

# Pour une évaluation des apprentissages scientifiquement fondée

**AVIS d'EXPERTS** en MESURE et ÉVALUATION des  
APPRENTISSAGES sur le PROJET de  
**POLITIQUE D'ÉVALUATION DES APPRENTISSAGES**

Serge P. Séguin, Ph.D.

Réjean Auger, Ph.D.

Jean-Yves Lancup, M.A.

Claudine Nézet-Séguin, M.A.

Carmen Parent, Ph.D.

Zeycan Yegin, M.A.



Faculté d'éducation  
Université du Québec à Montréal

# ***Pour une évaluation des apprentissages scientifiquement fondée***

## **SOMMAIRE**

- 1. Contexte de l'avis**
- 2. Tout n'est pas nouveau**
- 3. Un regard d'ensemble sur le développement scientifique de l'ÉVAP<sup>1</sup>**
- 4. Le discours conceptuel sur l'évaluation comme aide à l'apprentissage**
- 5. Les fonctions de l'ÉVAP**
- 6. Le discours sur la subjectivité**
- 7. Valeurs d'équité, qualités scientifiques et externalité**
- 8. Instruments de consignation**
- 9. Considérations sur la formation des enseignants<sup>2</sup> à l'ÉVAP**

---

<sup>1</sup> L'acronyme ÉVAP désigne l'évaluation des apprentissages.

<sup>2</sup> Dans ce document, le masculin est utilisé pour désigner toute personne, sans distinction de sexe.

## 1. Contexte de l'avis

En novembre 2000, le ministère de l'éducation du Québec (MÉQ) a soumis pour consultation un *Projet de politique d'évaluation des apprentissages* (ÉVAP). À l'Université du Québec à Montréal (UQÀM), milieu privilégié de formation initiale et continue d'enseignants, plusieurs instances, responsables à divers titres de cette formation, se sont prononcées dans des avis écrits qui ont été synthétisés dans l'avis officiel de la Faculté d'éducation de l'UQÀM.

L'actuel *Projet de politique d'ÉVAP* vise à renouveler la politique en vigueur depuis 1981 en s'articulant à la réforme de l'éducation actuellement désignée par l'expression *L'école, tout un programme*. Les documents officiels présentent la réforme comme s'appuyant sur le développement de théories scientifiques, notamment du cognitivisme et du socioconstructivisme, alors que toute préoccupation de s'appuyer sur des développements scientifiques éprouvés est absente du *Projet de politique d'ÉVAP*. À titre d'experts en ÉVAP, nous sommes donc interrogés par le *Projet de politique d'ÉVAP*. Nous sommes aussi interrogés par nos supérieurs, par nos collègues d'autres spécialités disciplinaires ou d'autres expertises pédagogiques et par nos étudiants, enseignants en formation. Les questions se résument en une seule: **quelle est notre position scientifique face à ce projet du MÉQ ?**

La politique d'ÉVAP, qu'il s'agisse de celle actuellement en vigueur ou du projet de nouvelle politique, décrit comment l'ÉVAP est orientée, encadrée et soutenue, de façon à permettre que les jugements portés sur les apprentissages supportent des décisions d'intervention ou de sanction pertinentes et équitables. Le présent texte prend la forme d'un avis d'experts en mesure et évaluation préoccupés de préciser les conditions de décisions pertinentes, fiables et équitables.

Pour exprimer notre position, nous ne reprenons pas, point par point, les éléments du projet de politique. Plutôt, nous retenons, en les regroupant thématiquement, les principaux aspects que le projet de politique néglige ou sous lesquels il semble aller à l'encontre de données scientifiquement éprouvées ou empiriquement démontrées. Ultimement, en effet, la politique d'ÉVAP, qui vise à énoncer les orientations et principes qui sous-tendent les pratiques d'ÉVAP en milieu scolaire, **devra assurer que l'évaluation des apprentissages soit scientifiquement fondée**. Il importe que les pratiques d'ÉVAP dépassent celles du sens commun de façon à assurer les valeurs de justice, d'égalité et d'**équité**, à fonder la **crédibilité** des titres ou diplômes décernés, ainsi que de la reconnaissance des compétences ou des acquis, sur les valeurs de transparence, de cohérence et de rigueur, notamment en appuyant tout jugement d'évaluation sur des informations pertinentes et suffisantes, valides et fidèles. Les regroupements retenus sont articulés autour de trois regards sur le développement scientifique de l'ÉVAP : une première vue d'ensemble, le discours conceptuel sur l'évaluation formative, l'objectivité et l'externalité particulièrement nécessaires en évaluation sommative. Nous y ajoutons des propos concernant les instruments de consignation et une réserve importante quant à l'utilisation de portfolios. Nous concluons cet avis sur quelques considérations relatives à la formation des enseignants en évaluation des apprentissages.

Notre **prise de position scientifique** intéressera sans doute le MÉQ ainsi que nos collègues avec qui nous partageons le mandat de former les enseignants à un large éventail de compétences pédagogiques. Mais elle s'adresse surtout aux étudiants en formation initiale et enseignants en formation continue que nous avons la responsabilité d'initier à l'ÉVAP.

## 2. Tout n'est pas nouveau

Dans les documents officiels relatifs à la réforme actuelle en éducation, le passé est souvent simplifié comme si rien ou presque rien n'avait été fait correctement il y a à peine quelques années. Par exemple, l'enseignement coopératif et la pédagogie par projet sont présentés comme des nouveautés alors qu'il s'agit de méthodes pédagogiques connues et utilisées. On sait aussi qu'elles ne sont pas toujours applicables, ni partout, ni pour tous les élèves, ni pour toutes les situations, ni pour toutes les compétences à développer. La réforme semble vouloir les imposer en toutes circonstances, niant ou contredisant un large pan des expériences et des connaissances acquises en pédagogie. Il est largement connu, et depuis longtemps appuyé par la recherche dans des études quantitatives ou qualitatives, qu'aucune approche didactique ou pédagogique unique n'a jamais pu être démontrée supérieure à toutes les autres approches en termes d'apprentissages plus nombreux ou de meilleure qualité. Il en est de même de toute théorie, méthode ou technique d'apprentissage ou d'enseignement qui tire son utilité ou son efficacité relative de divers agencements de caractéristiques contextuelles (enseignant, élèves, groupe, organisation, contenu, objectif).

De la même façon, on présente l'ÉVAP d'avant 1980 comme restreinte à la seule évaluation des connaissances. Cette prétention que l'élève n'était évalué que sur l'acquisition de connaissances, comme si l'école n'avait pas développé d'autres habiletés intellectuelles ou qu'on s'était privé de les mesurer, renie ou contredit certaines contributions scientifiques importantes dans le développement de l'ÉVAP. A-t-on oublié que, dès le début des années 1950, sous la conduite de pédagogues, de vastes travaux interdisciplinaires et internationaux ont été entrepris et allaient produire les taxonomies des comportements cognitifs développés à l'école ? Le premier niveau, le plus simple, est appelé CONNAISSANCE et ne fait appel qu'aux processus mémoriels de rappel ou de reconnaissance. Tout ce qui dépasse ce niveau premier devient du SAVOIR-FAIRE, des HABILITÉS, selon divers

niveaux gradués de signification, d'application ou de traitement des informations. Et si ces taxonomies ont alors été construites, c'est que déjà, dans les écoles, on cherchait à former les élèves et non seulement à les informer. Reconnaissons que, si certains déploraient que l'accent soit mis sur la mémoire, les éducateurs d'alors partageaient l'adage « une tête bien faite vaut mieux qu'une tête bien pleine ». Il paraît d'ailleurs simpliste, sinon absurde, de prétendre que les bouleversements d'après-guerre qui ont entraîné évolution scientifique et développements technologiques et de société n'aient résulté que d'une simple mémorisation de connaissances. Mais il fallait s'entendre sur les divers niveaux d'habiletés, sur ce que l'on voulait dire aux élèves, et aussi aux parents et aux autres enseignants, quand on affirmait vouloir qu'ils comprennent, appliquent, analysent, résument, critiquent, évaluent...D'où les taxonomies. C'est ainsi qu'au Québec, entre les années 1960 et 1980, les apprentissages visés par les programmes, et mesurés par les examens officiels, ont été graduellement formulés selon des niveaux d'habiletés définis qui ne se limitaient pas aux connaissances.

En 1981, le MÉQ se dote d'une *Politique générale d'évaluation pédagogique* visant à la fois à situer les pratiques d'ÉVAP au cœur de l'activité éducative et à les appuyer sur des valeurs de justice et d'équité. Depuis lors, les développements en ÉVAP au Québec s'adaptent à de multiples changements dont notamment : 1) la structure des programmes d'études, passant des programmes cadres aux programmes par objectifs; 2) les types de questions, offrant une combinaison de réponses choisies et de réponses construites; 3) l'interprétation de la mesure, donnant une large place à l'interprétation critériée; 4) le bulletin scolaire, qui s'enrichit d'un nouveau mode, le descriptif.

Depuis 20 ans, le MÉQ, les commissions scolaires et les universités ont consenti des efforts importants en ÉVAP pour la formation d'un personnel compétent, dont un grand nombre d'enseignants et de conseillers pédagogiques, pour le développement d'instruments conformes aux programmes et considérés comme possédant des standards de qualité parmi les plus élevés au monde. Des résultats

tout aussi importants ont été obtenus, les élèves québécois se comparant avantageusement à ceux de nombreux autres pays.

Le développement le plus significatif a été celui de la « définition du domaine ». Au début des années 1980, l'ÉVAP au Québec était mise au défi de résoudre les trois problèmes suivants : A) Comment mesurer et évaluer des habiletés ? B) Comment préparer des épreuves de synthèse ? C) Comment rendre l'interprétation critériée significative ? La « définition du domaine » résulte d'une stratégie mise au point au Québec pour répondre à ces trois questions. Elle désigne l'ensemble des éléments essentiels et représentatifs d'un programme d'études et permet d'assurer la cohérence entre les orientations ou l'esprit du programme et les conditions dans lesquelles la mesure des apprentissages doit être effectuée. Elle prend en compte tous les aspects du programme, sert de point de départ à la préparation d'une épreuve sommative de synthèse et de document de référence pour l'interprétation critériée des résultats.

La réforme actuelle semble ignorer tout ce qui a été fait au Québec en ÉVAP, acquis d'expérience et développement scientifique, en imposant le passage de l'évaluation d'objectifs d'apprentissage à l'évaluation de compétences. Les objectifs d'apprentissage énoncent des connaissances et des habiletés en termes de comportements observables qui constituent des manifestations de ces connaissances et habiletés. Le besoin d'évaluer des compétences différemment ou autrement que ce qui se fait actuellement en ÉVAP n'est pas démontré. L'articulation des compétences visées aux objectifs d'apprentissage des programmes d'études ne nous semble pas comprise.

Il existe de nombreuses définitions du concept de compétence. Il est sans doute admissible de proposer qu'elle soit un comportement global par lequel un ensemble de connaissances, d'habiletés et d'attitudes sont mobilisées pour agir dans un contexte de tâche ou d'activité selon des exigences de performance attendue exprimées par des conventions, critères, indicateurs, normes, règles ou standards. Les instruments mis au point pour l'ÉVAP, préparés dans le contexte des programmes d'avant la réforme, devaient faire appel à des problèmes ou à des mises en situation réalistes de la vie

courante; ils devaient mesurer des connaissances et des habiletés décrites dans les objectifs terminaux d'apprentissage, conformément à un cadre appelé définition de domaine. Comment prétendre que cela ne mesurerait pas des compétences ? Passons-nous réellement à une autre ère ? Ou n'est-ce pas que d'utiliser un concept encore vaguement défini de compétence (rendu vague parce qu'on le dénature en l'appliquant à la formation générale des niveaux primaire et secondaire; ou encore parce que l'on en nie les composantes essentielles de connaissances et d'habiletés) viserait à faire table rase de tout ce qui a été fait, réalisé, acquis d'expérience, avant la venue des réformistes ? Évidemment, au plan scientifique, l'expression ÉVAP demeure, que les apprentissages s'appellent connaissances, savoirs, savoir-faire, habiletés ou même compétences.

### **3. Un regard d'ensemble sur le développement scientifique de l'ÉVAP**

Mettre un frein à la subjectivité, cause d'injustice ou d'inéquité, de complaisance ou de discrimination non fondée, telle a été la raison d'être de l'avènement d'une démarche de distanciation, de démonstration rigoureuse, d'objectivité scientifique. On notera au passage que ce mouvement, amorcé au début du XX<sup>e</sup> siècle, a marqué le développement de toutes les mesures et évaluations de comportements humains, dans tous les domaines des sciences et des interventions humaines et sociales, et non seulement en éducation et en ÉVAP. Une interprétation subjective est susceptible de varier en fonction de la personnalité de chacun des interprètes. L'interprétation objective fait abstraction de préférences et est conforme à la réalité qu'elle décrit avec le plus d'exactitude possible. L'objectivité appuie l'équité : un jugement équitable est impartial, ne favorise pas l'un aux dépens de l'autre ni l'un par rapport à l'autre.

La solution au problème de la « subjectivité indésirable », considérée comme une erreur, est passée par le développement scientifique de mesures de comportements humains le plus possible exemptes d'erreurs. Ce développement a permis d'établir les limites de la



mesure (tout n'est pas mesurable) et d'étudier les conditions de validité et fidélité.

Une mesure est dite valide lorsque l'élément réellement mesuré est conforme à sa représentation. La validité résulte de l'adéquation ou de la conformité à ce dont il s'agit : un test est dit valide pour mesurer tel objet lorsque l'on peut démontrer qu'il mesure réellement cet objet. Par exemple, tel test mesure telle compétence parce que chacune des questions du test permet de mesurer tels ou tels aspects ou composantes de cette compétence. Mais en ÉVAP, la validité est aussi question de pertinence, à un double point de vue. La pertinence, c'est d'abord la congruence ou relation étroite entre les apprentissages évalués et les apprentissages visés dans le programme officiel concerné ou dans le programme tel que transcrit par l'enseignant dans le vécu de sa classe; la pertinence, c'est aussi la cohérence entre les apprentissages évalués et la finalité de la démarche d'évaluation. Ainsi, un examen peut être adéquat pour bien mesurer tels apprentissages, mais il est non pertinent si ces apprentissages ne sont pas visés ou n'appartiennent pas au programme concerné par l'ÉVAP. Ou encore, une procédure valide pour bien évaluer des apprentissages prescrits par le programme concerné, peut être pertinente ou non selon le but visé ou la fonction de l'ÉVAP; par exemple, des informations, ainsi que les procédures pour les recueillir, qui sont adéquates et pertinentes pour des décisions relevant de la fonction formative, peuvent ne pas être pertinentes pour des décisions relevant de la fonction sommative. C'est la validité qui permet de garantir que les apprentissages évalués sont pertinents aux décisions à prendre et de démontrer que l'ÉVAP porte bien sur ces apprentissages.

Quant à la fidélité, elle désigne tantôt l'exactitude ou la précision de la mesure, tantôt la constance de l'objet observé ou de l'observateur, tantôt l'homogénéité des composantes de l'instrument ou l'accord entre les évaluateurs. Une mesure est dite fidèle s'il est démontré qu'on obtiendrait le même résultat si l'on répétait la mesure du même objet dans des situations semblables; une observation est fidèle si elle est la même quel que soit l'observateur; un jugement est fidèle lorsque chaque évaluateur note de la même façon des

comportements identiques, ou lorsque plusieurs évaluateurs ont entre eux le même jugement sur un même comportement observé.

En science de l'ÉVAP, les qualités de validité et de fidélité doivent s'intégrer à toutes les étapes de la démarche. Validité et fidélité caractérisent donc les informations initiales (mesures et observations), l'interprétation de ces informations, les jugements qui s'y appuient et les décisions finales. Validité (conformité et pertinence) et fidélité (précision ou constance) ne peuvent se soustraire à des conditions d'objectivité. À leur tour, validité et fidélité constituent les conditions scientifiques d'une ÉVAP crédible (vraisemblable) supportant des décisions fiables (dignes de confiance) et équitables.

Au sens général, évaluer, c'est déterminer la valeur, le prix ou l'importance de quelque chose. La présence d'une mesure dans l'évaluation paraît tellement évidente que, pour *Le Petit Larousse* (1998), l'évaluation en enseignement désigne la « mesure à l'aide de critères déterminés des acquis d'un élève... ». Dans le domaine de l'ÉVAP, cette définition de sens commun simplifie la réalité à plus d'un titre : 1) comme les acquis sont des comportements humains, ils ne peuvent être mesurés directement mais seulement inférés, donc indirectement mesurés, à partir de manifestations observées; 2) les acquis, visés ou réalisés, sont tellement nombreux qu'ils ne peuvent tous être spécifiquement mesurés; il devient alors nécessaire de les regrouper en échantillons de manifestations, représentatifs de l'ensemble des manifestations possibles des apprentissages (ce que l'on peut désigner par le terme d'indicateurs); 3) des séries de mesures seules ne sauraient suffire pour déterminer la valeur des apprentissages : il faut aussi des jugements. Mais un jugement, c'est une estimation, une appréciation, contenant nécessairement une part de subjectivité. Comment faire pour que cette subjectivité ne soit pas inéquitable, qu'elle ne masque pas la réelle valeur de ce que l'élève a vraiment appris, voilà le défi qui a inspiré le développement scientifique de la mesure et de l'évaluation.

Dans la deuxième édition de son *Dictionnaire actuel de l'éducation* (Montréal : Guérin, 1993), l'auteur Renald Legendre, professeur-

chercheur et didacticien, consacre une vingtaine de pages aux concepts et définitions de l'évaluation, de divers types de mesures et d'interprétation, du jugement qui fait le lien entre l'interprétation des mesures et les décisions pédagogiques ou administratives en vue desquelles l'évaluation existe. Ainsi, l'ÉVAP, c'est l'appréciation du niveau d'apprentissage atteint par l'élève, c'est-à-dire de ses connaissances et de ses habiletés ou compétences en regard de celles visées par un programme d'études, en vue « de juger d'un cheminement antérieur et de prendre les meilleures décisions quant à un cheminement ultérieur ». Cette appréciation fait appel à un jugement fondé sur un ensemble d'informations qualitatives et quantitatives systématiquement recueillies, traitées et interprétées.

Le jugement, qui occupe une place centrale dans la démarche d'ÉVAP, permet d'appuyer des décisions, à la condition qu'il soit fondé sur des informations pertinentes et suffisantes. Ces informations sont des quantités ou des énoncés de qualité servant d'indicateurs des apprentissages réalisés. Chaque information prise en compte dans le jugement doit être valide, pertinente; et les informations appuyant un jugement doivent être suffisantes, qualitativement et quantitativement, dans leur ensemble, pour entraîner un jugement équitable sans qu'il soit nécessaire de recueillir d'autres informations.

En ce sens, donner toute la place qui revient au jugement de l'enseignant est un objectif louable, voire une décision politique acceptable, dans la mesure où l'enseignant exerce un jugement professionnel, c'est-à-dire en tenant compte de résultats de mesures aussi bien que de considérations complémentaires pertinentes relatives à l'élève lui-même, au programme d'études, à diverses circonstances. Les mesures (notes ou scores exprimant des quantités d'apprentissage mesurées par des examens, des contrôles ou des tests) ne sauraient être ignorées ou mises de côté : elles sont les assises objectives des jugements d'évaluation et des décisions pédagogiques ou administratives qui en découlent.

En résumé, l'ÉVAP est un jugement (ou un ensemble de jugements) porté sur les apprentissages et qui les met en relief. Ce jugement est fondé sur des informations rigoureusement recueillies sous forme

de mesures ou d'indicateurs dont les quantités sont interprétées au regard de critères ou de normes, en vue de prendre des décisions en termes de sanction ou d'actions à entreprendre. On admet que le jugement soit subjectif, mais il doit être conscient et argumenté, éclairé par les résultats, à défaut de quoi il ne sera qu'une simple opinion et risquera, si on y prend appui, d'entraîner des décisions inéquitables et non fondées.

#### **4. Le discours conceptuel sur l'évaluation comme aide à l'apprentissage**

Avant 1970, l'évaluation **quotidienne** existait en classe, le plus souvent de façon non formelle, ni systématisée, ni universellement instrumentée : elle faisait partie de l'arsenal pédagogique courant de tout enseignant, mais on ne savait pas encore qu'on allait l'appeler formative. L'évaluation telle qu'on l'entendait officiellement (on ne savait pas encore qu'on allait la désigner par le vocable « sommative »), c'était l'évaluation dont on parlait, sur laquelle on écrivait, que l'on étudiait scientifiquement ou systématiquement. Elle répondait aux objectifs des administrations scolaires et politiques d'alors qui cherchaient à conserver des traces et à rendre compte aux parents des apprentissages réalisés par leurs enfants. Cette évaluation sommative occupait tellement de place qu'elle occultait celle qui nourrissait le vécu pédagogique quotidien de la classe.

Au tournant des années 1970, sous l'impulsion à la fois des milieux d'éducation et des mouvements sociaux, on reconnaissait et donnait son importance à l'ÉVAP menée par chaque enseignant dans sa classe pour appuyer ses interventions pédagogiques là où il y a le plus de besoins et pour informer les élèves par rétroactions sans attendre les examens trimestriels ou autres. Les spécialistes de la mesure et de l'évaluation ont alors commencé à mettre au point des modèles et des méthodes pour l'interprétation critériée de mesures et pour la construction de tests et d'examens destinés à mieux encadrer les démarches d'évaluation formative. Car une évaluation formative, pour atteindre pleinement l'efficacité qu'on lui souhaite, nécessite des instruments de mesure qui garantissent aussi la

pertinence et la fiabilité des décisions pédagogiques qui en découlent.

Actuellement, en ce début de XXI<sup>e</sup> siècle, c'est le quotidien de la classe que les documents officiels veulent mettre en évidence, au point de refiler à l'enseignant la responsabilité de répondre aux exigences sociales des parents et des organismes qui demandent au système scolaire de leur dire ce que valent les apprentissages de chaque élève. Le discours actuel sur l'évaluation en classe ignore pourtant presque tout des gains des sciences de la mesure de comportements humains, en particulier d'apprentissages scolaires; il ignore le besoin de rigueur scientifique pour assurer pertinence et fiabilité des décisions et des instruments qui les supportent. Comment le discours a-t-il pu dévier de la sorte ?

D'abord, on a souvent, à tort, confondu l'interprétation critériée (comparant une mesure à un seuil pré-établi) et l'évaluation formative (qui a pour fonction l'amélioration des apprentissages), notamment en utilisant des expressions telles « évaluation critériée » ou encore « examens ou tests formatifs ». Cette confusion vient sans doute d'une conjoncture historique : pendant que la fonction formative de l'ÉVAP se démarquait des habituelles fonctions sommatives, l'interprétation critériée marquait son opposition à l'habituelle interprétation normative (comparant chaque élève aux autres de sa classe ou de son niveau). Aussi, la fonction formative est bien servie par l'interprétation critériée alors qu'elle n'a que faire des comparaisons normatives. Cependant, l'interprétation critériée est tout aussi utile à l'évaluation sommative. Il est possible d'interpréter, dans le bulletin scolaire, les notes de l'élève en les comparant à des seuils ou à des critères. De tels seuils sont toutefois difficiles à établir de façon à ce qu'ils soient bien ajustés à la fois aux attentes absolues des programmes et aux attentes relatives fixées par voie empirique pour des élèves ayant des caractéristiques données.

Ensuite, on a voulu, à tort, faire correspondre l'interprétation critériée non plus à un seuil pré-établi pour tous mais à des seuils différenciés selon les capacités individuelles. En interprétation critériée, la mesure observée et la mesure-critère doivent être de même nature,

ce qui n'est pas le cas des capacités individuelles et des apprentissages visés par les programmes d'études. Cette approche ne peut donc pas garantir l'équité des jugements d'évaluation, ni en fonction des élèves entre eux, ni au regard des exigences des programmes. Par conséquent, comparer un élève à lui-même demeure un discours que l'observation scientifique actuelle ne permet pas de fonder; ce discours apparaît démagogique et risque des abus de subjectivité.

Une autre dérive a eu un impact important : on a proposé des bulletins critériés mais avec un discours formatif, appelés « bulletins formatifs ». Depuis 1980, plusieurs se sont acharnés à caractériser les savoirs et les savoir-faire visés par les programmes de façon à justifier leur pertinence en évaluation soit formative, soit sommative. C'était prendre la lunette par le mauvais bout. Ce qui distingue les fonctions sommative et formative de l'ÉVAP, ce sont les finalités, c'est-à-dire la nature des décisions qui en découlent. En ce sens, **les bulletins ont en essence des fonctions éminemment sommatives**, répondant aux nécessités d'informer périodiquement les parents et les administrations scolaires sur la situation de chacun de élèves et d'en conserver des traces, des indications globales ou de synthèse, à des fins de sanction. Ces finalités ont peu à voir avec le contexte formatif de la gestion des apprentissages : pour guider le mieux possible ses interventions et les apprentissages des élèves, l'enseignant doit pouvoir observer le mieux possible la panoplie des apprentissages réalisés; en donner le détail exige d'en faire une longue liste en termes de savoirs et de savoir-faire (parfois aussi de savoir-être), de connaissances et d'habiletés ou de compétences.

Au tournant du millénaire, une autre dérive s'annonce lorsque l'expression «l'évaluation au service de la réussite» prend valeur de slogan. L'expression «au cœur de l'apprentissage», tout comme l'expression «au service de l'apprentissage», signifient «ayant pour but d'améliorer les apprentissages». Si l'expression «au service de la réussite» **peut** être synonyme de «au service de l'apprentissage», ne peut-elle pas signifier également «se servir de l'évaluation pour faire réussir» en passant sous silence ce qui n'est pas réussi ou ce qui l'est difficilement ? Plusieurs réactions publiques rendent

crédible cette dernière possibilité, dans le contexte de discours ministériels actuels fréquemment construits autour d'objectifs visant à réduire les taux d'échec ou à augmenter les taux de diplomation, de tels objectifs de performance étant déjà inscrits, aux niveaux postsecondaires, dans des contrats en échange de subventions. Pour éviter que cette expression ait valeur de slogan invitant à faire réussir tout le monde, il suffit de ne pas l'utiliser, d'autant plus que sa part positive est déjà incluse dans l'expression d'une évaluation **au coeur** ou **au service** de l'apprentissage.

Centrer l'ÉVAP à la fois sur la réussite des élèves et sur le jugement enseignant peut avoir un autre effet pervers. En effet, pour plusieurs analystes de la réforme, et nous en sommes, ce projet semble parfois découler d'une prémisse que tout enfant peut tout apprendre, pourvu qu'il en ait le temps et qu'il utilise les méthodes d'apprentissage qui lui sont les plus appropriées. Comme le système uniformise la durée, il suffirait de choisir les meilleures méthodes. Qui choisit les méthodes ? L'enseignant. Qui aide l'élève à découvrir ses processus d'apprentissages ? L'enseignant. Le voilà maintenant chargé d'une autre responsabilité. C'est lui qui déciderait de la réussite ou de l'échec des apprentissages. En cas d'échec, il n'aurait que lui-même à blâmer puisqu'il en serait doublement la cause : comme enseignant, puis comme évaluateur. Ainsi considéré, l'enseignant devient responsable de la valeur de justice sous laquelle on veut placer l'ÉVAP.

Selon une argumentation de sens commun, on pense éviter le piège d'une « subjectivité indésirable » en s'appuyant sur un jugement établi en petit groupe, par exemple par une équipe-cycle plutôt que par l'enseignant. Mais cette modalité risque, au contraire, de porter un jugement qui n'aurait de la validité que l'apparence et, en conséquence, de rendre l'exercice du jugement moins transparent. En ce cas, la validité apparente, c'est-à-dire l'impression qu'il y a validité parce que plusieurs individus se sont exprimés pour produire un jugement commun (pas toujours unanime), sert d'écran à la validité objective qui pourrait être absente ou présente. La validité objective appartient à la décision argumentée, supportée par des observations empiriquement cohérentes entre elles, pertinentes par rapport aux buts visés, suffisantes pour faire ressortir le caractère de

constance ou de récurrence et de fiabilité, suffisamment complètes dans les manifestations de la compétence visée pour en généraliser la qualité et la portée sur les apprentissages. Le jugement en petit groupe ne peut donc pas faire abstraction de cette démonstration sans risquer l'illusion d'une apparente validité qui masquerait des décisions inéquitables ou non fondées.

La prémisse que tout enfant peut tout apprendre semble aussi faire peu de place à la responsabilité qui incombe à l'élève lui-même en regard de sa propre réussite. Le seul endroit où on ouvre la porte à une affirmation de la responsabilité de l'élève, c'est lorsque l'on propose que l'élève prenne une part active dans l'évaluation de ses apprentissages pour qu'il reconnaisse ses forces et faiblesses, mais surtout, dit-on, pour qu'il découvre comment il apprend. Cependant, faire participer l'élève à son évaluation risque d'avoir un effet pervers en évaluation sommative. Théories et études convergent pour démontrer que la fonction formative de l'ÉVAP (aide à l'apprentissage) est mieux remplie lorsque l'élève participe à son évaluation, lorsque l'on fait appel à l'autoévaluation ou à l'évaluation par des pairs; cela aide aux échanges pédagogiques entre élèves et enseignant. Mais vouloir que l'évaluation par les pairs ou l'autoévaluation constituent des composantes d'une évaluation sommative contredit les études qui démontrent que, en contexte sommatif, l'autoévaluation et la coévaluation par des pairs font appel aux perceptions, au sentiment de plaisir et au jeu des relations sociales, bien davantage qu'aux faits observés.



## 5. Les fonctions de l'ÉVAP

Les politiques, comme les discours officiels, font surtout grand cas de la fonction formative, qui concerne l'enseignant dans sa classe et qui, au premier chef, supporte la relation pédagogique en vue de l'apprentissage. Mais une politique d'ÉVAP a pour but d'orienter, d'encadrer et de soutenir l'ÉVAP; elle ne peut donc se permettre de passer trop rapidement sur les fonctions sommatives de l'évaluation qui ont par ailleurs le plus d'impact sur l'accès aux études ultérieures, sur la carrière, sur l'insertion professionnelle et sociale. Alors qu'une seule fonction de l'ÉVAP est formative, trois sont éminemment sommatives et comportent des exigences d'uniformité : la reconnaissance de compétences, la reconnaissance d'acquis, la certification.

En contexte formatif, il importe que les jugements rendus s'appuient sur des observations et des mesures qui soient le plus objectives possibles. Cependant, le jugement de l'enseignant, dans ce contexte de la classe, même biaisé ou d'une subjectivité réactive, peut sembler admissible tant qu'il est au service de l'amélioration des apprentissages. Mais en contexte sommatif, le jugement est plus définitif, il est consigné, conservé, communiqué et utilisé en-dehors de la classe pour diverses décisions : autant de motifs pour que tout jugement sommatif soit appuyé sur des observations et des mesures objectives, valides et fidèles, qui en assurent la pertinence et l'utilité.

Il existe des différences importantes entre les rôles de l'enseignant d'aider les élèves à apprendre et de rendre compte; entre celui de juger de la qualité et de la quantité des apprentissages réalisés par les élèves et celui de statuer sur la promotion, la diplomation, l'accès à des études ultérieures ou supérieures, l'insertion sociale et professionnelle, le développement de carrière de chacun des élèves. Cela milite en faveur que l'évaluation sommative s'appuie sur une forme de référence externe qui garantit une base commune aux jugements d'évaluation, donc que l'évaluation sommative ne repose pas uniquement sur les épaules de l'enseignant.

De tout temps, l'enseignant a eu mission d'aide à l'apprentissage, de soutien, de facilitateur. L'apparition de l'évaluation formative comme entité conceptuelle ne niait pas les pratiques antérieures des enseignants; au contraire, elle reconnaissait une réalité et lui donnait un nom. Puis cette pratique est devenue un objet d'étude et de formation et elle a été considérée comme partie essentielle d'un acte pédagogique. L'évaluation formative s'arrête à la porte de la classe; elle se déroule entre l'enseignant et ses élèves. Vouloir imposer ou régir ce que l'enseignant doit faire en classe et comment il doit le faire pour le vécu et la gestion de ce qui se passe dans la classe apparaît contredire le discours même sur la professionnalisation de l'enseignant. L'ÉVAP qui exige d'être rigoureusement étudiée, contrôlée, décrite, encadrée, c'est l'évaluation sommative, celle qui rend compte de, celle qui est conservée et utilisée en-dehors de la classe.

## **6. Le discours sur la subjectivité**

L'histoire de l'éducation et la documentation scientifique sont truffées d'exemples où la subjectivité en ÉVAP conduit à l'inéquité, par abus ou par ignorance, toujours par croyance. Comme nous l'avons rappelé au début de la section 3, la subjectivité trônait en ÉVAP avant que, il n'y a pas encore un siècle, par recherche constante de rigueur et d'objectivité, la science fournisse à l'ÉVAP des moyens de plus en plus éprouvés de freiner cette subjectivité et les inéquités qu'elle pouvait engendrer. Que penser alors des discours actuels qui présentent la subjectivité comme inévitable, centrale, voire souhaitable, par rapport à une objectivité présentée comme purement formelle sinon fausse ?

La documentation scientifique en ÉVAP a démontré à plusieurs reprises que tout pourcentage, incluant le 60 %, n'a pas la même signification pour tous. Il en va de même pour l'interprétation des échelles de notes ou de jugements qui varie selon les enseignants lorsqu'ils observent les mêmes comportements d'élèves. Ces faits justifient le processus de modération des notes. De nombreuses études sur l'utilisation d'échelles descriptives, dites qualitatives, sur

lesquelles sont ordonnés des descripteurs de caractéristiques ou d'indicateurs de performance (de telles échelles sont fréquemment utilisées en langues, notamment pour la correction de productions écrites), ont démontré des écarts importants entre les correcteurs quant aux catégories utilisées pour noter ou qualifier les mêmes performances observées. Il est scientifiquement reconnu que, dans les meilleures conditions d'entraînement de correcteurs pour l'utilisation de telles échelles d'évaluation des réponses élaborées, les notes ou les catégories attribuées correspondent au mieux aux deux-tiers (en moyenne à la moitié) à la réalité des performances elles-mêmes; cette part descend à environ le quart lorsque les correcteurs ne sont pas entraînés. Dit autrement, lorsque les performances des élèves sont évaluées de cette façon, les différences entre les élèves sont dues pour moins de la moitié à leurs propres performances, et pour plus de la moitié à la subjectivité des correcteurs.

C'est donc manquer de réalisme que de supposer que l'objectivité n'est pas requise pour assurer pertinence et fiabilité des informations. Les informations non pertinentes ou non fiables, même en très grand nombre, seront toujours insuffisantes pour appuyer un jugement d'évaluation utile.

Le *Projet de politique d'ÉVAP* rapporte la citation suivante de Philippe Perrenoud (1998 : *Construire des compétences dès l'école*, Paris : ESF, p. 87) pour appuyer le jugement subjectif en niant toute recherche d'objectivité pour appuyer le jugement d'évaluation :

*Renoncer à standardiser l'évaluation, à s'abriter derrière une équité purement formelle, exiger et s'accorder pour soi-même la confiance voulue pour établir un bilan de compétences fondé sur un jugement expert plutôt que sur un barème.*

Clarifions dès maintenant que, selon la science de l'ÉVAP, l'équité ne réside pas dans la forme de ce qui est mesuré mais repose sur ce qu'un test mesure vraiment : les situations qu'il demande de résoudre sont démontrées correspondre à des apprentissages souhaités et considérés importants; les règles de correction assurent

le plus possible que toute performance reçoive la même cotation et la même interprétation quel que soit le correcteur.

L'équité, tout comme l'ÉVAP, peut être formelle par opposition à une équité non formelle ou informelle. En ÉVAP, l'équité est formelle lorsqu'elle est rigoureusement planifiée par des indicateurs et des instruments qui permettent de l'opérationnaliser et de la démontrer. Laisser la valeur d'équité au jugement particulier et subjectif, même professionnel, d'un seul enseignant, c'est la rendre informelle, non démontrable. Un jugement expert, pour être équitable, doit être conforme à un barème objectif sur lequel il s'appuie, c'est-à-dire un barème scientifiquement construit ou dont la signification est partagée par un ensemble d'experts. La pertinence, la fiabilité et l'équité doivent pouvoir être démontrées indépendamment du fait de connaître ou non les auteurs du jugement; c'est pourquoi ces qualités sont généralement supérieures en ÉVAP formelle qu'en ÉVAP informelle, offrant ainsi des conclusions plus certaines, épurées d'intuitions, de jugements personnels ou de réputations professionnelles.

La citation ci-devant ne témoigne donc pas d'une profonde connaissance ni d'une compréhension du développement scientifique de l'ÉVAP. Elle fait fi des expériences québécoises acquises en testing à des fins de sanction d'études ou de comparaisons internationales. Si l'auteur était lui-même enseignant dans une école québécoise, son jugement expert serait sans doute teinté par ses interprétations du programme, ses choix de méthodes et d'instruments, ses attentes personnelles, ses expériences passées et ses comparaisons entre sa classe actuelle et les autres classes, au Québec ou ailleurs dans le monde. Le reconnaît-il lui-même lorsqu'il écrit dans la revue *Mesure et évaluation en éducation* (Vol. 19, no 2, 1996, p. 82): « pour être sûr que les élèves ont progressé, il faut, tôt ou tard, mesurer les apprentissages » ?

Nous pensons qu'il faut développer la confiance en soi requise pour porter un jugement professionnel. Mais enfermer l'avenir et le cheminement de l'élève dans cette subjectivité apparaît comme une forme d'abus d'autorité et une négation de ce qui fonde l'équité. L'équité ne peut être fondée sur la subjectivité.

## 7. Valeurs d'équité, qualités scientifiques et externalité

Devant la nécessité que l'ÉVAP soit marquée aux valeurs de l'équité, la science de l'ÉVAP considère les valeurs de transparence (ce qui est attendu est clairement exprimé à l'élève et ce qui est évalué l'est tout autant), de cohérence (rapport étroit entre ce qui est attendu et ce qui est mesuré) et de rigueur (évaluation conduite avec exactitude et précision), comme des valeurs instrumentales pour assurer les valeurs fondamentales de justice, d'égalité et d'équité. Ainsi, dans un mouvement d'unification regroupant des représentants de ministères d'éducation et d'organismes professionnels, de commissions ou de conseils scolaires et d'administrateurs, de conseillers et de psychologues, d'enseignants et de chercheurs en éducation de tout le pays, un Comité consultatif mixte a produit en 1993 un recueil de *Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada* (University of Alberta : Center for Research in Applied Measurement and Evaluation, 22 p.). Ce recueil énonce les lignes directrices et les actions qui doivent servir d'appui à l'exercice du jugement professionnel en vue d'une évaluation juste et équitable des élèves. La pratique évaluative y est définie comme la cueillette et l'interprétation d'informations dans le but d'informer l'élève (ou ses parents) sur son apprentissage ou d'informer les personnes qui prennent des décisions d'intervention ou de sanction concernant les élèves.

On y observe que l'équité ne se limite pas à la pertinence et à la suffisance. Toutes les exigences des qualités dites scientifiques participent à assurer l'équité aux diverses étapes d'un processus d'évaluation : élaboration ou sélection de méthodes, cueillette d'informations, interprétation, jugement, notation, communication, conclusion. La science de l'ÉVAP a opérationnalisé les valeurs instrumentales recherchées selon diverses appellations et façons de faire. La transparence concerne la clarté des intentions et de la formulation des manifestations qui seront observées; à ce propos, des compétences formulées d'une façon incompréhensible

trahiraient cette qualité nécessaire et constitueraient donc un recul par rapport à des objectifs terminaux formulés en termes de comportements observables attendus. La cohérence définit la qualité scientifique dite de validité ou de pertinence. La rigueur, selon les termes d'exactitude et de précision qui la définissent, correspond à la qualité scientifique de fidélité.

Dire que le jugement d'évaluation doit reposer sur des informations pertinentes et suffisantes, c'est trop peu dire lorsqu'on passe sous silence que, pour obtenir de telles informations, il faut des instruments de collecte qui soient objectifs et valides, utilisés à plus d'une occasion et dans des contextes variés.

Les *Principes d'équité* ... sont divisés selon les deux grandes finalités fonctionnelles de l'ÉVAP : la première section regroupe les principes et directives des ÉVAP menées par les enseignants, en vue d'informer les élèves sur leurs apprentissages; la deuxième section regroupe les principes et directives des ÉVAP élaborées à l'extérieur de l'école en vue de leur utilisation par divers décideurs. Dans chacune des deux parties, plus d'une trentaine de lignes directrices sont énoncées sous cinq grandes orientations.

Quelles que soient les méthodes d'enseignement (coopération, pédagogie par projet ou autre) ou d'apprentissage (par déduction ou induction, par problème ou autrement), quel que soit le niveau ou le cheminement, que l'on soit ou non en accord avec le mouvement socio-constructiviste, l'enseignant intervient en tenant compte du groupe et des différences individuelles dans le groupe, en vertu de sa mission première de faciliter les apprentissages de ses élèves. Mais il vient un temps où l'on doit vérifier empiriquement si les apprentissages souhaités ou visés ont été réalisés et lesquels l'ont été. Une telle vérification relève de la fonction sommative de l'ÉVAP. Lorsqu'elle est mise en oeuvre, cette évaluation est externe à ce qui a été vécu en classe.

Les études connues concernant les nombreuses caractéristiques des élèves, telles la motivation scolaire, les connaissances antérieures, l'autoévaluation, les perceptions de l'enseignant, tendent à démontrer que tous ces aspects n'ont pas de liens

significatifs avec les apprentissages réalisés ou les compétences acquises à l'école. De même, des études comparant des approches pédagogiques ont démontré que des résultats positifs de certaines approches pédagogiques résident dans des conditions très contextualisées, telles les caractéristiques des élèves, de l'enseignant, du matériel didactique disponible et sont difficilement transférables et généralisables à un système d'éducation. En conséquence, limiter l'ÉVAP à ce qui se passe en classe n'a pas d'appui scientifique. Au contraire, l'expérience vécue et les développements scientifiques de l'ÉVAP militent en faveur de fonder l'ÉVAP sur des observations et des mesures qui soient indépendantes des particularités des relations pédagogiques et didactiques vécues en classe, mais en adéquation avec les connaissances, habiletés ou compétences inhérentes au programme d'étude. Après tout, ne vise-t-on pas à ce que l'élève soit capable de transfert de ses compétences dans d'autres contextes, hors de sa classe, de son groupe-classe, de son groupe-cycle ? Ceci invite à une évaluation à caractère décontextualisé.

## **8. Instruments de consignation**

Tout instrument de consignation doit répondre à l'un ou à l'autre des buts suivants, ou aux deux, à savoir :

- A) informer les parents** sur le cheminement scolaire de leur enfant : ce qu'il apprend ou a appris à l'école; ce qu'il devrait y apprendre ou y avoir appris mais qu'il n'a pas appris. Comme cette finalité appartient aux parents, ou à l'élève lui-même à la fin du secondaire, il faut que l'instrument de consignation les renseigne clairement;
  
- B) conserver des traces** pour faciliter le cheminement scolaire, la promotion, la sélection, la diplomation, le changement d'école ou de commission scolaire.

Deux types de remarques s'imposent ici. D'une part, aucun de ces buts ne concerne la communication pédagogique enseignant-élève. L'instrument de consignation ne vise pas à soutenir ni à rapporter cette communication pédagogique dans la classe. Il sert plutôt à conserver des résultats d'apprentissage afin de les faire connaître ou de les utiliser en-dehors de la classe. Une liste exhaustive de comportements à apprendre (objectifs, compétences, etc.) est inappropriée pour la conservation de traces à des fins de système. Une telle liste appartient à la classe, au support de la communication pédagogique entre l'élève et l'enseignant; elle peut être occasionnellement communiquée aux parents, mais elle sert d'abord à éclairer les décisions à prendre en vue d'améliorer les apprentissages.

D'autre part, il est d'usage d'utiliser des instruments de consignation (le plus couramment appelés **bulletins**) qui répondent à la fois aux buts A et B. On évite ainsi d'ajouter aux tâches cléricales et aux formulaires existants; surtout, on assure les élèves et les parents que le système scolaire conserve comme traces consignées les informations apparaissant dans les bulletins qu'ils reçoivent. Cela ne définit pas automatiquement les éléments (contenus, compétences ou habiletés, mesures, jugements, décisions) des bulletins; mais on peut se servir des multiples expériences québécoises pour démontrer que les bulletins visant les deux buts doivent posséder certaines caractéristiques de synthèse : la liste des éléments ne doit pas être trop longue, ce qui entraîne que chaque élément désigne un ensemble assez global (une matière ou une partie importante de matière) de sorte que le jugement qui est consigné sur cet ensemble constitue un jugement global, une sorte de moyenne de jugements partiels que l'enseignant pourrait porter sur les composantes de cet ensemble. Pour la consignation de résultats d'ÉVAP communiqués et utilisés en-dehors de la classe, seuls des jugements d'ensemble paraissent constituer des traces utiles au système et la majorité des parents semblent réclamer que de tels jugements d'ensemble leur soient communiqués.

Cette perspective appelle quelques commentaires sur les portfolios. Leur usage semble de plus en plus répandu et, surtout, certains en



font la promotion comme instrument d'évaluation ou de consignation. À la fois selon la documentation pédagogique et selon la littérature scientifique, le portfolio est toujours considéré comme un recueil, tantôt pour consigner et communiquer du matériel et des résultats d'évaluation, tantôt pour regrouper des travaux réalisés et sur la base desquels une évaluation d'ensemble peut être faite. S'en servir comme d'un instrument de communication permet de conserver des traces à la fois pour l'élève et pour l'enseignant, tout en exposant aux parents des éléments « sélectionnés » parmi tout ce qui a été fait. Ainsi considéré, le portfolio s'avère sans doute un outil pédagogique intéressant pour la communication enseignant-élève. Chaque enseignant aura à apprécier une telle utilité pédagogique à la lumière des énergies et du temps qu'il lui faudra consacrer pour assurer le suivi et la mise à jour des portfolios individuels. Sur le plan de la communication externe, certains parents peuvent trouver le portfolio intéressant... mais les parents québécois qui se sont exprimés veulent plus que cela et les portfolios ne sauraient servir adéquatement d'instruments de consignation pour le système scolaire. Se servir du portfolio comme instrument d'évaluation, c'est-à-dire un classeur servant à l'évaluation (recueil de travaux réalisés et sur la base desquels une évaluation d'ensemble serait alors faite) peut avoir un bon côté, celui de ne pas perdre de vue ce que l'élève a réussi ou réalisé de bien, d'excellent même, puisque le portfolio-recueil contiendra probablement surtout les productions illustrant les apprentissages réussis. Mais s'il ne contient pas aussi les illustrations de ce que l'élève ne réussit pas (parmi ce qu'il devrait réussir), on revient au dérapage d'un « portfolio au service de la réussite » et non au service de l'apprentissage, c'est-à-dire qui laisse croire que tout est réussi parce que les traces de ce qui ne l'est pas sont absentes, de même que les traces des difficultés que l'élève pourrait avoir.

En conséquence, nous reconnaissons l'utilité du portfolio comme moyen de communication pédagogique entre l'élève et l'enseignant, dont les parents peuvent être informés et auquel ils peuvent participer. Le contenu du portfolio constitue une trace de classe que l'enseignant utilise pour ajuster ses interventions au besoin et pour appuyer ses jugements sur les apprentissages (réalisés, difficiles, à faire), mais le portfolio lui-même n'est pas un instrument de

consignation pouvant servir au système scolaire. Le portfolio appartient à la classe et non au système. Il appartient à la gestion de classe et non à l'ÉVAP. Comme il s'agit d'un classeur, il peut contenir les travaux et examens faits et notés, avec les résultats de mesures et les rétroactions; il peut aussi contenir des informations quantitatives ou qualitatives sur des manifestations des acquis de l'élève. De tels éléments peuvent être pris en compte dans le jugement d'ensemble rendu par l'enseignant en évaluation sommative, ce qui ne fait pas du portfolio un instrument d'évaluation. L'évaluation sommative requiert des informations pertinentes et suffisantes, incluant des observations objectives, valides et fidèles, relatives à diverses compétences attendues et à leurs manifestations dans des contextes variés; elle requiert aussi des critères auxquels les informations doivent être comparées pour assurer un jugement juste et équitable.

## **9. Considérations sur la formation des enseignants à l'ÉVAP**

Généralement, du moins à l'UQÀM, la formation initiale à l'ÉVAP se limite à une initiation dans le cadre d'une seule activité de trois crédits, soit 45 heures de cours réparties le plus souvent sur deux à trois mois des quatre années que dure la formation à l'enseignement. Une telle contrainte exige que cette initiation cible ce qui assure validité et fidélité des informations, fiabilité et crédibilité du jugement d'évaluation, pertinence et équité des décisions, c'est-à-dire les fondements scientifiques de l'évaluation des apprentissages, que ces apprentissages soient appelés connaissances, habiletés ou compétences. Évidemment, ces fondements sont mieux compris, intégrés, lorsqu'ils sont accompagnés d'une démarche d'évaluation que l'enseignant en formation est appelé à planifier et à réaliser. Nos considérations sur la formation à l'ÉVAP sont ici regroupées sous trois volets circonscrits par l'objectivité, les nouvelles tendances, les jugements d'évaluation.

## **La recherche de l'objectivité**

Tout n'est pas mesurable et les scientifiques cherchent à identifier des manifestations encore inconnues et à rendre observable ce qui ne l'est pas encore. En ÉVAP, il est admis de limiter l'objet à ce qui est mesurable, de n'évaluer que les apprentissages mesurables et mesurés. Cette position, souvent adoptée depuis 20 ou 30 ans, est théoriquement et scientifiquement fondée. Mais limiter l'ÉVAP aux apprentissages mesurés ne signifie pas qu'on ne doit pas porter de jugement sur ce que l'on mesure. Nous pensons qu'une large part des inconvénients rencontrés au cours des dernières années est due au fait qu'on s'est souvent limité à mesurer sans juger. Il est plus facile d'exécuter des procédures que de porter un jugement et la mesure semble offrir une protection de neutralité moins susceptible d'être contestée. Réclamer le retour à un véritable jugement, c'est vouloir nuancer les résultats des mesures et non les ignorer. Intégrer dans l'ÉVAP les développements scientifiques du XX<sup>e</sup> siècle, c'est intégrer à la fois des mesures valides et fidèles et des jugements professionnels crédibles et fiables pour produire des décisions pédagogiques et administratives pertinentes et équitables.

En contexte de formation initiale ou continue, former des enseignants à l'ÉVAP, c'est les rendre capables de composer avec les concepts, les méthodes et les principes de base de la mesure et de l'évaluation; capables, dans un contexte défini par les fonctions connues d'une démarche d'ÉVAP, de construire un instrument de mesure, de recueillir des données, d'analyser les résultats, de les interpréter, puis de juger et de décider des actions à entreprendre; capables de comprendre les résultats consignés d'évaluation d'apprentissages; capables d'intégrer à leur enseignement un modèle centré sur des comportements (objectifs, savoirs, habiletés, compétences) en fonction desquels ils pourront organiser des situations efficaces d'évaluation des apprentissages. Pour soutenir ces objectifs de formation, nous avons régulièrement mis à jour des recueils de notes de cours et un plan de cours communs. Les principaux éléments, dont l'efficacité a été éprouvée sur plusieurs années et auprès de nombreux groupes d'enseignants en formation initiale ou continue, ont été consignés par trois d'entre nous dans six

modules et cinq fascicules regroupés dans une publication (Montréal : Les Éditions Logiques, 2000) intitulée *Formation de base en évaluation des apprentissages*.

### **Les discours des nouvelles tendances**

Les enseignants en formation sont notamment saisis des nouvelles tendances en ÉVAP dans une double perspective historique et critique : culture de l'évaluation; intégration à la didactique, à la pédagogie, à l'enseignement, à l'apprentissage; l'évaluation des processus d'apprentissage et de résolution, du progrès; participation de l'évalué; contexte signifiant, évaluation authentique et validité authentique; tandem processus-information; profils de scores. Le Québec n'échappe pas à plusieurs des influences de ces mouvements : on parle d'examens authentiques et on veut redonner l'importance au jugement professionnel de l'enseignant. Mais on ne sait pas comment exercer un jugement professionnel en ÉVAP à partir de concepts tels compétences, habiletés transférables, habiletés ou compétences transversales, pour qu'un tel jugement soit respectueux des principes de justice, d'égalité et d'équité. De plus, quels sont les indicateurs, critères et autres balises pour évaluer les manifestations de telles habiletés complexes au primaire ou au secondaire ?

On s'accorde par ailleurs à reconnaître la nécessité d'implanter un mécanisme national d'ÉVAP, pour continuer à renseigner la population, les responsables locaux de l'éducation et les éducateurs sur le rendement des différents établissements, dont les font partie les apprentissages des élèves. D'une part, on veut se préoccuper d'évaluer le processus et les stratégies personnelles de résolution de problèmes dans des situations authentiques et près de la réalité sociale; somme toute, on veut ajuster l'évaluation aux préoccupations de chaque individu. D'autre part, le fait que l'évaluation nationale soit porteuse d'informations pour les décideurs milite en faveur d'évaluations plus standardisées et moins contextualisées.

Les préoccupations québécoises actuelles en matière d'ÉVAP sont donc beaucoup influencées par des tendances dites modernes. Ces tendances permettent-elles de mieux évaluer ou constituent-elles un recul en ignorant les acquis scientifiques et d'expériences du vingtième siècle ? Appuyer des jugements d'évaluation davantage sur des perceptions d'apprentissages réalisés que sur des informations recueillies à partir d'instruments systématiquement construits, est-il progressiste ou rétrograde ?

### **Les jugements d'évaluation**

L'ÉVAP est souvent considérée comme un mal nécessaire : personne n'aime ni ne veut être évalué, jugé, y compris sur ses apprentissages scolaires. Tout élève dont l'ÉVAP démontre des faiblesses ou des échecs, a tôt fait d'en attribuer les causes à diverses particularités : épreuves, contenus, contextes, circonstances, conditions environnantes ou psychologiques, etc. De la même façon, les responsables d'évaluer les apprentissages éprouvent souvent de l'inconfort : ils craignent d'exercer un jugement qui pourrait être mal fondé, à partir d'informations biaisées, erronées ou insuffisantes, et qui pourrait avoir un effet négatif sur la vie future d'êtres humains, à court terme mais aussi à moyen ou long terme.

Par ailleurs, il est rare que tous les apprentissages visés soient maîtrisés complètement, ou qu'ils ne le soient pas du tout. Le plus souvent, il y a des choses que l'élève a apprises ou appris à faire, qu'il démontre dans un ou plusieurs contextes; et il y en a d'autres qu'il n'a pas apprises, quel qu'en soit le contexte. L'observation ou la mesure systématique permettent de cerner les différents objets et contextes pertinents à l'expression des apprentissages visés. Les informations ainsi recueillies constituent des témoins des apprentissages réalisés et le jugement d'évaluation les apprécie en termes de qualité ou de quantité par rapport aux apprentissages visés.

## **Conclusion**

Nous sommes en faveur de toujours chercher à améliorer la pertinence et la suffisance des informations, ce qui passe par leur validité et leur fidélité, pour assurer la justice et l'équité des jugements et de tout le processus d'évaluation des apprentissages. Mais en affirmant cela, nous rappelons au moins trois éléments d'importance :

- Évaluer pour faciliter l'apprentissage constitue un acte pédagogique qui correspond à la mission première de la scolarisation des jeunes et de la société. Mais la démarche et les contextes d'une telle ÉVAP formative se distinguent et doivent se distinguer de ceux mis en œuvre dans une opération d'ÉVAP sommative, par laquelle l'élève est appelé à démontrer à la société ce qu'il a appris ou ce qu'il a appris à faire à l'école.
- Il est utopique de vouloir tout observer ou tout mesurer, tous les objets et dans tous les contextes possibles. Mais ce n'est pas une raison suffisante pour discréditer la validité de l'ÉVAP et suggérer, éventuellement, de ne plus évaluer globalement les apprentissages scolaires.
- Plusieurs des moyens proposés pour la collecte d'informations sur les apprentissages scolaires (grilles d'observations, épreuves à réponses élaborées accompagnées de grilles de correction) semblent avoir tendance à rejeter les principaux acquis scientifiques : la subjectivité est alors plus présente dans la prise d'informations et on diminue le nombre de comportements mesurés ou le nombre de contextes dans lesquels ils le sont. Ces caractéristiques sont contraires aux conditions qui favorisent validité et fidélité, pertinence et suffisance. Il faut éviter le piège qui nous guette de ne plus tenir compte des évaluations parce qu'elles seraient devenues trop subjectives.

# Pour une évaluation des apprentissages scientifiquement fondée

## LES AUTEURS

Serge P. Séguin, Réjean Auger et Carmen Parent, tous trois professeurs au département des sciences de l'éducation (DSÉ) de l'Université du Québec à Montréal (UQÀM), détiennent chacun un grade de M.A. suivi d'un grade de Ph.D. en mesure et évaluation en éducation. Jean-Yves Lancup, Claudine Nézet-Séguin et Zeycan Yegin, tous trois chargés de cours au même DSÉ de l'UQÀM, détiennent chacun un grade de M.A. en mesure et évaluation en éducation. Ensemble, ces six auteurs assument la formation en ÉVAP de presque tous les diplômés de l'UQÀM en enseignement préscolaire, primaire, secondaire et en adaptation scolaire.

## Le LABFORM

### ***Laboratoire de recherche, analyse, développement et formation en mesure et évaluation***

Le **LABFORM** désigne un regroupement de chercheurs universitaires et le lieu de ce regroupement, dont la fonction première est la recherche dans le domaine de la mesure et de l'évaluation en éducation, laquelle peut être libre, subventionnée ou commanditée. Recherches et analyses ont souvent pour but le développement de stratégies et de méthodes.

Outre ces fonctions normales d'un laboratoire universitaire de recherche, le **LABFORM** a aussi comme fonction déclarée de constituer un milieu de formation pour les étudiants et étudiantes qui y travaillent et d'assurer des interventions de formation spécialisée commanditées par des organismes externes.



Faculté d'éducation, Université du Québec à Montréal

Juin 2001