

De l'évaluation aux compétences : mise en perspective de pratiques émergentes

Daniel Bain
Service de la recherche en éducation, Genève

INTRODUCTION. LES COMPÉTENCES OU « QUAND DIRE, CE N'EST PAS (NÉCESSAIREMENT) FAIRE »

Dans les « textes du savoir », en particulier les introductions aux plans d'études, le terme de compétence apparaît actuellement comme une sorte de joker pour désigner des objets d'enseignement ou d'apprentissage, des « choses à enseigner ou à apprendre », dans une perspective que l'on veut résolument rénovée. Malheureusement, le lecteur a généralement de la peine à situer ce terme parmi d'autres utilisés simultanément dans le co-texte, sans articulation aucune ou presque : savoirs, connaissances, notions, conceptions, capacités, aptitudes, maîtrise, mais aussi méthodes d'apprentissage, motivation, etc. Il doit souvent essayer de deviner, dans ce discours peu ou mal réglé, quelles sont les relations d'opposition, de complémentarité ou de synonymie entre ces termes.

Par ailleurs, dans leurs analyses ou leurs discussions, les spécialistes ou chercheurs manifestent les mêmes difficultés à appréhender le concept, même si c'est à un autre niveau, même si l'on s'efforce d'établir des définitions et des oppositions. Ils donnent l'impression de les sortir d'un chapeau, prenant rarement soin de préciser leurs références épistémologiques ou les raisons (pragmatiques, sociales ou idéologiques notamment) qui les poussent à mobiliser ce concept. On ne s'étonnera alors pas de constater les efforts naïfs des responsables scientifiques des enquêtes internationales

(cf. Weinert, 1998) embarqués dans une espèce de quête de la définition de la compétence qui recevrait l'aval de tous, dans la recherche vaine d'une conception « œcuménique » des dimensions à évaluer. D'autre part, ces mêmes spécialistes constatent un hiatus certain (manque de relation, manque de cohérence) entre les définitions et l'opérationnalisation qu'en donnent les instruments (tests) des enquêtes. Ce qui apporte de l'eau à mon moulin en justifiant ce que j'avance dans mon titre : « quand dire, ce n'est pas (nécessairement) faire ».

C'est précisément pour éviter ce travers que je propose ce parcours apparemment paradoxal de l'évaluation aux compétences. L'évaluation est un bon analyseur de ce que les acteurs sur le terrain entendent par compétences. Quand on construit ou propose un instrument ou un dispositif d'évaluation, il n'est pas possible de tricher, de rester dans le vague, le flou ou l'implicite. « Dis-moi comment tu évalues les compétences et je te dirai comment tu les définis effectivement ». « Les compétences, c'est ce qu'évalue votre test » plutôt que ce que votre discours prétend échafauder.

Cette entrée par l'évaluation va m'aider à mettre en perspective des pratiques émergentes et innovantes que j'observe dans mes activités de collaboration avec des enseignants de diverses branches. Ces pratiques se réclament plus ou moins explicitement de la volonté d'enseigner – ou plutôt de faire acquérir – des compétences généralement négligées par l'enseignement dit traditionnel au profit de la transmission de connaissances ou de savoir-faire sous formes d'algorithmes ou d'automatismes (pour caractériser sommairement l'opposition esquissée par les intéressés). J'ai choisi de faire ici une étude de cas en me centrant sur l'objectif *acquisition d'une démarche scientifique* et plus particulièrement *d'une stratégie expérimentale*, qui figure dans les plans d'études de biologie et de physique. Une telle compétence, dans toute sa complexité, m'apparaît particulièrement intéressante parce qu'elle se situe comme une *compétence d'action*, s'inscrivant dans une approche théorique très actuelle (*action competence model* ; cf. le résumé qu'en fait Weinert, 1998, p.10). D'autre part, elle est très caractéristique de la perspective rénovatrice mentionnée à l'instant. La démarche scientifique peut être définie brièvement comme une « procédure de résolution de problème, reproductible, totalement explicite, ne concluant rien qui n'ait été auparavant vérifié » (B. Darley, exposé aux maîtres de biologie, septembre 1997). La compétence expérimentale, qu'elle inclut, se traduit par la capacité de concevoir, d'organiser, de réaliser et d'interpréter une expérience à partir d'un problème, mais aussi d'en rendre compte dans un rapport écrit ou oral. Je n'en dirai pas plus puisque précisément les activités impliquées par la situation d'évaluation devraient préciser ce que l'on entend effectivement par là.

Mettre en perspective une *pratique émergente* implique que l'on précise de quoi elle émerge, de quelles pratiques, conceptions, représentations

ou habitudes elle cherche à se dégager ou à se distancer. Je commencerai donc par présenter en contraste partiel, à titre de contre-exemples démonstratifs, trois exemples d'évaluations prétendant saisir une compétence expérimentale. Ils résument et illustrent bien à mes yeux des conceptions qui ont souvent cours dans l'enseignement et dont je chercherai à montrer la médiocre pertinence. J'ai emprunté ces trois exemples à des enquêtes de type *survey* susceptibles d'avoir un effet en retour (*wash back effect*) sur les évaluations et l'enseignement des maîtres de sciences.

Dans les analyses qui suivent, mon point de vue sera celui d'un observateur-participant du fonctionnement des systèmes scolaires – en particulier du système genevois – depuis plus de 30 ans. Mon approche méthodologique est donc celle d'une observation participante des pratiques formatives et évaluatives, d'une herméneutique de ces pratiques, avec tout ce que cela peut comporter de subjectivité, de prises de position théorique, idéologique, voire partisane pas toujours consciente ou explicitée. Mais, de ce point de vue, je me sens en bonne compagnie dans le monde scolaire ou académique quand il s'agit de compétences. Par ailleurs, l'interaction directe avec les acteurs permet de saisir à travers divers incidents critiques ce qui se fait et se joue effectivement dans un domaine où on a tendance à se payer de mots.

LES COMPÉTENCES EXPÉRIMENTALES, EST-CE BIEN CE QUE MESURE VOTRE TEST ?

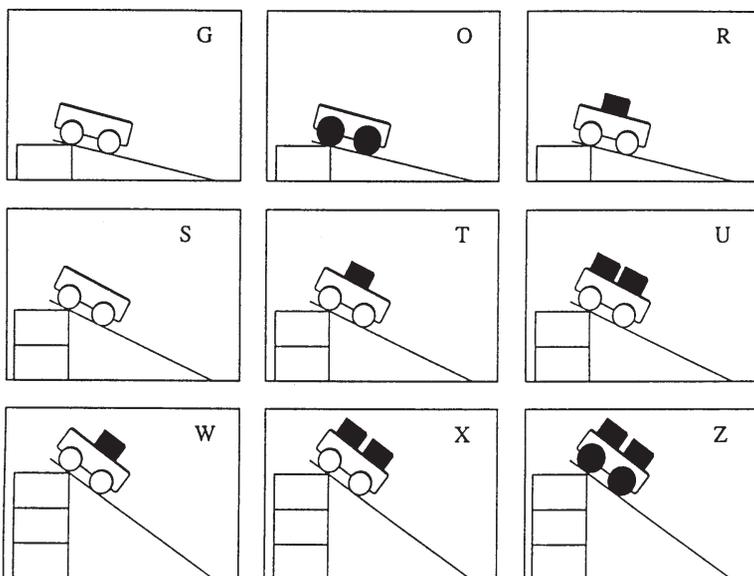
Faut-il parler de *la* compétence ou *des* compétences expérimentales ? Tout dépend du point de vue adopté sur le concept de compétence. Comme mon intention n'est pas ici d'explorer ce mot pour en faire la théorie (d'autres auteurs ont dû s'en charger dans cet ouvrage), je dirai simplement, en m'inspirant de Le Boterf (1998, chap. 2), qu'il s'agit d'une combinatoire complexe, d'un agencement coordonné, multidimensionnel, toujours à recréer, de la mobilisation de ressources multiples, de savoirs, de savoir-faire, de stratégies, d'habiletés manuelles, d'attitudes, de valeurs privilégiées... L'important dans cette amorce de caractérisation est de souligner les adjectifs *complexe* et *multidimensionnel*. Typiquement, certains tests ou items prétendant mesurer *la* compétence expérimentale (s')en donnent une représentation simpliste, abusivement unidimensionnelle. Pour ma part, j'aurais donc tendance à utiliser le plus souvent un pluriel, constatant que les conduites impliquées par ce concept participent des intentions à la fois rationnelles et scripturales définies par Rey (1996) à propos des compétences transversales.

J'emprunte mes deux premiers exemples à l'enquête TIMSS (Third Mathematic and Science Study, passée en 1995 ; Sterbler, Reusser, Ramseier,

1998), décrite comme « un effort significatif pour le travail sur les compétences dans le contexte de l'OCDE » (Salganik, Rychen, Moser & Konstant, 1998, p. 24). Le troisième, tiré d'une recherche américaine, a été présenté par L. H. Salganik dans une conférence au Service de la recherche en éducation (SRED) à Genève en décembre 1997. Les deuxième et troisième exemples sont présentés par les auteurs cités comme des solutions intéressantes pour améliorer l'évaluation des compétences expérimentales, parce qu'elles mettent les élèves dans des situations concrètes et plus authentiques de résolution de problèmes. Enfin, le troisième prétend introduire la dimension du travail en équipe.

Exemple 1 *TIMSS, test papier-crayon, item I12*

Les dessins ci-dessous montrent différents essais que Jean a effectués avec des chariots ayant des roues de taille différente. Il les a fait partir de différentes hauteurs et les blocs qu'il a mis dedans ont la même masse.



Il veut tester l'idée suivante : plus le chariot est lourd, plus sa vitesse au bas de la pente est grande. Quels sont les trois essais qu'il doit comparer ?

Exemple 2 *TIMSS, test pratique : une des six expériences proposées*

TABLETTES

*Résumé de la consigne**

L'élève travaille seul en suivant un protocole d'expérience. Celui-ci lui précise d'abord le matériel dont il aura besoin : eau chaude et froide, plusieurs verres (*becher*), quelques tablettes (d'un produit non identifié), une cuiller, une montre avec une trotteuse marquant les secondes, un thermomètre, un triple décimètre. Il doit « observer avec quelle rapidité les tablettes se dissolvent à différentes températures ». On lui demande d'écrire le plan de son expérimentation en indiquant ce qu'il veut mesurer, combien de mesures il va effectuer et la façon de représenter ses mesures dans un tableau. Pour préparer l'interprétation des résultats, on l'interroge sur l'effet de la température sur la vitesse de dissolution et on l'incite à donner une explication du phénomène.

* texte original en allemand ; mon adaptation.

Exemple 3 Situation expérimentale proposée par une enquête américaine (communication de L. H. Salganik au SRED, Genève, 1997)

Vendredi, tâche 2*

Titre : Rebond d'une balle

CONSIGNES

Vous travaillerez en équipes. Vous allez mesurer la hauteur de rebond d'une balle de tennis lâchée depuis quatre différentes hauteurs : 100 cm, 80 cm, 60 cm, 40 cm. Chaque membre de l'équipe aura une tâche à réaliser désignée par les numéros 1, 2, 3 ou 4. Votre affectation à une de ces tâches sera inscrite au tableau noir. Le numéro 3, l'observateur, doit s'assurer d'être assis bien en face de la règle métrique. Quand l'observateur repère la hauteur du rebond, cette hauteur est mesurée en partant du plancher jusqu'au bas de la balle.

[...].

N.B. Les tâches de chaque équipier sont décrites de façon détaillée dans la suite de la consigne, qui prescrit 3 essais pour chaque hauteur. Le tableau nécessaire pour enregistrer et organiser les résultats est fourni tout prêt (4 hauteurs x 3 essais).

* texte original en anglais ; ma traduction.

Chacun de ces exemples mériterait un commentaire approfondi. Je me contenterai ici d'une analyse globale et rapide, comptant sur l'esprit critique de mes lecteurs pour repérer plus en détail dans les trois échantillons de questions ci-dessus la réduction que ce type d'évaluation fait subir au concept de *démarche scientifique* et de *compétence(s) expérimentale(s)*.

- Tous les spécialistes de l'expérimentation vous le diront : la question de départ joue un rôle central et primordial dans la démarche heuristique. Or, paradoxalement, elle se situe en arrière-plan dans nos trois exemples. Dans le premier, elle est même parachutée à la fin de la consigne. De même, dans le deuxième, elle n'apparaît qu'une fois présentée la liste du matériel. Dans le troisième, les élèves savent qu'ils vont mesurer des rebonds, mais pour quoi faire ? En fait, dans les trois cas, le problème lui-même, en tant que question ayant du sens ou de l'intérêt pour l'expérimentateur, n'apparaît pas. En quelque sorte, l'expérience ne sert à rien... et à personne. La capacité à résoudre un problème est ainsi déconnectée de ses dimensions motivationnelles au sens large (opinions, valeurs, intérêts, attitudes, habitudes). En possession de la réponse, l'élève ne sait d'ailleurs guère qu'en faire, sinon la transformer en bonne note.
- La démarche heuristique est plus ou moins préprogrammée à travers la consigne, et non pas à inventer. On vérifie si l'élève est capable de suivre des consignes plus que de créer son plan d'expérience et de le modifier de façon dynamique. Aucune manipulation préalable n'est possible (ex. 1 et 3) alors que, selon mon expérience, des adultes enseignants dans la même situation s'empressent d'« essayer pour voir » et pour comprendre de quoi il retourne. Dans l'exemple 2, une question finale porte sur une modification éventuelle du plan primitif, mais manifestement une telle modification risque d'apparaître aux yeux des intéressés comme la rectification d'une erreur de départ plutôt que comme un avatar heuristique légitime. La référence sous-jacente semble bien être le modèle linéaire OHERIC (Observation, Hypothèse, Expérimentation, Résultats, Conclusion), modèle dont l'inadéquation à la pratique expérimentale est patente, mais qui continue à avoir du succès auprès des enseignants... et des évaluateurs (Bain, 1990).

- Les facteurs à considérer, à faire varier ou à neutraliser, sont donnés, directement ou indirectement, à travers les manipulations à faire ou encore la liste du matériel. Il n'y a apparemment qu'une façon de parvenir à la solution juste : appliquer correctement un raisonnement de type formel, bernardien (« toutes choses égales par ailleurs »), reposant sur la combinatoire des facteurs présentés.
- La situation est abstraite au sens premier de ce terme. Dans l'item I12, par exemple, on a retiré des problèmes tout ce qui peut matériellement faire problème : fixation et frottement des roues, position du lestage, détermination de la position du chariot au départ, repère à choisir étant donné des angles différents du plan, mesure de la vitesse (ex. 1).
- Le problème central de la mesure et des erreurs de mesure est soit ignoré ou négligé (ex. 1 et 2), soit court-circuité (ex. 3 : trois essais sont prescrits dans la consigne ; pourquoi trois, au juste ?). L'organisation des résultats est indûment facilitée par l'allusion à un tableau à établir (ex. 2) ou par la fourniture de ce tableau avec ses deux dimensions étiquetées (ex. 3).
- L'élève est seul face au problème (ex. 1 et 2) ; c'est même apparemment une condition essentielle de ce type d'évaluation, la compétence étant conçue comme une caractéristique spécifiquement individuelle. Là où l'on prétend réintégrer dans l'activité expérimentale une dimension de travail d'équipe (ex. 3), c'est de façon véritablement caricaturale. La consigne décrit ce qu'un sociologue appellerait une *division du travail* plutôt qu'une réelle *collaboration*. L'entrepreneur-expérimentateur est en l'occurrence l'adulte qui conduit et organise l'évaluation... et l'expérience. On passe évidemment à côté de l'aspect social du travail en équipe (négociation des rôles, régulation des interactions en cours de tâche, par exemple), avec tout ce que cela implique d'interrelation entre le cognitif (adaptation aux contraintes de la tâche), l'affectif (préférences personnelles pour telle activité) et le social (optimisation du travail d'équipe en vue de la réussite de l'action).

On alléguera qu'une partie de ces caractéristiques sont déterminées par la situation d'évaluation, par la nécessité de simplifier et de standardiser les conditions d'observation, de correction, de cotation et d'appréciation pour assurer la comparabilité des résultats. Oui, mais la mesure obtenue ainsi reflète-t-elle bien ce que l'on cherche à évaluer ; en d'autres termes, est-elle valide ? Et c'est finalement ce qui importe le plus au pédagogue qui s'intéresse à l'évolution des apprentissages faits par ses élèves dans le domaine heuristique.

UNE PRATIQUE ÉVALUATIVE ÉMERGENTE : LES TESTS PRATIQUES EN BIOLOGIE

Ces dernières années, le groupe de biologie du Cycle d'orientation genevois s'est efforcé d'adapter l'évaluation commune aux objectifs d'apprentissages – et donc aux compétences – qu'il cherche à promouvoir. Insatisfait de l'évaluation des savoir-faire par les traditionnelles épreuves communes papier-crayon, il a introduit des *tests pratiques* contrôlant en particulier la capacité des élèves à inventer, à mener et à exploiter une expérience¹.

Exemple 4 Test pratique de biologie
(Cycle d'orientation, Genève, 1996)

Courrier des lecteurs*

Plusieurs camarades de mon équipe de volleyball m'ont affirmé que les sportifs ont un volume respiratoire plus grand que les non-sportifs. Qu'en est-il en réalité ?

Anna Bolize

[...]

Comment t'y prendrais-tu pour répondre à Anna de façon aussi scientifique que possible ?

* texte fictif (et annoncé comme tel) collé sur une page du courrier des lecteurs d'une revue scientifique pour jeunes.

Ce test a connu diverses versions. Il a eu comme caractéristique de partir de questions relativement ouvertes telles que celle reproduite dans l'exemple 4. La consigne demandait aux élèves, regroupés par équipes, d'organiser l'expérience de façon à obtenir une (voire des) réponse(s) à la question. Ce test s'étant déroulé de façon très diverse d'une classe à l'autre, il n'est pas

1. Pour une illustration de la pédagogie préconisée, on se référera notamment au film réalisé par le Centre de production audiovisuel du Département de l'instruction publique genevois sous le titre « Quand on sait pas, on peut pas répondre » ; conception pédagogique : A. Conti et J.-L. Rimaz ; réalisation : P. Conscience.

question de décrire dans le détail ni de résumer ici l'ensemble de l'opération. Comme le laisse bien entendre dans mon titre le terme de « pratique émergente », il ne s'agit pas de présenter ce dispositif d'évaluation comme abouti, exemplaire, mais bien comme une tentative de faire un bilan plus valide des compétences visées par le plan d'études, de mettre en cohérence objectifs de formation, enseignement et évaluation. Mon intention est de montrer à la fois l'intérêt et les difficultés de l'opération.

Par souci de brièveté, plutôt que de décrire le dispositif sous ses diverses réalisations (cf. à ce propos A. Conti, 1997), je vais en caractériser les aspects principaux en relation avec le thème : l'évaluation face aux compétences, pour en saisir moins la réalité que les tendances, les ouvertures ou la problématique.

- Les objectifs couvrent assez largement les aspects de ce que l'on appellera rapidement les compétences expérimentales ; les élèves doivent parcourir tout le processus heuristique : de la formulation d'une question (hypothèse) susceptible d'être traitée scientifiquement (cf. supra la définition de la démarche scientifique) au compte rendu final faisant le point sur la démarche et les résultats.
- Le problème est motivé et contextualisé pour accrocher les élèves (Conti, 1997, p. 24), en recourant par exemple à un article de presse (« Check-up de Miguel Indurain » ou lettre fictive de l'exemple 4). Dans une classe, le maître précise qu'il n'a pas la réponse et que celle-ci ne se trouve pas directement dans les livres. C'est donc une « vraie question » ! L'expérience montre que les élèves doivent passer d'une attitude scolaire : « il s'agit de trouver ou deviner la bonne réponse » à une attitude plus heuristique-scientifique d'exploration : « la solution est à construire ».
- Les éléments de l'hypothèse (variable indépendante et dépendante) doivent être précisés : l'opposition *sportif vs non-sportif* est-elle dichotomique (point de coupure ?) ou continue (temps hebdomadaire de pratique ou d'entraînement) ? Comment mesurer une capacité respiratoire ?
- Le dispositif de mesure et d'enregistrement des données doit être construit de toutes pièces à partir des registres des conduites acquises par les élèves à d'autres occasions (mesure d'un volume, conception d'un tableau ou d'un graphique). Les élèves doivent affronter les vraies difficultés : comment définir et placer les dimensions ou les axes des tableaux et des graphiques (*vs* placer les données dans des cases toutes préparées). Ils sont incités à mobiliser de façon fonctionnelle les moyens appris ou enseignés antérieurement plutôt qu'à la manière conventionnelle qu'induisent les questions d'épreuves classiques (cf. exemples *supra*).
- Le test comporte une phase de travail individuel et une phase de travail par groupes de 2 ou 3 élèves. Les équipes doivent être constituées par les élèves (négociation), en sachant qu'une partie de l'évaluation (sous

forme de grille) prendra en compte le travail de groupe et le comportement de chacun dans la collaboration. Soulignons que dans la philosophie de cette approche didactique, il ne s'agit pas d'une simple modalité de travail. À travers cette contrainte, les enseignants essaient de faire passer le concept d'*équipe de recherche* coordonnant ses travaux (sens étymologique de *collaboration*) en vue de la réussite d'un projet.

- Le test lui-même, la façon de présenter le problème et les modalités de travail proposées aux élèves ont fait l'objet d'une discussion entre les maîtres concernés dans les établissements qui ont choisi cette forme d'évaluation. De plus, l'application a pu varier d'une classe à l'autre en fonction du « génie propre » ou des conceptions pédagogiques du maître et en fonction des élèves qu'il avait : présentation et formulation du problème, types d'interventions du maître, alternance de travail individuel et en groupe. On renonce donc à une standardisation de la situation d'évaluation, dans le but de mieux adapter celle-ci à l'objectif de contrôle des compétences, ce qui fait naturellement problème si l'on cherche à comparer les classes.
- La situation est présentée comme un « test séquence d'apprentissage » et se caractérise comme un projet de classe. Très ouverte, elle autorise les « essais pour voir », les rectifications de tirs. Mais aussi des stratégies heuristiques très diverses, comme le montre l'analyse de ce qui s'est passé dans les classes (Conti, 1997, pp. 26 et ss.) : recherche documentaire et argumentation sur cette base, enquête, observation, expérimentation.
- La tâche contraint les élèves – mais aussi le maître – à gérer la complexité et l'imprévu qu'implique toute recherche de ce type. Activités et matériel sont en étroite dépendance. L'expérimentation ne se réduit pas à la reprise d'une hypothèse classique et d'un dispositif bien connu, pour lesquels des procédures et du matériel existent tout prêts. Une bonne coordination au sein de chaque équipe est nécessaire pour la réussite des opérations.
- La forme du produit demandé se rapproche d'un projet ou d'un rapport d'expérience (selon l'étape ou le développement donné à l'opération), texte explicatif dont les aspects formels ont pu être étudiés aux leçons de français. S'agissant d'une forme socialisée, d'un genre textuel existant, on peut lui appliquer pour l'évaluation des références moins arbitraires que les critères scolaires habituellement appliqués.
- Ce n'est plus l'aspect formel de la réponse qui est important mais son degré de pertinence par rapport à la tâche. L'appréciation ne porte d'ailleurs plus seulement sur un produit-résultat condensé dans une réponse finale, juste ou fausse, mais aussi et surtout sur le processus.

- Les discussions à l'intérieur de chaque équipe et l'obligation de rédiger un compte rendu à l'intention de la « communauté scientifique » (concept introduit et commenté dans certaines classes à de telles occasions), et non plus seulement pour le maître, induisent une certaine décentration à la fois par rapport au point de vue propre et par rapport à la situation scolaire habituelle. Les descriptions et argumentations demandées incitent les élèves à des prises de conscience et à des réflexions à vertu métacognitive. Le dispositif évaluatif les pousse à « rendre compte pour se rendre compte », selon une formule qui m'est chère et qui montre l'articulation entre interaction sociale et métacognition (Bain, 1997).
- La situation-problème contraint les expérimentateurs-apprentis à faire flèche de tout bois, à mobiliser, à transférer et à intégrer des connaissances, des habiletés, des procédures ou plus généralement des opérations et des schèmes acquis ailleurs qu'au cours de biologie : dans la vie de tous les jours ou dans les cours de travaux manuels (montages), en observation scientifique (stratégie expérimentale), en mathématiques et informatique (traitement des données), en français (compte rendu, texte explicatif), en dessin (schémas).

Sur le plan épistémologique, des attitudes et des représentations, un tel dispositif évite de renforcer, comme d'autres formes d'évaluation,

l'image d'un savoir établi, d'une masse de connaissances inscrites dans les livres et les manuels dont nous exhumons parfois une démonstration expérimentale. [...] Cela demande en revanche de poser un problème nouveau aux élèves, un problème dont la réponse ne se trouve ni dans le cours, ni à la bibliothèque, ni sur le visage du maître (Conti, 1997, p. 33).

PROBLÈMES ET LIMITES DE CETTE APPROCHE DU « TEST SÉQUENCE D'APPRENTISSAGE »

En tant que pratique explicite, publique et sociale, l'évaluation constitue un bon révélateur des problèmes de toute sorte. Elle oblige le théoricien ou le praticien à renoncer aux camouflages du discours pédagogique ordinaire. Une opération comme celle à laquelle je me réfère ici sommairement, suivie attentivement par une commission du groupe de biologie, tient du défi et de la gageure ; les collèves et maîtres volontaires ne se sont d'ailleurs pas bousculés au portillon (13 collèves sur 17 ont choisi un test « méthodologique » plus simple, contrôlant une expérience ou une manipulation faite en classe). Par les réactions et les analyses qu'elle a suscitées, elle a mis en évidence divers problèmes entrant dans la problématique : l'évaluation face aux compétences. J'aborde ici quelques-uns de ces problèmes sans prétendre être exhaustif.

Le premier obstacle concerne l'apprentissage des compétences elles-mêmes. Évaluer des compétences complexes, c'est bien ; encore faut-il que les élèves aient reçu l'enseignement adéquat ! L'analyse des résultats par Conti (1997) à partir d'une soixantaine de copies d'élèves montre des différences importantes entre les classes. « On peut soupçonner que le type d'enseignement prodigué par les maîtres est principalement responsable de ces différences » (p. 38). Toutes les classes n'ont pas eu les mêmes opportunités d'exercer le processus heuristique dans sa complexité. À partir de ce constat, on peut se demander comment la situation va évoluer du point de vue institutionnel. L'effet en retour de ce type d'évaluation sur l'enseignement va-t-il influencer réellement celui-ci ?

Un deuxième groupe de problèmes est d'ordre docimologique : comment assurer aux appréciations l'objectivité et la fidélité souhaitables et aux résultats la généralisabilité nécessaire ?

Il faudrait d'abord se mettre d'accord sur l'objectif de ce type de contrôle. Pour les enseignants, il s'agit essentiellement de vérifier si les objectifs d'apprentissage fixés par le plan d'études ont été atteints. Mais aussi de récolter des informations utiles pour réguler leur enseignement. Mais aussi, à l'occasion, de disposer de quelques éléments pour l'orientation en fin d'année². Dans la pratique, il faut peut-être se résigner à ce que ces évaluations aient à la fois des fonctions de bilan certificatif, de régulation formative et de pronostic. Sur le plan docimologique, la solution réside probablement dans la diversification des regards portés sur les conduites observées : prise en considération de la distance entre comportements manifestés et comportements attendus, analyse du processus menant aux réussites ou aux échecs, hypothèses sur les opérations mobilisées et sur les obstacles rencontrés, mise en relation avec des résultats antérieurs... Pour remplir ces différentes fonctions, les observations évaluatives doivent impérativement se référer à l'ensemble de la séquence didactique, qui met en perspective les données recueillies (Bain & Schneuwly, 1993). Ce qui suppose un observateur-évaluateur expérimenté, comme nous allons le voir.

Un autre obstacle docimologique tient au fait qu'on a affaire à un processus séquentiel dont toutes les étapes sont liées. Un mauvais choix à un moment donné peut compromettre la suite des opérations. Cela implique l'intervention judicieuse du maître, amené en cas de besoin à faire diverses suggestions ou à donner certains coups de pouce, pris en compte dans l'appréciation finale.

Pour l'appréciation du travail tout au long de la séquence, les auteurs du test ont proposé deux grilles d'évaluation : individuelle et par groupe.

2. Je laisse ici de côté les objectifs de l'institution (contrôle des maîtres, évaluation du système de formation...), pour ne pas rendre l'analyse encore plus complexe.

Classiquement, ces grilles distinguent diverses conduites repères (« entrée en matière sans/avec référence à une théorie explicative » ; « les élèves s'écoutent... ») évaluées (quantification intensive) au moyen d'un curseur. Pour les concepteurs du test, la principale difficulté consiste à choisir les bons repères. Pour les enseignants, le problème majeur est de transcrire les informations pour chaque élève de leur classe (en fait, une demi-classe), puis de reconstituer à partir des traces recueillies par la grille et par le protocole d'expérience ou journal de bord une image de l'ensemble des activités de chacun. Le plus gênant est de ne plus pouvoir justifier « objectivement » son évaluation finale en s'appuyant sur un total de points. La discussion de références critérielles, primordiales dans ce type de test, met en évidence des différences importantes entre maîtres en ce qui concerne l'objet d'évaluation et du même coup les compétences (Qu'est-ce qu'une « référence à une théorie explicative » ? une référence au cours, à une représentation construite par l'élève ?).

Un autre doute important plane sur une telle évaluation. Portant sur un problème particulier, ses résultats sont-ils généralisables à d'autres situations expérimentales ? L'expérience acquise en *observation scientifique* comme d'autres travaux dans ce domaine incite à des réserves importantes : la mise en œuvre des compétences est en partie spécifique au problème et à son contexte. Et il n'est généralement pas possible de multiplier les contrôles. Mais il est vrai que cette réserve est valable aussi pour les épreuves classiques papier-crayon.

Je passe sur bien d'autres problèmes qui se posent dans ce genre de test pratique. Alors, que faire ? Il s'agit d'améliorer le dispositif d'évaluation en se centrant sur les compétences de l'évaluateur plutôt que sur les qualités docimologiques d'un instrument mythique fournissant clés en main un diagnostic. J'aimerais esquisser quelques pistes allant dans ce sens dans mes conclusions.

EN GUISE DE CONCLUSION : QUELQUES SUGGESTIONS POUR UNE ÉVALUATION PLUS VALIDE AU SERVICE DE LA CONSTRUCTION DES COMPÉTENCES EXPÉRIMENTALES

Il faut d'abord être conscient qu'en passant du premier type de test décrit aux pratiques évaluatives présentées comme émergentes, il y a changement de perspective, rupture à la fois épistémologique, pédagogique et idéologique. Cette rupture a naturellement son parallèle dans les conceptions des compétences et de la pédagogie qui correspondent à ces types d'évaluations. Je n'insisterai pas sur le fait qu'on passe d'une conception de la construction du savoir fortement empreinte de behaviorisme à une

perspective de tendance socio-constructiviste. Quelques auteurs proposent une schématisation ou théorisation permettant de caractériser l'opposition entre les perspectives illustrées par les items de TIMSS d'une part et le test pratique de biologie d'autre part. J'en citerai brièvement trois.

Le Boterf (1998) met en évidence deux modèles de la compétence, qui interfèrent actuellement dans les pratiques de gestion. L'analogie avec le domaine pédagogique – et avec les pratiques évaluatives décrites ci-dessus – est immédiate ; je cite le texte pour que le lecteur puisse en juger (remplacer dans la citation ci-dessous *sujet* ou *professionnel* par *élève*) :

Dans le modèle « A », hérité des conceptions taylorienne et fordienne, le sujet est considéré comme un opérateur dont la compétence se limite à savoir exécuter des opérations conformes à la prescription. La compétence se limite à un savoir-faire descriptible en termes de comportement attendu et observable. Une compétence se décrit avec une maille étroite cohérente avec la parcellisation des savoir-faire. [...].

Dans le modèle « B », préfigurant peut-être ce qui se met en place avec la montée de l'économie de service, le sujet est considéré davantage comme un acteur que comme un opérateur. Le professionnel compétent est celui qui sait aller au-delà du prescrit, qui sait agir et donc prendre des initiatives. Face à des actions complexes, il est considéré qu'il y a plusieurs façons d'être compétent et que diverses conduites peuvent être pertinentes. La conduite ne se réduit pas à un comportement. La maille choisie pour décrire la compétence est assez large ; elle reconnaît la faculté pour le sujet d'enchaîner des ressources et des actions. Elle ne confond pas la compétence et les ressources (savoir, savoir-faire...) à mobiliser pour la construire [...] (p. 91).

Sur le plan sociologique et idéologique, les deux conceptions de l'évaluation et des compétences (ainsi que des pédagogies qui les sous-tendent) correspondent aux deux *codes du savoir scolaire* de Bernstein (1975) :

- le *code-série*, qui différencie et hiérarchise les choses et les gens, définit des frontières strictes entre les domaines de savoirs et les statuts des personnes, favorise la compétition, structure de façon rigide la transmission des connaissances et sépare espaces scolaire et social (*things must be kept apart*) ;
- le *code intégré*, qui constitue ou reconstitue les liens entre les choses et entre les personnes à partir d'une finalité commune, articule les différents domaines de connaissances ainsi que leurs apprentissages, favorise la collaboration et rapproche l'école du monde extérieur (*things must be put together*)³.

3. Pour des développements sur ces codes du savoir scolaire et leur traduction dans les divers aspects du fonctionnement de l'école, cf. Bain, 1979, pp. 512-518.

Sur le plan des choix pédagogiques et d'évaluation, le modèle de Desjardins, repris par Bélair (1999, p. 17), oppose de façon analogue :

- des *stratégies militaire* (orientée par l'enseignant, pédagogie de l'empreinte, évaluation sanction) ou *populaire* (orientée par le produit, pédagogie du conditionnement, évaluation notation) ;
- à des *stratégies ouverte* (orientée par le processus, pédagogie constructiviste, gestion de l'erreur) ou *libre* (orientée par l'apprenant, pédagogie humaniste, auto-évaluation).

Les pratiques émergentes impliquent donc une certaine conversion (changement de direction et de conviction) épistémologique et pédagogique. Comme le suggère Bélair (1999, p. 117) une telle « trans-formation », un tel passage à de nouvelles approches, ne peut se faire qu'en recourant à une formation adéquate. Au « quoi faire ? » de tout à l'heure les biologistes du Cycle d'orientation genevois ont répondu par la création d'un réseau de formation continue. Il s'agit notamment par des interventions de spécialistes, des travaux en équipes et des échanges :

- d'améliorer les connaissances scientifiques des maîtres quant aux objets d'enseignement ; dans le domaine des conduites expérimentales, certains en sont restés à des conceptions stéréotypées de type OHERIC et risquent bien de reproduire dans leur enseignement et leurs évaluations des représentations simplistes de la compétence expérimentale ;
- d'apporter un éclairage épistémologique sur les principales notions figurant au plan d'études ; quand on s'occupe de transmission et d'apprentissage de connaissances, il est indispensable d'étudier la façon dont celles-ci sont construites dans diverses perspectives et à diverses époques ;
- d'élaborer ensemble des séquences didactiques définissant clairement des objectifs d'apprentissage, des activités et des moyens de contrôle (quitte à les renégocier en cours de séquence, si nécessaire) ;
- de poser clairement les enjeux et les modalités des évaluations communes, de préparer des tests pratiques du type de ceux présentés ci-dessus, d'en discuter non seulement les résultats et les problèmes, mais aussi les exploitations possibles ;
- de saisir cette occasion particulièrement opportune de l'évaluation pour se demander quelles compétences on cherche à construire, à renforcer et pour remettre en cause, le cas échéant, les objectifs et les méthodes de l'enseignement.

Les problèmes proprement docimologiques de validité et fiabilité des résultats ne sont pas tous résolus pour autant. Comme ils sont massifs, je ne peux qu'esquisser quelques pistes. Dans la perspective émergente évoquée ci-dessus à titre d'exemple (on en trouverait d'analogues dans la didactique des langues), la priorité est donnée nettement à l'enseignement et aux apprentissages sur l'évaluation : faut-il le regretter ? L'obsession évaluative pollue littéralement nos activités didactiques. Dans les évaluations de type sommatif-certificatif considérées comme indispensables, les validités de contenu et de *construct* doivent donc être privilégiées étant donné leur effet en retour sur l'enseignement ; elles impliquent des situations-problèmes du type du *test pratique* présenté ci-dessus. L'appréciation et l'interprétation des résultats, pour améliorer la fiabilité de l'évaluation, devraient bénéficier des repères apportés par :

- l'auto-évaluation de l'élève, l'évaluation mutuelle dans l'équipe, la coévaluation du travail par le maître et l'élève⁴, la discussion des observations faites avec d'autres maîtres, dans un travail d'équipe, cette multiplication des points de vue aidant à relativiser celui du maître évaluateur ;
- la mise en relation des conduites enregistrées lors du contrôle avec d'autres observées antérieurement au cours de l'enseignement ; l'intégration du contrôle dans la séquence didactique elle-même, qui permet de situer les observations dans une évolution (Bain & Schneuwly, 1993) ; la constitution d'un portfolio (ou dossier progressif : Bélair, 1999, pp. 69-114) peut y contribuer efficacement ;
- la compétence d'analyse clinique acquise par le maître dans la formation continue décrite ci-dessus.

C'est sur ce dernier point que je proposerais d'investir surtout. Il faudrait finalement renoncer, comme on le fait trop souvent, à déléguer à un instrument, même sophistiqué, la tâche de juger, de décider ou même de réguler. On a tout intérêt à ce que la compétence évaluative (aux différents sens de ce terme) revienne prioritairement aux enseignants si l'on veut une bonne adéquation entre les objectifs visés par le plan d'études, les compétences exercées par les apprentissages, les méthodes utilisées par l'enseignement et l'évaluation.

4. Sur le rôle de l'élève dans l'évaluation (formative) et les stratégies d'autoévaluation, évaluation mutuelle, coévaluation, cf. notamment Allal, 1993 et Allal & Michel, 1993.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (1993). Régulations métacognitives : quelle place pour l'élève dans l'évaluation formative. In L. Allal, D. Bain & Ph. Perrenoud (Éd.), *Évaluation formative et didactique du français* (pp. 81-98). Neuchâtel : Delachaux et Niestlé.
- Allal, L. & Michel, Y. (1993). Autoévaluation et évaluation mutuelle en situation de production écrite. In L. Allal, D. Bain & Ph. Perrenoud (Éd.), *Évaluation formative et didactique du français* (pp. 239-264). Neuchâtel : Delachaux et Niestlé.
- Bain, D. (1979). *Orientation scolaire et fonctionnement de l'école*. Berne : P. Lang.
- Bain, D. (1990). OHERIC, le choc. *Lettre LDES, numéro spécial pour le 10^e anniversaire du LDES*, 13.
- Bain, D. (1997). La métacognition au banc d'essai de la didactique. In B. Noël, D. Martin, P.A. Doudin & D. Bain (Éd.), *La métacognition : un concept utile pour mieux apprendre ?* (pp. 27-37). Neuchâtel : Institut de recherche et de documentation pédagogiques.
- Bain, D. & Schneuwly, B. (1993). Pour une évaluation formative intégrée dans la pédagogie du français : de la nécessité et de l'utilité de modèles de référence. In L. Allal, D. Bain & Ph. Perrenoud (Éd.), *Évaluation formative et didactique du français* (pp. 51-79). Neuchâtel : Delachaux et Niestlé.
- Bélair, L. M. (1999). *L'évaluation dans l'école. Nouvelles pratiques*. Paris : ESF.
- Bernstein, B. (1975). *Class, codes and control, vol. 1, Towards a theory of educational transmissions*. London : Routledge & Kegan.
- Conti, A. (1997). *Analyse des tests pratiques de biologie 8^e S/Option 2 (1995-1996)*. Genève : Commission Évaluation, groupe de biologie, Cycle d'orientation.
- Le Boterf, G. (1998). *De la compétence à la navigation professionnelle*. Paris : Les Éditions d'Organisation.
- Piaget, J. & Inhelder, B. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris : PUF.
- Rey, B. (1996). *Les compétences transversales en question*. Paris : ESF, Collection Pédagogies.
- Salganik, L. H., Rychen, D. S., Moser, U. & Konstant, J. W. (1998). *Projects on Competencies in the OECD Context : Analysis of Theoretical and Conceptual Foundations. Draft*. Berne : OECD/OCDE, DeSeCo.
- Stebler, R., Reusser, K. & Ramseier, E. (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenz, Erträge aus dem TIMSS-Experimentiertest. *Éducation et Recherche*, 20, 1, 28-54.
- Weinert, F. E. (1998). *Concepts on Competence, first version*. Berne : OFS, OCDE, DeseCo document de travail.