Niveaux évalués	Échantillon	Nombre d'élèves	Langue 1 ⁶	Français (L2)	Mathématiques	Moyenne générale
Djibouti	2 ^e AF	APC	161	/	5,13	4,51	4,82
		N-APC	174	/	3,76	3,09	3,43
Mauritanie	6 ^e AF	APC	133	4,38	3,74	4,97	4,36
		N-APC	104	3,95	2,51	2,81	3,09
Madagascar	1 ^{re} AF	APC	676	5,46	/	3,29	4,38
		N-APC	352	4,54	/	2,03	3,29

Une première remarque s'impose eu égard à l'ensemble des données : le niveau des notes est relativement bas puisque toutes les moyennes générales sont inférieures au seuil de 50%. D'une part, il faut relever que ces trois pays sont parmi les plus défavorisés de toute la planète : à l'Indicateur de développement humain⁷ (PNUD, 2004), sur 177 pays référencés, Madagascar se place à la 150^e place, la Mauritanie à la

⁶ Mauritanie : langue arabe ; Madagascar : langue malgache.

⁷ L'Indicateur de développement humain, publié par le PNUD chaque année, prend en compte l'espérance de vie à la naissance, le taux d'alphabétisation des adultes, le taux brut de scolarisation combiné (du primaire au supérieur) et le PIB par habitant.

152^e et Djibouti à la 154^e. D'autre part, tant à Madagascar qu'en Mauritanie, des épreuves « ressources » ont également été appliquées permettant d'évaluer les objectifs spécifiques de manière séparée⁸ : les résultats obtenus par les élèves à ces épreuves sont tout aussi faibles, tant pour les échantillons expérimentaux « APC » que pour les échantillons témoins « Non-APC »⁹. Ces faibles résultats se retrouvent d'ailleurs dans les études internationales et doivent – pour mieux les comprendre – être mis en relation avec les tailles des classes de ces pays, où il n'est pas rare de rencontrer des classes de 100, voire de 200 élèves !

En termes d'efficacité, ces résultats convergent vers le même constat : la mise en œuvre de modules d'intégration permettant un travail systématique sur des situations complexes permet de gagner environ 1,5 point sur 10 – soit 15% – dans des épreuves d'évaluation des compétences constituées de situations complexes. Toutes les différences observées sont (très) significatives. Dans deux cas – le Français à Djibouti et le Malgache à Madagascar – les scores moyens des échantillons expérimentaux passent au-dessus de la barre des 50%. De manière générale, il faut d'ailleurs constater qu'un gain moyen de 15% présente pour ces pays défavorisés un caractère exceptionnel : en effet, il signifie que de nombreux élèves qui étaient en échec auparavant peuvent réussir grâce à l'approche par les compétences de base, et cela même sur la base d'épreuves de niveau taxonomique supérieur.

5. Conclusion

Pour des pays ayant inscrit leur système éducatif dans la perspective de l'approche par les compétences de base, l'évaluation des acquis scolaires par des situations complexes est la seule pertinente puisqu'elle est la seule à susciter la mobilisation et l'intégration de ressources pour résoudre des problèmes.

Même dans une perspective strictement formative – et nous sommes convaincu que c'est cette évaluation-là qui est la plus importante –, l'utilisation de situations complexes permet d'optimiser le diagnostic des difficultés puisqu'elles permettent de disposer et de traiter des informations non seulement sur la maîtrise ponctuelle d'objectifs spécifiques, mais aussi sur leur mobilisation et leur intégration dans la résolution de problèmes.

L'élaboration, l'utilisation et la correction de telles épreuves présentent cependant un ensemble de difficultés qui nécessitent encore de nombreux travaux pour pouvoir être pleinement résolues.

Les difficultés principales tiennent

⁸ Ce sont des épreuves correspondant à la « phase 3 » des travaux effectués à l'ULB par l'équipe de Bernard Rey (REY, CARETTE, DEFRANCE & KAHN, 2003) permettant d'évaluer ce que ces auteurs appellent des « procédures de base ou compétences élémentaires ».

⁹ Dans ces épreuves « ressources », les élèves des échantillons expérimentaux obtiennent des résultats soit équivalents soit supérieurs à ceux des échantillons témoins. On ne constate donc pas de déperdition au niveau des apprentissages classiques.

- à la définition claire et précise des paramètres de la famille de situations permettant de garantir la meilleure équivalence possible entre les différentes situations complexes, même si cette équivalence ne sera jamais totale ;
- à la définition des critères d'évaluation qui doivent être à la fois pertinents et indépendants ;
- à la définition des indicateurs et de leur mode de gestion pour apprécier les critères dans un système qui soit à la fois pertinent et faisable ;
- au processus de correction des productions des élèves et d'appréciation des critères qui nécessite de la part des correcteurs un détachement par rapport à leurs pratiques habituelles ainsi qu'un travail spécifique pour réduire au mieux l'inévitable désaccord intercorrecteurs.

Ces difficultés réelles n'empêchent pas l'élaboration et l'utilisation d'épreuves par situations complexes. Elles invitent à la vigilance et à la recherche de solutions. Puisse cette intervention y contribuer.

6. Références

- ADEN, H.M. & ROEGIERS, X. (2003), *À quels élèves profite l'approche par les compétences de base ? Étude de cas à Djibouti*. Inédit accessible sur <http://www.bief.be>.
- ALLAL, L., ROUILLER, Y., SAADA-ROBERT, M. & WEGMULLER, E. (1999). « Gestion des connaissances orthographiques en situation de production textuelle », *Revue Française de Pédagogie*, n°126, janvier-février-mars 1999, 53-69.
- BOURDIEU, P. & PASSERON, J.-C. (1970). *La reproduction. Éléments pour une théorie du système d'enseignement*. Paris : Éditions de Minuit.
- CRAHAY, M. (1997). *Une école de qualité pour tous*. Bruxelles : Labor.
- CRAHAY, M. & DETHEUX, M. (2005). « L'évaluation des compétences, une entreprise impossible ? (À propos de la résolution de problèmes mathématiques complexes) », *Mesure et évaluation en Éducation*, Vol.28, n°1, 57-78.
- DE KETELE, J.-M. (1984). « Objectif et évaluation : comparaison et confrontation de quelques terminologies utilisées », *Bulletin de Pédagogie Universitaire*, Kinshasa.
- DE KETELE, J.-M. (1996). « L'évaluation des acquis scolaires : quoi ? pourquoi ? pour quoi ? », *Revue Tunisienne des Sciences de l'Éducation*, 23, 17-36.
- DE KETELE, J.-M. (2000). En guise de synthèse : Convergences autour des compétences, in BOSMAN, C., GERARD, F.-M., ROEGIERS, X. (Éds). *Quel avenir pour les compétences ?* Bruxelles : De Boeck Université, pp.187-191.
- DE KETELE, J.-M. (2001a). Place de la notion de compétence dans l'évaluation des apprentissages, in FIGARI, G., ACHOUCHE, M. (Éds). *L'activité évaluative réinterrogée. Regards scolaires et socioprofessionnels*. Bruxelles : De Boeck Université, pp. 39-43.
- DE KETELE, J.-M. (2001b). Enseigner des compétences : repères, in JADOULLE, J.-L. & BOUHON, M. (Éds). *Développer des compétences en histoire*, Louvain-la-Neuve-Bruxelles : Université catholique de Louvain et Ministère de l'Éducation, de la Recherche et de la Formation, 13-22.
- DE KETELE, J.-M. & GERARD, F.-M. (2004). « La validation des épreuves d'évaluation selon l'approche par les compétences », *Mesure et évaluation en Éducation*, à paraître.
- D'HAINAUT, L. (1983). *Des fins aux objectifs*, Bruxelles-Paris : Labor-Nathan.
- DOLZ, J. & OLLAGNIER, E. (Éds) (2002). *L'énigme de la compétence en éducation*, Bruxelles : De Boeck Université.

- FOUREZ, G. (1999). « Compétences, contenus, capacités et autres casse-têtes », *Forum des pédagogies*, mai, 26-31.
- GERARD, F.-M. (2004). *Tout dépend de ce qu'on veut évaluer*, Inédit, publié sur www.bief.be.
- GERARD, F.-M. (2005). « Évaluer des compétences, ou ne pas se tromper de cible », *Liaisons*, n°40, Février 2005, Beyrouth, Liban, 7-9.
- JONNAERT, Ph. (2002). *Compétences et socioconstructivisme - Un cadre théorique*, Bruxelles : De Boeck.
- LE BOTERF, G. (1994). *De la compétence. Essai sur un attracteur étrange*. Paris : Éditions de l'organisation.
- LEGENDRE, M.-F. (2001). « Sens et portée de la notion de compétence dans le nouveau programme de formation ». *Revue de l'AQEFLS*, 23(1), 12-30
- O/ DIDIYE, D., EL HADJ AMAR, B., GERARD, F.-M. & ROEGIERS, X. (2005). *Étude relative à l'impact de l'introduction de l'APC sur les résultats des élèves mauritaniens*, Inédit, accessible sur <http://www.bief.be>.
- PERRENOUD, Ph. (1997). *Construire des compétences dès l'école*. Paris : ESF.
- PROGRAMME DES NATIONS UNIES POUR LE DÉVELOPPEMENT (2004). *Rapport mondial sur le développement humain 2004*, Paris : Economica.
- RAJONHSON, L., RAMILJAONA, F., RANDRIANIRINA, P., RAZAFINDRALAMBO, M.H., RAZAFINDRANOVONA, O., RANOROVOLOLONA, E. & GERARD, F.-M. (2005). *Premiers résultats de l'APC : invitation à continuer...*, Inédit, accessible sur <http://www.bief.be>.
- REY, B. (1996). *Les compétences transversales en question*. Paris : ESF.
- REY, B., CARETTE, V., DEFRANCE, A. & KAHN, S. (2003). *Les compétences à l'école — Apprentissage et évaluation*, Bruxelles : De Boeck.
- ROEGIERS, X. (2000). *Une pédagogie de l'intégration*. Bruxelles : De Boeck.
- ROEGIERS, X. (2003). *Des situations pour intégrer les acquis*, Bruxelles : De Boeck.
- ROEGIERS, X. (2004). *L'école et l'évaluation*, Bruxelles : De Boeck.
- SCALLON, G. (2004). *L'évaluation des apprentissages dans une approche par les compétences*. Bruxelles : De Boeck Université.
- TILMAN, F. (2000). « Qu'est-ce qu'une compétence ? », *Exposant neuf*, n°2/2000, 28-31.