

3. TECHNIQUES DE MESURE DANS L'AUTOÉVALUATION

(DIX TECHNIQUES D'AUTO-ESTIMATION DE LA QUALITÉ DE SES RÉPONSES)

Dieudonné LECLERCQ, Jean-Luc GILLES

3.1 Introduction

Depuis des décennies, l'évaluation pédagogique a emprunté à la psychologie l'arsenal de ses méthodes, de ses instruments et des techniques statistiques. Même des apports aussi récents et importants que la théorie de la généralisabilité (Cronbach *et al.*, 1972 ; Cardinet *et al.*, 1975) ou que le modèle de Rasch (Rasch, 1960 ; Wright & Stone, 1979) ont d'abord été formulés dans un contexte psychométrique. Globalement, ces apports ont incontestablement été bénéfiques à l'éducation et continueront à l'être, même si, sur certains points, de regrettables confusions sont possibles, comme « le dangereux mythe de la courbe de Gauss » dénoncé par De Landsheere (1976). À côté de la psychométrie, cependant, se constitue progressivement ce que Carver (1974) a appelé l'éduométrie, c'est-à-dire un ensemble d'approches, de concepts, d'instruments et de techniques spécifiques aux problèmes pédagogiques, à l'évaluation des apprentissages individuels.

A. Vers une éduométrie instrumentée

Souvent, les évaluateurs en éducation pourraient être comparés à des chimistes qui travailleraient avec une pelle à charbon. Or on sait que la mesure précise des masses a amené des progrès décisifs en chimie. Si nous voulons comprendre un peu mieux les mécanismes d'apprentissage et de traitement de l'information, il est urgent que nous travaillions avec un plus grand degré de nuances, une plus grande précision que par le passé, que nous élaborions les loupes, les balances, les microscopes de la pédagogie. Les probabilités subjectives, ou les degrés de certitude qui en sont une forme dérivée, peuvent être l'un de ces outils. Des logiciels existent pour faciliter l'exploitation des informations issues de l'utilisation des probabilités subjectives : CERT (Boxus *et al.*, 1991), ELI (Van Lenthe, 1993), SCANTEST (Gilles, 1998). Bien entendu, se posent les problèmes de la validité, de la fidélité, de la sensibilité (ou acuité) des mesures permises avec ces nouveaux instruments (Leclercq, 1993).

Plutôt que de dire aux étudiants : « Répondez uniquement si vous savez, et omettez si vous ignorez », ce qui est une consigne ambiguë (De Finetti, 1965), on leur permettra d'exprimer leurs doutes et leurs certitudes (Shufford *et al.*, 1966 ; Adams et Adams, 1961). Au lieu de se limiter à demander une réponse à une question, on favorisera la production éventuelle de plusieurs réponses hiérarchisées selon leur vraisemblance (Leclercq, 1975 ; Bruno, 1990).

B. En finir avec la correction *for guessing*

Depuis près d'un siècle (les OTIS tests aux USA datent de 1914), les chercheurs en éducation ont essayé de « rendre équitable » la notation des étudiants à partir d'examens constitués de questions à choix multiple (QCM). Si l'on désigne par k le nombre de solutions proposées (par exemple $k = 4$), par p la probabilité de désigner la solution correcte, par H le fait « donner une réponse au Hasard » et par I l'expression « étant donné que », on a : $p | H = 1/k$ (ici = 0,25). Si l'on a prévu un tarif (T) en cas de réponse correcte (TC) de + 1 point et un tarif en cas de réponse incorrecte (TI) de 0 point, alors l'étudiant complètement ignorant a intérêt à répondre au hasard, ce qui entraîne un score oscillant autour de 25 points sur 100, si toutes les QCM ont 4 solutions proposées (dont une correcte). Ce « gonflement » du score (qui aurait dû être 0/100), dû à ce que certains appellent « la probabilité automatique », rend les scores à des QCM incomparables aux scores obtenus à des questions ouvertes. On connaît la vieille parade à ce problème : la « correction pour divination » (*correction for guessing*) qui consiste à retirer des points en cas de réponse incorrecte, bref à fixer un tarif TI négatif. Depuis Hevner (1932), on sait (pour démonstration, voir Leclercq, 1986) que l'on doit fixer TI à $-1/(k-1)$ pour que le score d'une personne qui répondrait partout au hasard oscille autour de zéro. C'est la *correction for guessing* classique qui a pour objectif de rendre nulle l'espérance mathématique. Il en découle qu'un apprenant qui ne sait rien a *autant* intérêt à omettre qu'à répondre au hasard. Leclercq (1987) a montré que la *correction for guessing* (c.f.g.) est un instrument à bannir car il présente quatre défauts que ne présentent pas les probabilités subjectives qui, elles : (1) sont basées sur un modèle théorique plus pertinent que la c.f.g., (2) sont un principe de notation plus équitable que la c.f.g., (3) sont plus formatives (pour l'apprenant) que la c.f.g. (4) sont plus informatives (pour l'apprenant et l'enseignant) que la c.f.g.

C. Une récente confusion scientifique à surmonter

Une large partie de la littérature (à très forte majorité anglo-saxonne) concernant les degrés de certitude (*confidence marking*) a été consacrée à trouver un système de notation des questions à choix multiple (QCM) qui résolve de façon élégante et scientifique le vieux problème de la « correction pour divination ». La définition d'Ebel (1968) est très révélatrice de cette limitation de l'intérêt pour les degrés de certitude à ce seul problème technique : « La pondération par degré de certitude (*confidence weighting*) est une façon spéciale de répondre aux questions d'un test [...], et une façon spéciale d'attribuer des points à ces réponses. En bref, on demande à l'étudiant d'indiquer non seulement quelle est la solution [...] qu'il croit correcte, mais en plus dans quelle mesure il est certain de l'exactitude de cette solution. Il recevra plus de points pour une réponse correcte donnée avec un degré de certitude élevé qu'avec un degré de certitude faible. Mais la pénalisation

(points négatifs) pour une réponse incorrecte donnée avec un degré de certitude fort devra être suffisamment lourde pour décourager les prétentions de certitude non fondées ».

Ce principe de notation a, en fait, une assez longue histoire (Hemmon, 1911 ; Hollingworth, 1913 ; Trow, 1923 ; Hevner, 1932 ; Jacobs, 1968). Les recherches sur le sujet ont culminé, aux États-Unis, dans la période 1960-1970, puis ont été abandonnées. Quatre raisons expliquent la période de « creux » de ces recherches.

Raison 1 : On a cru que l'enjeu principal était docimologique (une technique d'attribution de points), alors qu'il est mahétique (lié à l'apprentissage), épistémologique (conception de la connaissance et de son utilisation sociale).

Raison 2 : Les modèles sous-jacents sont mal connus ou mal acceptés. Nombreux sont ceux qui, par crainte (mal justifiée) du subjectivisme, hésitent à s'engager dans cette voie. Ils ignorent qu'il existe des techniques permettant d'étudier objectivement la subjectivité. Or l'influence de la personnalité sur l'autoestimation a été étudiée depuis longtemps (Hevner, 1932 ; Wiley et Trimble, 1936 ; Swineford, 1938 ; Jacobs, 1971). Plutôt que de se plaindre de l'« intrusion » de l'affectif dans le cognitif, le formateur et le chercheur devraient y voir une voie enfin objective d'approcher cette facette individuelle, et, par la disponibilité d'un nouveau type de données sur l'apprenant, de nouvelles possibilités d'interventions et de compréhension, notamment de la métacognition (Jans & Leclercq, 1997). Par exemple, la tendance à la surestimation souvent décrite dans la littérature semble pouvoir se nuancer lorsqu'on compare les tendances à sur et sous-estimer ses compétences chez les filles et les garçons non entraînés à l'utilisation des degrés de certitude (Gilles, 1995).

Raison 3 : Les utilisateurs sont peu conscients de la nécessité, pour recueillir des données valides, de s'en tenir à des méthodes spécifiques et rigoureuses. La plupart des données fournies à ce jour par la littérature mondiale sont de peu d'intérêt, car récoltées dans de mauvaises conditions méthodologiques (consignes « ordinales » et non « métriques », barèmes de points « élégants » et non « calculés selon la théorie des décisions »).

Raison 4 : Il existe une grande confusion quant aux interprétations à donner aux nouvelles informations recueillies. De très nombreuses recherches, notamment des traitements statistiques concernant la validité et la fidélité des nouvelles « mesures » obtenues, sont sans objet, car une grave équivoque pèse sur la notion même de « mesure » (en fait le nouveau score n'est pas une mesure, mais un « paiement » par « combinaison », pondérée, de deux mesures : la compétence et le réalisme).

3.2 Les consignes dichotomiques (C1)

Van Naersen et Van Beaumont (1965) dactylographiaient une QCM comme suit : Quelle est la capitale de la France ? Z 1 – Lyon / Z 2 – Paris / Z 3 – Marseille.

Rappelons que Z est la première lettre de *zekerheid* (certitude en néerlandais). Les apprenants peuvent choisir une solution (2 par exemple) soit en entourant seulement le chiffre 2 (pas sûr = 2) ou en encerclant Z et 2 (sûr = Z 2). Ces deux auteurs et Sandbergen (1971) ont utilisé le barème suivant : TC = +1 et TI = 0, avec Z, TC = + 2 et TI = -1. Ce barème est conforme à la théorie des décisions.

Dans la théorie des décisions, l'étudiant est supposé choisir le comportement (le degré de certitude) qui lui rapporte le score attendu (SA) le plus élevé, calculé par la formule $SA = (p.TC) + (q.TI)$ où p est la probabilité (subjective) de réussir et q celle (complémentaire) de se tromper. On peut ainsi voir qu'avec ce barème, en deçà de 50% de certitude, l'étudiant a intérêt à répondre sans Z et, au-delà avec Cross et Frary (1977) ont montré que rares sont les situations d'ignorance totale et beaucoup plus fréquentes celles de connaissance partielle. Il est donc très rare que les étudiants répondent totalement au hasard (Leclercq, 1987, p. 128-130).

3.3 Les consignes ordinales (C2)

On peut inviter les étudiants (Jacobs, 1971 ; Leclercq, 1973) à fournir un degré de certitude entre 0 et 3, avec la consigne suivante : indiquez 0 pour « pas sûr du tout », 1 pour « peu sûr », 2 pour « sûr », 3 pour « très sûr ». De telles consignes sont vagues. Ce qui est très sûr pour un apprenant peut être peu sûr pour un autre. Puisque l'expérimentateur ne connaît pas l'interprétation exacte que différents apprenants ont donnée de « sûr », deux réponses accompagnées du même degré de certitude ne peuvent être comparées, qu'elles aient été données par deux apprenants différents ou par la même personne. En effet, des contrastes de facilité entre des questions successives peuvent amener l'apprenant à réviser son interprétation des divers degrés de certitude quand il passe d'une question à l'autre. Malgré ces faiblesses, ce type de consigne a été abondamment utilisé, avec divers barèmes. Jacobs (1971) utilise le barème suivant : « je devine » (+1 ; 0), « assez sûr » (+2 ; -2), « certain » (+3 ; -3). Bien que commodes, ces barèmes doivent être abandonnés car ils ne respectent pas les critères élémentaires de la théorie des décisions. En effet, le score attendu, ou S.A., de certains degrés de certitude centraux ne doivent jamais être choisis car ils ne sont jamais supérieurs à celui des autres degrés de certitude.

3.4 *Les consignes par zones régulières d'une échelle d'intervalles (C3)*

Pour échapper aux critiques qui viennent d'être adressées aux échelles ordinales, l'expression du degré de certitude doit se faire en termes de probabilités. Par ailleurs, la sensibilité (ou acuité) d'un apprenant dans l'estimation de ses chances de succès n'est pas fine au point qu'il puisse distinguer entre 0,372 et 0,373. C'est pourquoi on se contentera de réponses probabilistes assez globales, correspondant à des zones sur l'axe des probabilités. Une des consignes utilisées par Leclercq est la suivante avec son barème de tarif (conforme à la théorie des décisions) : « sûr entre 0 et 25% » (0 ; 0), « entre 25 et 50% » (+3 ; -1), « entre 50 et 75% » (+4 ; -2), « entre 75 et 100% » (+5 ; -5). Shuford *et al.* (1966) proposent d'utiliser les 10 digits (0, 1, 2, 3, 4,... 9) pour désigner 10 intervalles (0 à 10 %, 10 à 20 %, 20 à 30 %, etc.) réguliers.

3.5 *Les consignes par étoiles (C4)*

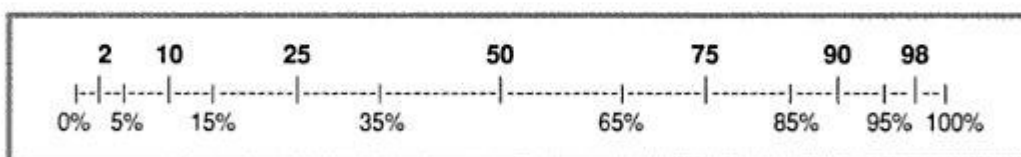
De Finetti (1965) décrit un « système à cinq étoiles » où l'apprenant doit répartir cinq étoiles (ou astérisques) sur les diverses solutions proposées d'une QCM, chaque étoile équivalant à une probabilité de 0,20. Il existe seulement sept façons de distribuer cinq étoiles. Michael (1968) a utilisé un « système à dix étoiles » où le score à la question vaut le nombre d'étoiles attribuées à la solution correcte, ce nombre étant divisé par 10. De ses données expérimentales, Michael conclut qu'un test noté selon des procédures classiques devrait être 1,7 fois plus long pour atteindre la même fidélité dans les mesures qu'avec la procédure des dix étoiles.

3.6 *Les consignes par rapports (C5)*

Edwards (1968), entre autres, a recommandé d'utiliser une échelle logarithmique pour définir les degrés de certitude. Voici un exemple d'une telle échelle : A = 1 chance sur 1000, B = 1/100, C = 1/10 ; D = π , E = \int , F = Ω , G = 9/10, H = 99/100, I = 999/1000. Cette échelle à neuf degrés correspond aux propriétés logarithmiques de la perception humaine. Leclercq (1983) a montré que cette sensibilité logarithmique ainsi que le « nombre magique 7 » de Miller (1956) s'appliquent aussi aux probabilités subjectives, comme ils s'appliquent à la vision, au toucher, à l'audition, au goût, etc. L'échelle d'Edwards paraît fondée, car nous savons que des différences aux extrémités de l'échelle sont plus perceptibles qu'au centre. Ainsi, un étudiant moyen peut faire la distinction entre 1 chance sur 100 et 10 chances sur 100 (9 % de différences) alors qu'il aura bien plus de difficultés à faire la différence entre 33 % et 42 % (la même différence de 9 %, mais au milieu de l'échelle des probabilités).

3.7 *Les consignes par zones irrégulières et asymétriques (C6)*

Gardant ceci à l'esprit, on peut définir des zones de certitudes inégales, mais symétriques. Voici l'échelle actuellement utilisée par Leclercq, Gilles et Jans (1999) qui considèrent cette technique comme le meilleur compromis étant donné leurs objectifs, leurs types d'étudiants, leurs modalités d'utilisation :



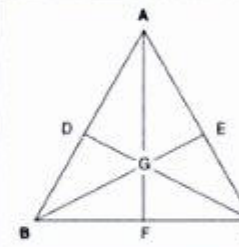
Dans le cas d'un Vrai-Faux Généralisé (VFG), chaque question étant binaire (vrai-faux), les deux degrés inférieurs à 50 % sont, en pratique, inutiles. Même dans ce cas extrême, il reste à l'étudiant le choix entre 4 degrés de certitude !

3.8 *Les consignes par échelles continues (C7)*

Cette procédure, recommandée par De Finetti (1965) sous le nom de « Continuous Confidence Marking » autorise l'apprenant à exprimer son degré de certitude avec la précision qu'il veut : il peut aussi bien répondre par « 30 % » que par « 31,27 % ». Ce mode de réponse est intéressant quand la continuité est nécessaire. Baker (1969) a développé une technique qui garantit ce principe. Il présente sur l'écran vidéo d'un ordinateur, les quatre histogrammes correspondant aux probabilités (exprimées en pourcentages) attribuées à chacune des quatre solutions proposées d'une QCM. Quatre boutons permettent à l'apprenant de changer chaque valeur (la modification est immédiatement affichée sur l'écran). Les trois autres valeurs sont automatiquement adaptées (par une règle proportionnelle) de sorte que la somme des quatre valeurs soit toujours égale à 100. Dans cet exemple, les changements n'ont pas besoin d'être effectués pas à pas, mais peuvent être continus. Du point de vue de la notation (l'attribution des points), on n'utilise plus une matrice de tarifs, mais une fonction mathématique. C'est aussi ce que font Shufford (1993) et Dirkzwager (1993).

3.9 *Les consignes par techniques du contour triangulaire (C8)*

De Finetti (1965) a proposé des consignes où l'apprenant doit désigner la seule solution correcte parmi trois solutions proposées (A, B et C).

<p>L'apprenant doit se positionner :</p> <ul style="list-style-type: none"> • soit sur un des sommets (A, B ou C) d'un triangle équilatéral ; • soit exactement entre A et B (réponse D) ; • soit exactement entre A et C (réponse E) ; • soit exactement entre B et C (réponse F) ; • soit au centre du triangle (réponse G). 	
---	---

Évidemment, de telles réponses doivent être interprétées comme suit :

« D = j'élimine C et j'hésite également entre A et B (ou je réponds A à 50 %, B à 50 % et à C à 0 %). »

« G = j'hésite entre les trois solutions proposées »

Bruno et Baxter (1989) utilisent le même principe que De Finetti, mais en offrant trois positions intermédiaires (G, H et I) entre A et B au lieu d'une seule. Ils offrent de même trois positions, entre A et C (D, E et F) et trois autres positions entre B et C (J, K et L), plus un point central (M), ce qui fait, au total, 13 possibilités (au lieu de 7 précédemment). Une telle procédure offre des possibilités de nuances considérables, puisque, G signifie : A à 75 %, B à 25 % et C à 0 %. Une telle approche ne fonctionne *que* pour trois solutions proposées.

3.10 Les consignes par fractiles (ou fourchettes) (C9)

Quand la réponse à une question porte une valeur numérique repérable sur un continuum (par exemple, un poids, une surface, un prix, une date, une durée, une vitesse, etc.), la procédure des fractiles est une manière élégante et formative de faire face au problème de la connaissance partielle. On demande à l'apprenant de fournir, pour chaque question, deux réponses (c'est-à-dire deux fractiles) : la limite inférieure et la limite supérieure d'un intervalle (ce qui constitue une « fourchette »). Plus l'intervalle est étroit, plus l'apprenant est sûr de lui et plus le risque est grand que la solution correcte « déborde » de l'intervalle ; donc plus grand doivent être le tarif (positif) en cas de succès et le tarif (négatif) en cas d'échec. Quand l'intervalle n'inclut pas la valeur correcte, on parle de « surprise ». On observe souvent un grand nombre (imprévu) de surprises. Par exemple, alors qu'on demande aux apprenants de fournir un intervalle tel qu'il inclue la réponse correcte dans 80 % des cas, on observe souvent un taux de succès inférieur à 50 %. Ce phénomène a été appelé par Pitz (1974) « hyperprécision dans les intervalles ».

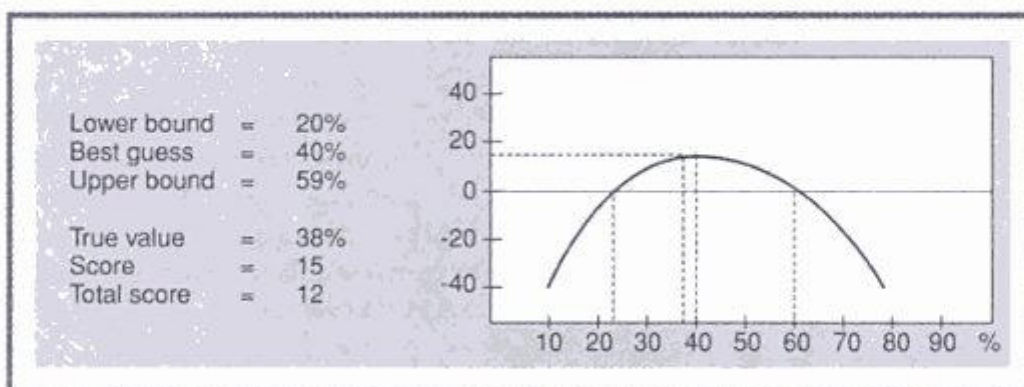
3.11 Les consignes par ajustement d'une distribution de probabilités (C10)

Van Lenthe (1993 ; 1994) permet de répondre à des questions appelant des réponses en pourcentages, par son logiciel ELI (les trois premières lettres du mot anglais *elicitation*).

« Quel est le pourcentage d'enfants de 11 ans qui passe plus de 2 heures par jour devant la TV en France, dans le sondage de mai 1966 ? »

Après avoir reçu la question (sur le haut de l'écran), l'étudiant voit s'afficher (sur le bas de l'écran) une distribution de probabilités, centrée sur 50 % et avec des « bords » éloignés de 15 % (donc 35 % et 65 %). L'étudiant peut (avec les touches → et ←) déplacer horizontalement cette courbe. La voici déplacée pour que son centre (la Moyenne) soit 40 %. Avec les touches ↑ et ↓, l'étudiant peut aussi la rendre plus pointue ou plus aplatie.

Quand l'étudiant a confirmé sa réponse, la réponse correcte s'affiche (sous forme de barre verticale, ici à 38 %) ainsi que le score (sous forme de barre horizontale (ici 12 points).



3.12 Conclusion

Les différents états de connaissances partielles mesurables à l'association d'un degré de certitude à une réponse autorisent un diagnostic cognitif plus subtil, via l'analyse spectrale (Jans et Leclercq, 1998) et les graphiques de réalisme (Leclercq, 1993). Cette amélioration de la sensibilité de l'outil d'évaluation contribue également à une mesure plus subtile de modifications intra-individuelles (par exemple entre un pré et un post-test). Ainsi, il est fréquent qu'après un débat, les réponses des participants n'évoluent pas, mais que leurs certitudes, elles, soient affectées.

Quatre conditions nous paraissent cependant indispensables.

- Parmi les consignes énumérées ci-dessus, seules celles qui sont numérotées de C3 à C10 peuvent constituer ce que Shufford *et al.* (1966)

appellent des « procédures admissibles de mesure des probabilités (subjectives) ».

- Le choix entre l'une d'elles dépend du but de la recherche ou de l'action pédagogique, des moyens disponibles, du degré de familiarité de l'apprenant et du formateur avec ce genre de pratique. Par exemple, il est inutile de permettre à l'apprenant de donner des degrés de certitude très détaillés (l'échelle continue des degrés de certitude) si le formateur n'a pas les moyens de les traiter et de donner une information appropriée en retour. Si un taux de certitude de 37,24% est traité comme s'il valait 30%, l'expression nuancée constitue une perte d'énergie.
- Il n'y a pas encore de règles et principes d'autoestimation de ses compétences qu'on puisse enseigner comme on a pu, par exemple, enseigner la « test wiseness » (Slakter *et al.*, 1970) car cette performance varie avec les contenus, les enjeux, etc. Par contre l'apprentissage de cette habileté métacognitive se fait par l'ajustement de comportements d'autoestimation après avoir été confronté aux conséquences de nos jugements, bref par conditionnement opérant. Gilles (1997) observe une augmentation significative de la moyenne des scores obtenus à l'indice de réalisme lors de trois entraînements à l'autoestimation de ses compétences chez des étudiants universitaires. L'entraînement de jeunes enfants ou des étudiants non entraînés devrait recourir à des consignes comportant peu de degrés de certitude et une procédure de notation limpide (qui puisse être appliquée à la main par l'étudiant lui-même). Si un expérimentateur veut mesurer de subtiles modifications dans les connaissances, il devra, bien sûr, utiliser un plus grand nombre de degrés de certitude.
- Dans ce contexte de familiarisation, des feedbacks métacognitifs (indices de réalisme) sont cruciaux pour permettre à l'apprenant d'améliorer l'autoestimation de ses compétences. Rapportés à des normes (Gilles, 1996), ces feedbacks lui permettront de se situer par rapport aux performances d'une population cible.

4. LA MÉTACOGNITION EN PERSPECTIVE

Linda ALLAL

Les contributions précédentes abordent l'autoévaluation en se référant — plus ou moins explicitement — au concept de la métacognition. L'autoévaluation comprend toujours une démarche de « retour sur soi » (Allal & Michel, 1993) qui est fortement apparentée aux processus identifiés dans les recherches sur la métacognition. Je propose ici de mettre en perspective certaines lignes de force et certaines dérives qui caractérisent les usages du concept de métacognition.

Rappelons que selon la définition de la métacognition proposée par Flavell en 1976, ce terme recouvre deux aspects du fonctionnement cognitif : les connaissances que le sujet possède de ses propres processus cognitifs et sa régulation active de ces processus afin d'atteindre un but. Les propositions de Flavell ont suscité un important développement de travaux théoriques et de recherches portant sur les différents types de métaconnaissances élaborées par le sujet et sur les régulations métacognitives intervenant dans l'apprentissage et dans la résolution de problèmes. Malgré quelques signes de déclin d'intérêt pour ce champ relevés il y a une dizaine d'années (Noël, Romainville & Wolf, 1995), on constate actuellement une forte expansion des publications relatives à la métacognition et à son rôle dans le domaine de l'éducation. On peut citer, entre autres, dans le contexte européen, la parution récente d'ouvrages collectifs sous la direction de Doudin, Matin et Albanese (1999) et de Grangeat et Meirieu (1997), la réédition actualisée de l'ouvrage de Noël (1997), la publication de numéros spéciaux des revues *Instructional Science* (1998) et *European Journal of the Psychology of Education* (1998), les articles parus dans de nombreuses revues scientifiques et pédagogiques.

4.1 Une dérive...

Dans bon nombre de publications destinées aux enseignants, on parle de la métacognition comme si l'on s'agissait d'un phénomène extérieur au processus de pensée de l'élève, comme si la métacognition était une activité ou outil pédagogique dont l'enseignant pouvait se servir pour stimuler l'apprentissage des élèves. Voici quelques exemples d'énoncés qui reflètent une telle perspective :

- pratiquer la métacognition avec les élèves pour leur apprendre à réfléchir (Barth, 1996) ;
- métacognition, une aide au travail de l'élève (Grangeat & Meirieu, 1997) ;
- faut-il initier les élèves, les équipes pédagogiques à la métacognition (Gather Thurler, 1997) ?

Ces énoncés donnent l'impression que l'intervention de l'enseignant est nécessaire pour permettre aux élèves de construire une démarche métacognitive ou d'accéder à la métacognition en tant qu'objet de savoir. On introduit ainsi une discontinuité entre le concept de métacognition et la signification du terme cognition. Si le mot « cognition » est substitué à « métacognition » dans les trois énoncés, on voit assez vite le décalage : il paraît en effet quelque peu absurde de dire qu'on « pratique » la cognition à l'école, que la cognition est une « aide » au travail de l'élève, que l'école va « initier » les élèves à la cognition.

Les textes fondateurs du champ de la métacognition (Flavell, 1976 ; Brown, 1987) postulent une continuité entre cognition et métacognition. La métacognition est un processus cognitif qui se caractérise par un objet particulier (le fonctionnement cognitif) et par un niveau élevé d'intentionnalité, de réflexion et de régulation active. La frontière entre cognition et métacognition n'est cependant pas facile à discerner (Allal & Saada-Robert, 1992 ; Noël, Romainville & Wolfs, 1995) précisément parce que ces processus se situent sur un continuum, au lieu d'être deux phénomènes disjoints.

La construction des compétences métacognitives est le fruit de l'ensemble des interactions de l'élève avec son environnement physique et social ; cette construction s'amorce même si l'enseignant n'a rien fait de particulier pour la stimuler. Cela étant dit, les interventions de l'enseignant peuvent avoir un impact important si elles créent des conditions qui favorisent la réflexion métacognitive et facilitent l'émergence des régulations métacognitives. Des pratiques d'enseignement et d'évaluation allant dans ce sens peuvent sans doute contribuer au développement de l'autonomie de l'élève et à l'acquisition de stratégies d'apprentissage plus efficaces (Grangeat, 1998). Il faut éviter toutefois de confondre le développement métacognitif de l'élève avec l'acquisition de conduites « scolairement correctes ». Certains comportements déviants par rapport aux normes scolaires s'appuient sur des compétences éminemment métacognitives, telle la capacité d'un élève de faire semblant d'écouter la leçon tout en pensant à autre chose, ou la stratégie de tricherie d'un élève qui reflète un raisonnement métacognitif plus élaboré que celui exigé pour la simple préparation à l'épreuve.

4.2 *Deux approches contrastées...*

Les travaux qui portent sur l'autoévaluation, ou qui traitent plus largement de l'implication de l'élève dans le processus d'évaluation formative, ont souvent recours aux éléments conceptuels mis en évidence par les recherches sur la métacognition (Allal, à paraître). Il convient de différencier deux approches dans ce domaine, selon l'aspect de la métacognition qui est exploité en priorité.

Dans une première approche, on cherche surtout à conceptualiser le rôle des métaconnaissances dans des situations autoévaluatives. Cette approche est illustrée par les travaux de Noël (1997) et de Wolfs (1996) qui utilisent des questionnaires ou entretiens pour étudier les jugements métacognitifs des élèves ou étudiants par rapport à leur capacité d'entreprendre ou de réussir une tâche donnée, ainsi que leurs estimations de la confiance qu'ils accordent à leur démarche. Ces recherches montrent que les jugements métacognitifs reflètent des métaconnaissances qui varient selon leur degré d'explicitation, d'analyse et de conceptualisation.

Une deuxième approche est axée davantage sur le rôle des régulations métacognitives dans des activités d'apprentissage. Certaines recherches tentent de préciser les opérations de régulation qui interviennent lorsque l'élève s'implique dans une démarche d'évaluation intégrée dans la réalisation de la tâche (Allal, 1993 ; Allal, Rouiller & Saada-Robert, 1995). Les travaux de Campanale (1997) offrent un cadre conceptuel pour l'analyse des situations d'autoévaluation en termes de trois mécanismes de régulation de complexité croissante :

- le recul : le sujet évalue un produit par rapport à un référentiel et introduit des rectifications de surface pour assurer la conformité du produit aux critères ;
- la distanciation : en évaluant son produit, le sujet explicite avec l'aide d'autrui les procédures mises en œuvre et parvient ainsi à une modification de ses démarches et à une révision approfondie du produit ;
- la décentration : à travers l'interaction avec autrui, le sujet élabore des modèles interprétatifs qui transforment ses représentations plus profondes et lui permettent d'envisager de nouveaux produits et procédures.

On peut relever une limitation associée à chacune de ces approches. Dans le premier cas, on risque de confondre la métacognition avec des jugements autoévaluatifs assez statiques sans prise sur l'action du sujet. Dans le second cas, on analyse les mécanismes de régulation sans forcément identifier avec précision les représentations cognitives et métacognitives qui déclenchent et guident ces mécanismes. Il y aurait sans doute une complémentarité à rechercher entre les deux approches pour mieux saisir les complexités des processus d'autoévaluation.

4.3 *Une perspective à développer...*

La majorité des travaux sur la métacognition sont centrés sur les facteurs intellectuels (cognitifs et métacognitifs) qui interviennent dans des tâches finalisées d'apprentissage ou de résolution de problèmes. Le recours à ces travaux dans le champ de l'autoévaluation a donc tendance à renforcer une vision strictement intellectuelle des démarches évaluatives de l'élève. Or les recherches récentes sur l'autorégulation des apprentissages en situation scolaire (Zimmerman, 1995) insistent de plus en plus sur la nécessité de relier les dimensions cognitive et métacognitive aux dimensions conative et motivationnelle. Plusieurs modèles postulent des relations interactives entre ces deux dimensions (Boekaerts, 1997 ; Hartman & Sternberg, 1993 ; McCombs & Marzano, 1990).

Les implications pour l'étude des processus d'autoévaluation sont de deux ordres :

- l'interprétation des démarches autoévaluatives d'un élève devrait tenir compte non seulement de ses compétences métacognitives mais aussi de ses stratégies motivationnelles (par exemple, sa capacité de gérer le « stress », sa volonté de persévérer face à des obstacles, ses attributions causales) ;
- les modes d'interaction sociale et les outils proposés comme support à l'autoévaluation devraient être conçus en tenant compte non seulement des stratégies métacognitives permettant de résoudre le problème mais aussi de la volonté des apprenants de se concerter ou de recourir aux outils à disposition.

L'autoévaluation fait ressortir les tensions inévitables entre « ce que je voudrais faire » et « ce que je suis en mesure de faire », entre ma volonté d'investir dans la tâche et les outils dont je dispose pour réussir.