

l'évaluation des étudiants et les principales conceptions de la mesure et de l'évaluation

par Gérard SCALLON *

L'évaluation des étudiants n'est pas une préoccupation nouvelle au Québec. Depuis quelques années cependant, elle est devenue un point névralgique dans de nombreuses discussions ou remises en question. Alors qu'elles étaient dans le passé sous la responsabilité presque exclusive des administrateurs, les activités d'évaluation exercées par les enseignants et les étudiants prennent de plus en plus d'importance. Mais ce n'est qu'une tendance parmi plusieurs, car certaines conceptions de l'enseignement remettent parfois en question le bien-fondé même de l'évaluation des étudiants. Dans ce débat, le profane, l'homme de la rue et même plusieurs professionnels de l'éducation éprouvent une certaine confusion.

Dans le domaine de l'évaluation des étudiants, la mise en marché de concepts, comme la mesure normative, la mesure critériée, l'évaluation « formative », l'évaluation « sommative » ou l'évaluation continue, ne se réalise pas sans problème. Ces concepts ne sont pas d'hier mais leur utilisation est nouvelle, et leur injection dans notre système d'éducation risque d'augmenter la confusion si l'on ne se préoccupe pas de bien les clarifier au point de départ.

Nous nous proposons dans cet article d'établir quelques distinctions qui apparaissent à la base de toute discussion au sujet du bien-fondé de

l'évaluation des étudiants et qui semblent nécessaires à la compréhension de plusieurs termes utilisés. Nous n'avons pas la prétention de présenter ici un texte qui soit savamment documenté et qui traite adéquatement des problèmes que soulève l'évaluation des étudiants en milieu scolaire. Notre préoccupation sera plutôt de définir le plus simplement possible un certain nombre de termes qui sont de plus en plus utilisés dans les écrits qui traitent de l'évaluation des étudiants.

Mesurer et évaluer

Une distinction qui nous apparaît fondamentale est celle que nous devons établir entre *mesurer* et *évaluer*. Nombre de théoriciens de l'évaluation s'entendent pour affirmer que mesurer est une opération qui consiste essentiellement à associer des symboles à des objets, à des événements ou à des personnes selon des règles précises. Les symboles qui ont été utilisés au cours de cette opération s'appellent, une fois l'opération terminée, les résultats de la mesure. Dans la plupart des cas, ces symboles sont des nombres qui pour

* L'auteur est professeur en mesure et évaluation de la faculté des Sciences de l'éducation de l'Université Laval.

l'instant sont purement descriptifs. Ils doivent leur signification à l'unité de mesure qui les accompagne. Ainsi, on peut associer le nombre 7 à une table parce que cette table a une longueur de 7 pieds. Dans ce cas, le pied, qui est l'unité de mesure utilisée, a été appliqué un certain nombre de fois sur la longueur de la table. Le nombre 7 prend donc sa signification en vertu de sa relation à une longueur standard qui est le pied. La plupart des résultats de mesure obtenus en éducation n'ont pas cette propriété, à cause de la difficulté de trouver une unité de mesure pouvant servir de standard. Par exemple, un résultat comme « 40 bonnes réponses » à une épreuve de *rendement* est difficile à interpréter, parce que les 40 questions correctement résolues pour obtenir ce résultat ne constituent pas nécessairement des unités égales.

Évaluer, d'autre part, est une opération qui consiste à porter des jugements de valeur ou à accorder des valeurs à des objets, à des événements ou à des personnes en comparant ces objets, événements ou personnes avec un critère quelconque. Cette opération peut aussi s'appliquer aux résultats d'une mesure. Une fois l'opération terminée, on donne le nom d'évaluation aux jugements de valeur qui ont été portés. Des expressions comme « excellent », « médiocre », « supérieur à la moyenne », « succès » ou « échec » etc. sont des exemples d'évaluations rencontrées en éducation. Il faut comprendre que l'évaluation d'un résultat de mesure est d'autant plus importante que ce résultat est difficile à interpréter, à cause de l'absence d'une unité de mesure pouvant lui donner une signification.

Quelques exemples nous feront mieux saisir la distinction que nous venons d'établir entre mesurer et évaluer. Quelqu'un nous informe qu'un de ses amis est âgé de 90 ans. L'expression *90 ans* (ou le nombre 90) est le résultat d'une mesure. Pour associer le nombre 90 à l'individu concerné dans notre exemple, on a sans doute appliqué une règle précise qui consiste à déterminer combien de fois une certaine unité de mesure, l'année d'existence, est contenue dans l'écart qui sépare deux moments précis: la date de naissance de l'individu et la date où nous nous trouvons présentement. Si notre interlocuteur nous avait dit qu'un de ses amis est très âgé, il nous aurait alors communiqué une évaluation. Voici un autre exemple. Une personne nous apprend qu'elle a rencontré quelqu'un dont la hauteur physique est de 8 pieds. La règle qui

a été utilisée ici pour associer le nombre 8 à un individu nous est connue. Si cette même personne nous avait dit qu'elle a rencontré un géant, elle nous aurait communiqué alors son propre jugement de valeur, c'est-à-dire son évaluation de la grandeur physique de l'individu qu'elle a rencontré. On peut voir aisément, avec les deux exemples qui précèdent, que le résultat de la mesure est purement descriptif et ce sans qu'aucun jugement de valeur ne soit porté intentionnellement. Il s'agit là de l'une des principales caractéristiques de la mesure qui distingue celle-ci de l'évaluation. On objectera que la distinction n'est pas très claire dans les deux exemples, car tout le monde sait qu'un homme âgé de 90 ans est très âgé et qu'un individu mesurant 8 pieds de hauteur est un géant. C'est que nous faisons face ici à une situation très fréquente dans la vie courante, et dans laquelle le résultat d'une mesure s'inscrit facilement dans un cadre de référence qui correspond à notre expérience vécue. Nous connaissons tous l'unité de mesure qui a été utilisée dans chaque exemple: l'année d'existence dans le cas de l'âge et le pied dans le cas de la grandeur physique. De plus, nous connaissons les possibilités actuelles de la longévité humaine de même que la répartition de la grandeur physique chez un très grand nombre d'individus. Ce sont ces deux informations, l'une concernant l'unité de mesure utilisée et l'autre concernant les possibilités ou les limites du phénomène étudié, qui nous permettent très souvent de porter notre propre jugement de valeur à partir uniquement d'un résultat de mesure. Notons que dans les deux exemples qui précèdent, les jugements de valeur « très âgé » et « géant » sont des évaluations relatives à un ensemble d'individus dont on connaît la longévité et la grandeur physique. Le critère qui a servi de base à de telles évaluations est la caractéristique d'un ensemble d'individus. Il suffit que cette caractéristique soit modifiée pour que l'évaluation qui lui est relative soit elle aussi modifiée. Par exemple, si l'on parvenait un jour à prolonger la vie humaine jusqu'à 200 ans, alors le jugement de valeur « très âgé » ne conviendrait plus pour une personne de 90 ans.

Dans certains contextes, l'unité de mesure et les possibilités du phénomène étudié ne font pas partie de notre expérience vécue. Il devient alors difficile, sinon impossible, d'évaluer nous-mêmes à la suite du résultat d'une mesure. Prenons une situation fictive comme exemple. Un astronaute revient d'un voyage interplanétaire et nous

apprend que le premier habitant qu'il a rencontré sur la planète BX-12 mesurait 3 CVZ de hauteur (le CVZ étant l'unité de longueur utilisée par les habitants de la planète BX-12). Nous sommes en face du résultat d'une mesure et il nous est impossible d'évaluer la grandeur de cet habitant ou de porter un jugement de valeur comme: « très petit », « de grandeur moyenne » ou « très grand ». C'est que nous ne possédons ni l'une ni l'autre des deux informations qui sont essentielles à l'évaluation. D'abord, nous ne connaissons pas le CVZ et ensuite, nous n'avons pas la moindre idée de la grandeur physique de l'ensemble des habitants de la planète BX-12. Supposons que notre astronaute nous fournit des indices quant à cette deuxième information: il nous dit que la grandeur physique des habitants de cette planète varie entre 3CVZ et 9 CVZ. Déjà ce renseignement additionnel nous permet de déduire que le premier habitant qu'il a rencontré, avec ses 3 CVZ, se situe fort probablement parmi les plus petits habitants de la planète BX-12. Il faut noter que cette évaluation est limitée au cadre, de référence défini par la grandeur physique des habitants de la planète concernée. Si nous devions évaluer la grandeur de notre habitant en fonction de nos dimensions humaines, il nous faudrait connaître l'unité de mesure, le CVZ. Supposons que le CVZ équivaut approximativement à trois pieds de longueur. Nous sommes capables maintenant de déduire que l'un des plus petits habitants de la planète BX-12 serait considéré comme un géant sur notre terre. Ces deux jugements de valeur n'ont rien de contradictoire et illustrent le caractère relatif de l'évaluation.

Le contexte que nous venons de décrire, dans lequel nous ne connaissons pas l'unité de mesure et les possibilités du phénomène étudié, est très fréquent en éducation. C'est le problème que nous devons résoudre lorsqu'il s'agit d'évaluer l'intelligence d'une personne, ses aptitudes, ses intérêts ou la qualité de son rendement scolaire à partir d'un résultat de mesure. Le plus souvent, ce résultat de mesure est un nombre qui indique une quantité de bonnes réponses à un ensemble de questions posées dans un test ou dans une épreuve. Ainsi, un résultat de 60 obtenu par un étudiant à une épreuve de rendement, ne nous indique rien de la qualité de cet étudiant ou de sa compétence. Pour porter un jugement de valeur, il nous faudrait connaître l'une ou l'autre des deux choses. Si nous étions familiers avec la nature des problèmes que l'étudiant a pu résoudre dans cette épreuve, nous pourrions

alors inférer les capacités ou les habiletés dont cet étudiant peut faire preuve. Il s'agit ici d'une évaluation qui fait abstraction des résultats obtenus par les autres étudiants. De plus, notre connaissance de la nature des problèmes résolus ou du contenu de l'épreuve est un compromis face à la difficulté d'identifier une unité de mesure ou un standard. Si, d'autre part, nous connaissions les possibilités de résultats à cette épreuve, telles qu'elles ont été révélées par les résultats obtenus par d'autres étudiants, nous pourrions alors porter un jugement de valeur comme « excellent », « médiocre », « supérieur à la moyenne » ou « normal ». Il s'agit, dans ce deuxième cas, d'une évaluation dont le critère de base est la caractéristique d'un ensemble d'étudiants.

Le critère pouvant servir de base à l'évaluation d'un étudiant est une notion fondamentale dans la distinction qui sera établie entre la mesure critériée et la mesure normative. Nous y reviendrons un peu plus loin dans cet article.

Une définition fonctionnelle de l'évaluation

Nous avons défini l'évaluation comme une opération distincte de la mesure. Cette définition est incomplète, car elle ne met pas en évidence le rôle de l'évaluation. On ne porte pas de jugements de valeur comme « excellent » ou « inférieur à la moyenne » uniquement pour le simple plaisir d'évaluer ou pour cataloguer. Malheureusement, trop de textes sur le sujet entretiennent la fausse perception que l'évaluation est une activité suffisante en elle-même et ouvrent ainsi la porte à des discussions non justifiées. Il existe plusieurs définitions de l'évaluation qui nous permettent d'échapper à cette difficulté. Toutes ces définitions ont un caractère commun. L'évaluation y est vue comme un ensemble de procédés qui consistent à identifier la nature de certaines décisions, à recueillir les informations appropriées en vue de communiquer des données utiles à ceux qui devront prendre ces décisions.

Les décisions qui sont prises en éducation sont trop nombreuses pour en dresser une liste exhaustive. Le choix d'un procédé d'enseignement, l'implantation et le maintien d'un régime pédagogique particulier à un niveau quelconque, l'admission des étudiants dans une institution

d'enseignement, l'inscription à un programme d'études et l'organisation de rythmes particuliers d'apprentissage pour des étudiants déficients sont des exemples variés de prises de décision. Les nombreuses décisions qui concernent chaque étudiant peuvent être prises par plusieurs personnes: enseignants, parents, administrateurs, conseillers d'orientation, étudiants, employeurs etc. Ces décisions, pour être prises de façon éclairée, doivent être précédées d'une évaluation appropriée. Le concept de prise de décision est donc fondamental dans le domaine de l'évaluation pour deux raisons importantes. Premièrement, il nous oblige à réaliser que l'évaluation n'est pas une fin en elle-même mais un moyen qui est au service d'une prise de décision. Deuxièmement, les divers modes d'évaluation que nous rencontrons de nos jours se définissent par le contexte qui caractérise certaines prises de décision. C'est ainsi que nous distinguerons la mesure normative de la mesure critériée, et l'évaluation sommative de l'évaluation formative.

Mesure normative & mesure critériée

Dans le domaine de l'apprentissage et du rendement scolaire, il existe deux critères de comparaison qui peuvent servir de base à l'évaluation des étudiants: (1) la caractéristique d'un groupe d'étudiants et (2), un seuil de rendement ou de performance. Voyons par un exemple comment chacun de ces deux critères de comparaison peut être appliqué à l'évaluation d'un résultat de mesure. Un étudiant a obtenu un résultat de 65 à une épreuve de rendement. Si le critère de comparaison adopté pour évaluer cet étudiant était l'ensemble des résultats obtenus à la même épreuve par ses confrères de classe, nous pourrions constater par exemple qu'un résultat de 65 se situe parmi les 10 pour cent de résultats les plus faibles qui ont été observés dans la classe. Nous pourrions alors évaluer cet étudiant comme étant nettement inférieur à la moyenne et c'est un jugement de valeur basé sur la caractéristique du groupe auquel cet étudiant appartient. Si, d'autre part, le critère de comparaison pour évaluer l'étudiant qui a obtenu un résultat de 65 était un seuil de performance correspondant à un résultat de 60 par exemple, nous pourrions alors affirmer que l'étudiant a réussi l'épreuve peu importe les résultats d'autres étudiants à cette épreuve. Dans ce deuxième cas,

la notion de réussite est une évaluation basée sur un seuil de performance (ou critère de performance).

L'évaluation en fonction d'une caractéristique de groupe est dite « normative », car le plus souvent cette caractéristique s'exprime par la moyenne des résultats obtenus par un ensemble d'étudiants à une épreuve de rendement. La moyenne constitue alors une norme avec laquelle le résultat de chaque étudiant est comparé. L'évaluation en fonction d'un seuil ou critère de performance est dite critériée. Les qualificatifs « normative » et « critériée » ne sauraient être appliqués à la mesure comme telle, celle-ci étant exempte de tout jugement de valeur. Pourtant, l'emploi d'expressions comme « mesure normative » (*norm-referenced testing*) et « mesure critériée » (*criterion-referenced testing*), s'est vite généralisé dans les textes qui ont été publiés sur le sujet. Il faut comprendre ici que l'expression « mesure normative » signifie un type d'instrument de mesure (test, examen ou épreuve) utilisé en vue d'une évaluation normative, et que l'expression « mesure critériée » signifie un type d'instrument de mesure utilisé en vue d'une évaluation critériée. Cette interprétation est justifiée par le fait qu'il est impossible de déterminer, uniquement à partir de l'apparence physique des questions qui le composent, si un test est normatif ou critérié. Pour ce faire, il faut absolument connaître le mode d'évaluation pour lequel ce test a été conçu.

La différence entre la mesure normative et la mesure critériée ne se situe pas uniquement au niveau du mode d'évaluation utilisé. Le mode de communication des résultats de mesure, la rédaction des questions et le contexte de la prise de décision sont également d'autres aspects pour lesquels la mesure normative et la mesure critériée sont nettement différentes.

Dans le cas de la mesure normative, on se préoccupe surtout de communiquer des résultats qui reflètent le mieux possible la position relative occupée par chaque étudiant dans un groupe donné, qu'il s'agisse d'une classe, de plusieurs classes ou d'un ensemble d'étudiants ayant subi le même examen. C'est pourquoi la note initiale obtenue par chaque étudiant à un examen est convertie en rang centile, en rang stanine, en rang cinquième ou en score-type. Dans plusieurs institutions d'enseignement, la note est convertie en lettres A, B, C, D ou E selon la région autour

de la moyenne à laquelle le résultat de l'étudiant appartient.

Dans le cas de la mesure critériée, l'accent est mis sur la description des habiletés ou capacités maîtrisées par chaque étudiant, en vue de faciliter une évaluation en fonction d'un seuil de performance. Avec une épreuve composée de questions mesurant la même habileté, on rapportera la note initiale de chaque étudiant tout en indiquant le seuil de performance exigé. Avec une épreuve mesurant des habiletés différentes, on tentera de décrire le mieux possible le niveau de compétence de chaque étudiant en regard de chaque habileté mesurée. La communication des résultats de la mesure critériée n'implique pas nécessairement des informations quantitatives et peut revêtir des aspects qualitatifs. On peut concevoir, par exemple, un bulletin d'information contenant en énumération les habiletés maîtrisées par un étudiant à la suite d'une séquence d'apprentissage. Malgré les diverses formes que peut revêtir la communication des résultats, il est important de souligner que dans la perspective d'une évaluation critériée, l'évaluation de chaque étudiant est totalement indépendante de l'évaluation des autres étudiants.

Du côté de la rédaction des questions et de la construction d'épreuves de rendement, la différence entre la mesure normative et la mesure critériée se manifeste par une différence dans l'importance accordée aux différences individuelles. Dans le cas de la mesure normative, les questions sont rédigées en vue d'accentuer les différences qui peuvent exister entre les étudiants. Ce but est atteint à la limite lorsque parmi les résultats obtenus, il n'existe pas deux résultats qui sont identiques. Dans le cas de la mesure critériée, la recherche de différences individuelles n'est pas une préoccupation majeure. Dans le cas de la mesure normative comme dans le cas de la mesure critériée, la rédaction de questions d'examen doit s'inspirer des objectifs d'un programme ou d'un cours. Cependant, en mesure normative, la recherche de différences individuelles amène parfois des concessions importantes au niveau de la rédaction de questions, concessions qui peuvent faire dévier ces questions des objectifs pédagogiques visés.

Le contexte de la prise de décision est sans aucun doute l'aspect le plus important à consi-

dérer dans la distinction qu'il faut établir entre la mesure normative et la mesure critériée. Dans certaines situations, la prise de décision implique le choix d'un nombre limité de candidats parmi un très grand nombre. C'est ce qui se produit en pratique lorsqu'une institution d'enseignement ne peut admettre chaque année qu'un nombre limité d'étudiants parmi un très grand nombre d'étudiants qui ont demandé leur admission. Le marché du travail comporte une situation semblable lorsqu'il s'agit de choisir un candidat parmi plusieurs qui postulent un emploi. Dans de telles situations où la prise de décision implique un choix limité, on accorde la préférence aux meilleurs candidats qui se présentent et la décision est précédée la plupart du temps d'une évaluation normative. Cette évaluation fait suite à une épreuve ou à un examen conçus dans le but de faire ressortir les différences individuelles. Il y a d'autres situations cependant où la position relative d'un individu dans un groupe donné n'est pas pertinente à une prise de décision. Par exemple, pour devenir pilote d'avion, un minimum de perception visuelle est requis. Le fait de savoir qu'un candidat possède la meilleure vision dans un groupe donné ne garantit pas qu'il possède le minimum exigé. Dans un contexte comme celui-ci, la décision d'admettre le candidat (en supposant qu'il n'y a pas de contraintes dans le nombre d'admissions possibles), doit être précédée d'une évaluation du candidat qui est faite en relation avec un seuil de perception visuelle déterminé selon les exigences du métier et non pas en relation avec la perception visuelle d'autres individus. Dans le domaine de l'apprentissage scolaire, la même situation se présente si on se préoccupe de déterminer le degré d'atteinte de certains objectifs pédagogiques. Au niveau de chaque objectif, nous pouvons être intéressés à savoir si un étudiant possède un minimum de compétence, et il n'y a pas de restrictions quant au nombre d'étudiants qui peuvent avoir ce minimum de compétence. Dans un contexte comme celui-là, une mesure conduisant à une évaluation critériée est la plus appropriée.

Nous venons de voir avec quelques exemples que dans certains cas la prise de décision peut être précédée d'une évaluation normative, alors qu'en d'autres cas la prise de décision peut être précédée d'une évaluation critériée. C'est véritablement le contexte de cette prise de décision qui nous permet de dire si une épreuve de rendement est normative ou critériée.

Évaluation formative & évaluation sommative

La notion de critère de comparaison pouvant servir de base à l'évaluation d'un étudiant, nous a amenés à établir une distinction importante entre deux types de mesure: la mesure faite en vue d'une évaluation normative et la mesure faite en vue d'une évaluation critériée. Cette distinction est utile lorsque nous nous limitons à l'évaluation d'un étudiant réalisée à un moment très précis. Si nous considérons l'ensemble des évaluations auxquelles chaque étudiant peut être soumis au cours de sa scolarité, il nous faut maintenant distinguer entre deux approches: l'approche formative et l'approche sommative. Les deux approches se distinguent l'une de l'autre selon trois points de vue: premièrement, selon la nature des décisions qui sont prises à l'endroit de chaque étudiant; deuxièmement, selon l'instant précis où l'évaluation est réalisée dans le déroulement d'un programme d'études ou d'un cours et, troisièmement, selon la nature même des habiletés qui sont évaluées.

L'évaluation formative est un système d'évaluation qui consiste à recueillir, à plusieurs occasions pendant le déroulement d'un programme d'études ou d'un cours, des informations utiles dans le but de vérifier périodiquement la qualité de l'apprentissage des étudiants. À chaque occasion, l'évaluation est conçue dans le but de décider si un étudiant peut continuer dans le programme ou dans le cours, ou s'il doit retourner en arrière pour combler des lacunes du côté de son apprentissage. Une telle décision est prise par l'enseignant et dans certains cas par l'étudiant lui-même. En évaluation formative, le contenu de chaque évaluation faisant suite à une unité d'enseignement doit être en relation directe avec des objectifs pédagogiques clairement énoncés et qui sont particuliers ou spécifiques à cette unité. La préoccupation majeure de l'évaluation formative est de mesurer, pour chaque étudiant, le degré d'atteinte de chacun des objectifs et aussi de comparer ce degré d'atteinte avec un seuil de performance déterminé. La mesure permettant une évaluation critériée de l'étudiant est sans doute le type de mesure le plus approprié, soit pour décider si l'étudiant doit entreprendre une nouvelle unité d'enseignement, soit pour lui prescrire un enseignement correctif s'il y a lieu.

L'évaluation sommative est une évaluation qui survient à la fin d'une étape importante dans la scolarité de l'étudiant comme la fin d'un cours, la fin d'un programme d'études ou d'un cycle. Contrairement à l'évaluation formative, l'évaluation sommative ne vise pas à dépister des difficultés d'apprentissage dans la perspective d'y apporter des correctifs. Elle a plutôt un caractère terminal et elle est souvent associée à des activités comme la certification, l'attestation de capacités, la promotion, le diplôme, etc. Les prises de décision reliées à l'évaluation sommative sont nombreuses et ne suivent pas nécessairement de façon immédiate le moment de l'évaluation. Par exemple, un étudiant est évalué à la fin de ses études de niveau secondaire, mais cette évaluation influencera éventuellement l'obtention d'un emploi quelques années plus tard, lorsqu'il se présentera dans le monde du travail. C'est pourquoi, dans le contexte de l'évaluation sommative, les institutions d'enseignement produisent des rapports d'évaluation qui sont transmis aux parents et aux administrateurs et qui permettent dans certains cas de différer la prise de décision. Nous serions tentés d'établir un lien nécessaire entre l'évaluation sommative et la mesure normative. Il est vrai que le contexte de certaines prises de décision qui font suite à une évaluation sommative commande habituellement une comparaison entre plusieurs étudiants, ce qui est du domaine de la mesure normative. Il n'est pas exclu cependant que le responsable d'un service d'admission ou qu'un employeur soient intéressés à connaître les capacités ou les habiletés maîtrisées par un étudiant, à la fin d'une étape importante de sa scolarité, dans le but de les comparer à celles qui sont jugées indispensables au succès dans un programme d'études ou dans une profession. Le contexte de la décision peut justifier l'emploi de mesures critériées dans l'évaluation sommative des étudiants. Enfin, le contenu de l'évaluation sommative implique des habiletés plutôt complexes et pouvant être évoquées dans une foule de situations variées. En évaluation formative, on s'intéresse à des habiletés qui sont particulières ou spécifiques à une petite unité d'enseignement et d'apprentissage. Cette dernière distinction entre l'évaluation sommative et l'évaluation formative n'est pas facile à établir dans la pratique, étant donné la complexité du domaine traitant des objectifs pédagogiques lorsqu'il s'agit de distinguer un objectif général d'un objectif spécifique.

Avant de terminer cet exposé, il nous reste deux termes à aborder qui peuvent faire l'objet d'une certaine confusion. Il s'agit de l'évaluation continue et de l'évaluation diagnostique. L'évaluation continue présente toutes les apparences de l'évaluation formative mais, avant de l'affirmer, il nous faut prendre certaines précautions. Quant à l'évaluation diagnostique, qui elle aussi semble s'apparenter à l'évaluation formative, elle est devenue, avec la création de cette dernière, une entité distincte.

À cause du caractère périodique des évaluations qui y sont réalisées, l'évaluation dite continue présente toutes les apparences extérieures de l'évaluation formative. Si on donne le nom d'évaluation continue à un système qui consiste à consigner et à accumuler plusieurs évaluations partielles dans la perspective d'une évaluation globale, on ne saurait qualifier de formative une telle évaluation. Il s'agit plutôt d'une modalité de l'évaluation sommative par laquelle l'évaluation finale est remplacée ou complétée par une somme d'évaluations partielles espacées dans le temps. Pour être considérée comme formative, l'évaluation dite continue doit être caractérisée par un ensemble de décisions à court terme qui concernent la qualité de l'apprentissage de chaque étudiant. De plus, l'accumulation d'évaluations partielles est une activité superflue en évaluation formative, étant donné que chaque évaluation est suivie immédiatement d'une prise de décision. Il demeure possible que, dans certains milieux, on veuille donner le nom d'évaluation continue à une forme d'évaluation véritablement formative. Seule une analyse des prises de décision nous permettrait de l'affirmer.

L'évaluation diagnostique se distingue de l'évaluation formative par l'objet du diagnostic et par le moment précis où elle est réalisée en relation avec un cours ou un programme d'études. Alors que l'évaluation formative a pour but de dépister des déficiences ou des difficultés éventuelles dans l'apprentissage, l'évaluation diagnostique a pour but d'expliquer ces déficiences ou ces difficultés au niveau des aptitudes, des intérêts ou de la motivation par exemple. L'objet de l'évaluation diagnostique est plus global que celui de l'évaluation formative. Les difficultés étudiées en évaluation formative sont spécifiques à un apprentissage donné. De plus, l'évaluation diag-

nostique est généralement réalisée au début d'une séquence importante d'enseignement, alors que l'évaluation formative s'étend sur toute la durée de cette séquence.

Conclusion

Dans cet article, nous avons tenté de définir et d'expliquer les principaux concepts de base qui s'inscrivent dans les approches les plus récentes à l'évaluation des étudiants. Nous avons voulu éviter la présentation sèche d'un vocabulaire technique et la spéculation sur des problèmes soulevés par l'évaluation des étudiants en milieu scolaire. Pour ce faire, nous avons abordé des modalités de mesure et d'évaluation plutôt que des techniques. De plus, nous avons exposé un point de vue fonctionnel de l'évaluation qui est basé sur la notion de prise de décision, tout en laissant le lecteur libre d'accepter ou non le bien-fondé de certaines prises de décision ●

Références

- Airasian, Peter W. et Madaus, George F., *Criterion-Referenced Testing in the Classroom*, Measurement in Education, National Council on Measurement in Education, Vol. 3, N° 4, 1972.
- Bloom, B.S., Hastings, J.T. et Madaus, G.F., *Handbook on Formative and Summative Evaluation of Student Learning*, New York: McGraw Hill Book Co., 1971.
- Popham, W. James, Editor, *Criterion-Referenced Measurement, an introduction*, Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.
- Sullivan, Howard J., « Objectives, Evaluation and Improved Learner Achievement », *Monographs Series on Curriculum Evaluation*, N° 3 (Instructional Objectives), Chicago: Rand McNally and Co., 1969, pp. 81-87.
- Vanhoven, James B., « Reporting Pupil Progress: A Broad Rationale for New Practices », *Phi Delta Kappan*, Vol. 53, N° 6, 1972, pages 365 et 366.