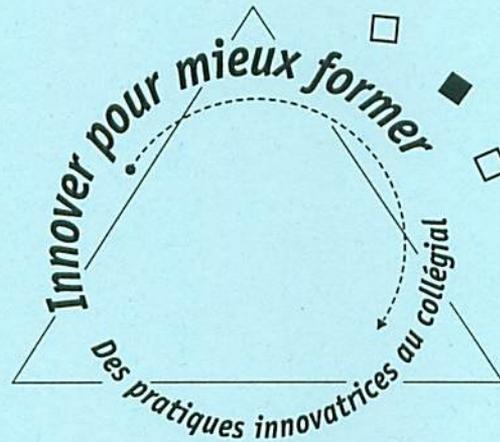


Actes du 17^e colloque de l'AQPC



5B30

**L'évaluation synthèse de programme :
ce n'est pas fini !**

Dany LAVEAULT, Ph. D.
Université d'Ottawa



**Association québécoise
de pédagogie collégiale**

L'ÉVALUATION SYNTHÈSE DE PROGRAMME : CE N'EST PAS FINI !

Dany LAVEAULT, Ph.D.
Université d'Ottawa

Que tous ceux et celles qui croient en avoir décousu avec l'épreuve synthèse se détrompent. Ce n'est pas parce que les premières épreuves synthèses ont été rédigées et notées qu'il sera maintenant possible d'entrer paisiblement dans une nouvelle routine de fin de session. D'abord, parce qu'il existe encore de nombreuses confusions en ce qui concerne l'ÉSP. Elles sont plus évidentes que jamais suite aux premiers choix qui ont été faits et mis à l'essai. Ensuite, parce pressés par l'urgence de produire l'ÉSP, peu de questions ont été soulevées quant aux multiples usages qui pourraient être faits des résultats par la suite. Cette question a été d'autant plus escamotée que l'enjeu "ÉSP" n'est pas totalement réglé dans ses impacts administratifs et de relations de travail. Aussi, tant ceux qui sont en faveur de l'ÉSP que ceux qui sont "tièdes" à son égard pour de multiples raisons, trouveront-ils dans cet exposé des motifs pour conclure qu'il y a encore un sacré boulot à accomplir.

Une chose semble sûre pourtant. Personne ne souhaite faire ce travail pour rien. Il serait dommage qu'au bout de cinq ans, après avoir évalué tant de finissant(e)s, nous n'en sachions guère plus sur la qualité des programmes d'études, sur la qualité de la formation ainsi que sur la préparation des élèves au marché du travail ou aux études universitaires. Pour ne pas être en retard d'une nouvelle réforme, c'est dès maintenant qu'il faut songer à réfléchir sur les choix actuels. La façon dont sont traités les résultats à l'ÉSP sont gages de son utilité dans l'avenir. D'où l'importance de mettre à profit ces premières expériences pour améliorer le tir.

Cette première mise à l'essai de l'ÉSP n'aura certainement pas été satisfaisante à tous les égards. Que devons nous faire pour la mettre à profit ?

1. L'analyse des premiers résultats peut nous conduire à réviser plusieurs aspects de l'outil : la procédure de correction, le système de notation, le contenu même de l'épreuve et les critères et les seuils employés.
2. Elle peut aussi nous aider à jeter un regard critique sur la formation dispensée à l'élève : cours offerts, niveaux des standards, programme d'étude, voire même les politiques d'admission. Dans une telle optique, l'ÉSP joue non seulement le rôle d'une évaluation sommative pour les élèves mais aussi celui d'évaluation formative pour les programmes d'étude.

Bref, c'est l'occasion de conforter certaines des décisions prises ou encore de réviser ces choix. D'ailleurs, plusieurs de ces choix résultent de compromis départementaux et d'équilibres précaires entre les différentes conceptions en présence. Plusieurs ont pu attendre cette première expérience pour éprouver leur choix ou celui des autres. Somme toute, le prototype d'ÉSP demande donc à être nourri des résultats de cette première mise à l'essai.

Conforter ou réviser nos choix

Tenir compte des conséquences de nos choix

Il existe entre les tests de dépistage de suicide et l'ÉSP une étrange parenté : tous deux possèdent un *taux de base* relativement élevé. Par exemple, si je devais prédire quel pourcentage de l'ensemble de la population risque de se suicider, je pourrais facilement prédire 0 % et je ne me tromperais que très rarement. En fait, le taux d'erreur serait le même que le taux de suicide. Le taux de base nous renseigne donc sur la probabilité qu'un événement se produise dans la population d'intérêt. Pour que le test soit utile et valide, il faut que le *taux de sélection*, c'est-à-dire la proportion des sujets que l'on retient suite aux résultats obtenus à une épreuve, améliore ce taux de base de façon significative. Or, dans le cas du suicide, la probabilité qu'un événement se produise est tellement faible qu'il est fort difficile de produire une épreuve capable d'améliorer une prédiction qui serait faite à partir d'un taux de sélection de 0 %!

Posons la question maintenant en ce qui concerne l'ÉSP. Quel *devrait* être le taux de sélection de l'ÉSP considérant que le taux de base dans la population est relativement élevé. Le taux de base serait ici la proportion des élèves qui ayant réussi tous leurs cours seraient assez compétents pour être admis dans une université ou encore qui feraient un travail acceptable du point de vue d'un novice sur le marché du travail.

Ce taux de base est sans doute fort élevé et c'est pourquoi le taux de sélection doit l'être aussi. Il est difficile de le connaître réellement, car il faudrait pour cela savoir le pourcentage des personnes qui ont les compétences pour réussir dans l'emploi ou le programme universitaire qu'ils ont choisi. En guise de compromis, nous pouvons nous rabattre sur ce qui est visé, donc conçu comme réaliste. En effet, si l'on prend en considération l'avis au ministre du Conseil supérieur de l'éducation (1992), on proposait alors de "viser à ce que d'ici l'an 2000, au moins 60 % de Québécois et des Québécoises aient obtenu un des diplômes d'études collégiales — de l'enseignement régulier ou du secteur adultes — avant l'âge de 25 ans — ...". Un tel objectif implique que le taux de base des élèves qui ont réussi tous leurs cours de formation collégiale soit très élevé : de l'ordre de 90 % et plus.

En effet, en décidant ou non d'accorder le diplôme sur base des résultats à l'ÉSP, il y a essentiellement trois situations qui peuvent se produire :

1. Les élèves qui ont les compétences pour réussir sur le marché du travail ou à l'université, sont choisis et ceux qui ne les ont pas sont rejetés. Le *taux de succès* de l'ÉSP se mesure donc au pourcentage des élèves promus qui ont effectivement les compétences pour réussir sur le marché du travail ou à l'université.
2. Les élèves qui n'ont pas les compétences nécessaires pour réussir sur le marché du travail ou à l'université et qui sont promus : on les nomme les *faux positifs*.
3. Les élèves qui ont les compétences nécessaires pour réussir sur le marché du travail ou à l'université et qui ne sont pas promus : on les nomme les *faux négatifs*.

À l'heure où l'objectif est d'accroître le taux de réussite dans les Collèges et de réduire les redoublements de cours, il est difficile d'envisager des taux de sélection qui ne seraient pas très élevés. D'autant plus qu'une sélection préalable a déjà été effectuée dans le cadre des cours. Alors qu'il est habituel de concevoir des pourcentages de réussite de 60 à 90 % dans les cours de première année, nous savons pertinemment que ces taux s'améliorent avec la progression des élèves dans le curriculum. D'ailleurs, le résultat est perceptible dans l'enseignement : enseigner à un groupe de 1^{ère} année, ce n'est pas la même chose qu'enseigner à un groupe de 2^e ou 3^e année.

Il est étrange que le taux de sélection attendu de l'ÉSP soit si peu discuté. En fait, tout indique qu'il devrait être relativement élevé : d'abord, parce que le taux de base est lui-même relativement élevé. Pourrait-on affirmer qu'il devrait être de l'ordre de 98 ou 99 % après exercice du droit de reprise ?

Nous devrions à tout le moins tenter de répondre à cette question avant l'entrée en fonction de l'ÉSP comme instrument d'attestation du D.E.C. Il existe une raison supplémentaire pour cela. Peu importe en effet le taux de base réel ou attendu dans la population, le taux de sélection a un impact sur le type d'erreur de décision que nous commettons. Souhaite-t-on minimiser l'erreur qui consiste à accorder le diplôme à un élève qui ne le mériterait pas ou l'erreur qui consiste à refuser le diplôme à quelqu'un qui le mériterait ? Sans doute, enseignants et élèves ont des points de vue différents. L'augmentation du taux de sélection réduit le risque de la première erreur, mais augmente le risque de la seconde : ce qui, on le comprend, pourrait ne pas faire l'affaire des élèves.

Alors, si une très forte proportion d'élèves qui s'inscrivent à une ÉSP sont censés la réussir, quelle est son utilité ? Pourquoi investir tant d'efforts pour discriminer entre élèves, alors que nous savons déjà à l'avance quel sera le résultat de la plupart d'entre eux et qu'il changera peu de choses au taux de succès du classement final ? Pour continuer le parallèle avec le taux de base du suicide, serions-nous en train d'utiliser un test de dépistage du suicide chez les abonnés réguliers des Club Med dont la probabilité de passer à l'acte, je suppose, frôle le 0.0000 % ?

Si l'on prend en ligne de compte les effets cumulatifs des taux de sélection dans chaque cours, puis à l'ÉSP, il serait difficile de justifier des taux à l'ÉSP qui ne seraient autrement que très élevés. Mais précisément, diront certains, ces taux sont trop élevés : ils permettent à trop d'élèves qui n'ont pas les compétences nécessaires d'obtenir le diplôme. Même si ces personnes peuvent avoir raison, il faut se poser la question de la transparence : serait-il adéquat d'exiger des taux de sélection inférieurs avec des politiques d'admission libérales et l'attestation année après année de la réussite cours par cours ?

Choisir la validité et la fidélité dont on a réellement besoin

Or, toute cette discussion autour du taux de base et du taux de sélection permet d'établir clairement une chose. L'ÉSP ne crée pas la situation idéale pour différencier entre eux tous les élèves. Avec un tel taux de sélection, elle ne peut certes pas donner des résultats valides pour tous les élèves. Et pourquoi d'ailleurs serait-ce si important ?

L'ÉSP n'est pas un *test de sélection*. Elle est conçue non pas pour différencier des personnes, mais pour différencier des compétences et attester de leur intégration. À quoi sert-il de savoir lequel de deux élèves ayant obtenu la même moyenne de 88 % dans l'ensemble de leurs cours obtiendra le résultat le plus élevé à l'ÉSP ? Il ne faut pas s'attendre à ce que le taux de réussite à l'ÉSP soit faible chez de tels individus. Est-il possible d'attester simplement que les deux maîtrisent et intègrent toutes les compétences essentielles ? Cet objectif d'évaluation ne peut s'atteindre qu'avec un instrument critérié qui discrimine bien au seuil des compétences minimales et non au seuil des compétences exceptionnelles. La validité et la fiabilité des résultats est davantage recherchée au niveau des compétences essentielles et au seuil minimal de compétences : non pas pour toutes les compétences et pour tous les seuils.

L'ÉSP doit être considérée comme une *épreuve de maîtrise*. Comme dans toutes les situations de jugement de maîtrise, il existe peu de variance dans les résultats des élèves. C'est précisément parce qu'une très large proportion d'entre eux maîtrise à un seuil minimum requis l'ensemble des compétences à évaluer. Or, s'il n'est pas important de discriminer entre les individus dits forts et ceux dits faibles, il est par contre essentiel de discriminer au seuil de maîtrise, soit entre ceux qui atteignent les standards minimaux et ceux qui ne les atteignent pas. Pour être valide, une ÉSP doit nous fournir des résultats qui permettent de discriminer adéquatement au seuil de maîtrise. Se tromper d'un ou deux points est bien plus grave aux environs de 60 qu'aux environs de 90 % ou de 30 %.

Développer des règles de décision valides et fiables

Au moment de noter l'élève, il n'y a rien de plus pratique qu'un chiffre. En plus de la précision apparente de la mesure, celle-ci permet de classer les élèves. Pour les tenants de la note chiffrée, il est primordial de reconnaître l'effort des meilleurs élèves. Or, si l'objectif n'est pas de classer les élèves, mais de différencier les compétences qui les caractérisent, il n'y a rien de plus ambigu qu'un score global chiffré et unique. En voulant souligner le mérite des meilleurs élèves et encourager ceux-là qui ont souvent le moins besoin d'être motivés, on risque d'occulter le véritable problème. Par exemple, si l'on doit mesurer l'élève sur plusieurs compétences, le classement devient difficile. Il est rare qu'un élève soit le meilleur (ou le moins bon) dans tout. Il faut alors trouver une formule de pondération compliquée pour classer les élèves : ce qui revient à ne plus différencier les compétences et à fondre les résultats dans une note globale.

Quadrature du cercle ? Ceci démontre qu'il est difficile et très onéreux de poursuivre deux objectifs d'évaluation à la fois : différencier les élèves et différencier les compétences. Le premier objectif n'est pas seulement difficile à réaliser, à cause du taux de sélection élevé, mais de plus contraire à l'esprit même de l'épreuve synthèse qui est de différencier les compétences au seuil de réussite. En se préoccupant trop du classement aux niveaux supérieurs, il y a risque de voir promus des individus qui ne le méritent pas (*faux positifs*) pas ou encore de ne pas accorder la promotion à des individus qui le méritent (*faux négatifs*). C'est ce qui risque de se produire si les efforts de notation sont dispersés et que l'on n'accorde pas plus d'attention aux critères de compétence et aux seuils minima de réussite.

Il existe un temps pour classer les élèves et un autre pour les certifier. À certains moments, il ne sert plus à rien de classer. Que diriez-vous d'inscrire les résultats d'examen sur les certificats de pratique médicale ? le rang obtenu par le médecin dans sa promotion ? D'inscrire sur les étiquettes de conserve, les résultats des propres tests de qualité de la compagnie qui les fabrique ? Que diriez-vous d'un test de dépistage du suicide qui mettrait en rangs les individus selon leur score total au test et qui enverrait automatiquement les cinq premiers à l'Hôpital psychiatrique ?

Parce que l'ÉSP se situe dans un contexte d'évaluation à partir de *standards* et non à partir de résultats de *mesure*, son approche est beaucoup plus qualitative. Et c'est sans doute très bien ainsi. Il y a déjà amplement de résultats de mesure quantitative dans le relevé de notes de l'élève. De plus, si l'on doit mesurer l'intégration des apprentissages faisant intervenir plusieurs compétences à la fois, il est difficile de ne pas faire appel à un *profil de résultats* plutôt qu'à un *résultat global*, étant donné la nature multidimensionnelle de la performance. Des règles de décision assez simples peuvent servir à attester de la réussite ou de l'échec.

Choisir les outils de décision appropriés : les seuils

Le seuils de réussite permettent de traduire en termes de décision concrètes notre appréciation, *qualitative et quantitative*, des résultats de l'élève. C'est aussi le moyen de fixer nos exigences et indirectement de déterminer le taux de sélection. Le niveau auquel nous fixons le ou les seuil(s) et les règles de décision que nous nous donnons, sont à la base même du taux de succès de l'ÉSP : des seuils pas assez exigeants et voilà que nous accordons la promotion à des personnes qui n'ont pas les compétences ; des seuils trop exigeants se traduiront par le refus du diplôme à des personnes qui réussiraient sur le marché de l'emploi ou l'université. D'où l'importance primordiale de s'entendre sur la validité et la fidélité des résultats au seuil de réussite.

Je n'entrerai pas dans le détail quant aux différentes procédures pour établir un seuil de réussite. Je crois qu'il est plus pertinent de répondre aux questions suivantes :

1. Y aura-t-il un ou plusieurs seuils ?
2. Dans le cas de seuils multiples, comment sera prise la décision finale ?
3. Faudra-t-il atteindre le seuil minimum dans toutes les compétences ou à certaines compétences seulement ?
4. L'atteinte du seuil dans une compétence peut-il compenser le fait que le seuil ne soit pas atteint dans une autre ?
5. Le niveau d'exigence sera-t-il uniforme pour toutes les compétences ou variera-t-il d'une compétence à l'autre ?

La réponse à toutes ces questions est essentielle pour se donner des règles de décisions claires. Mais, là encore, la modération a meilleur goût. Si la validité tend à s'accroître pour le plus grand nombre d'élèves si l'on multiplie le nombre de seuils, la fidélité, par contre, tend à diminuer. Même chose si l'on accroît indûment le nombre de critères à évaluer. Si la différence entre une maîtrise exceptionnelle et une maîtrise très bonne à la compétence #10 vous intéresse, l'augmentation du nombre de critères et du nombre de seuils est la voie qu'il vaut suivre. Mais une telle façon de faire va-t-elle améliorer la validité et la fidélité de la seule véritable décision importante à prendre : l'élève maîtrise-t-il ou ne maîtrise-t-il pas les compétences essentielles de son programme d'études ? La réponse risque d'être non.

Quelle note ou quel résultat communiquer ?

Le choix d'un système de notation doit être le reflet d'une conception de l'évaluation et non seulement répondre à des considérations pratiques. Quel système de note convient le mieux à la conception de l'évaluation sur laquelle repose l'ÉSP ?

En considérant un système de notation, deux types de renseignements sont importants :

1. choisir ce qui apparaîtra au bulletin ou au dossier de l'élève ;
2. choisir ce qui lui sera communiqué sur l'étendue de sa réussite.

Cette façon de procéder est courante. L'élève doit être prévenu à l'avance de ce que l'on attend de lui et des règles de décision qui seront suivies pour porter un jugement définitif sur sa performance. La note de cours qui apparaît au bulletin est la résultante d'une succession complexe de jugements et l'élève est généralement informé dans le détail des motifs qui lui ont mérité une appréciation plutôt qu'une autre. Il y a donc fort à parier que ce qui apparaîtra au bulletin, dans le cas de l'ÉSP, suivra le même procédé : le résultat final inscrit sera la résultante d'un processus d'évaluation plus détaillé dont l'élève aura pu prendre connaissance.

Il existe à ce sujet une discussion fort complète préparée par Robert Howe et parue dans le numéro de Mars 1997 de *Pédagogie collégiale* sur les avantages et inconvénients des systèmes de notation gradués et dichotomiques. Je ne vais donc pas revenir sur les pour et les contre relevés par l'auteur. Je voudrais faire valoir cependant comment un résultat dichotomique peut être tout aussi pertinent sinon plus qu'un résultat gradué, ce en quoi ma conclusion diffère de celle de l'auteur de l'article.

Voici sur quoi je base mes conclusions :

1. Il ne semble faire aucun doute que, pour plusieurs auteurs (Fortin et Raymond, 1996 ; Laliberté, 1995 ; Tremblay, 1994) l'ÉSP est une évaluation *sommative et critériée*. Il n'y a donc pas vraiment avantage à employer une note chiffrée ou encore une échelle graduée comportant un nombre restreint d'échelons. D'abord, parce que même s'il est important pour l'élève qui a raté l'ÉSP de savoir de combien il a raté la cible, il est encore plus important de l'informer sur les critères précis qu'il a ratés. Cela, une note chiffrée ou graduée ne l'assure pas.
2. Un autre argument veut qu'un score gradué rende mieux justice aux efforts des élèves vis-à-vis d'éventuels employeurs ou universités. Cette affirmation serait juste si l'on pouvait démontrer l'équivalence inter-institutionnelle et la continuité, d'une année à l'autre, des résultats. Or, rien de moins sûr. De fait, une note graduée comporterait selon moi plus de dangers que d'avantages. Le risque serait grand d'adopter une attitude centrée sur le score à l'ÉSP, qui éclipserait alors deux ou trois ans d'efforts continus dans les cours. Sur quelle base devrait-on préférer un élève qui a repris deux cours et réussit l'ÉSP avec 75 % dans un Cégep à un autre qui n'a jamais repris de cours et qui a réussi l'ÉSP avec 68 % dans un autre Cégep ?

3. Il est plus facile d'atteindre à l'équivalence inter-institutionnelle, à la fiabilité inter-correcteurs et à la stabilité d'années en années si les critères et les seuils sont clairement explicités et si le jugement se fait à partir d'une échelle dichotomique *réussite* ou *échec* qu'avec des échelles à plusieurs échelons ayant recours à des seuils multiples.
4. Enfin, insister pour discriminer entre élèves qui ont déjà réussi tous leurs cours s'avère une préoccupation qui risque de faire dévier l'objectif réel de l'ÉSP : discriminer, parmi les élèves, ceux qui ont réalisé un minimum d'intégration de leurs apprentissages de ceux qui ne l'ont pas fait. C'est au *seuil de maîtrise* que nous devons concentrer nos efforts.

Je ne crois pas que que l'ÉSP doive servir à recommander un finissant plutôt qu'un autre aux universités ou aux employeurs. Les comités de sélection des employeurs et les comités d'admission des universités poursuivent des objectifs qui leur sont propres. Pourquoi un employeur ne privilégierait-il pas les résultats à des cours bien précis, plus directement reliés aux besoins de l'emploi, que les résultats à l'ÉSP ? Nous devons former des finissant(e)s qui répondront à des exigences d'emploi variées et qui seront capables de s'adapter aux conditions changeantes du marché. Nous devons attester de leurs compétences à l'intérieur d'un large spectre.

Quelle fonction ? sommative, formative ou les deux ?

Alors que la fonction sommative de l'ÉSP dans l'évaluation des élèves ne fait aucun doute, sa fonction formative est moins bien définie. Certains sont d'avis que la préparation à l'ÉSP ne doit pas attendre au dernier trimestre : après tout si l'intégration des apprentissages doit être évaluée à la fin des études, c'est tout au cours de celle-ci qu'elle se prépare. D'autres sont d'avis que le résultat à l'ÉSP doit également fournir une sorte de rétroaction pour permettre à l'élève de savoir ce qu'il a ou n'a pas réussi. Ce souci est particulièrement justifié dans le cas de ceux qui échouent. Avant de se présenter à nouveau à l'ÉSP, ils devraient recevoir une forme de rétroaction qui indique les critères qu'ils n'ont pas atteints et ce qu'ils doivent entreprendre pour réussir la fois suivante. Mais, il ne faut pas oublier que l'ÉSP n'est pas conçue prioritairement comme un instrument d'évaluation formative des élèves, même si, pour ceux qui ratent, le retour sur les résultats obtenus pourrait donner lieu à une certaine forme de rétroaction.

Le rôle de la fonction formative de l'ÉSP dans l'évaluation des programmes et des cours est encore moins claire que dans le cas des élèves. Il serait dommage en effet d'évaluer tous ces finissant(e)s d'année en année en fermant les yeux sur les faiblesses qui se répètent et en n'intervenant pas sur les éléments du programme d'études ou de la coordination des enseignements qui pourraient améliorer la situation. Dans ce sens, l'ÉSP peut jouer un rôle formateur utile dans l'évaluation des programmes, car si les élèves ne sont pas infaillibles, les programmes d'études et les prestations de cours ne le sont pas non plus.

Il n'est donc pas surprenant de retrouver cette fonction de l'ÉSP dans *le Guide général pour les évaluations des programmes d'études réalisés par la Commission d'évaluation de l'enseignement collégial* (1994). L'ÉSP y est mentionnée à deux reprises comme intervenant dans l'évaluation de l'efficacité du programme d'étude (critère 5). En effet, au sous-critère 5.4 on peut lire :

Une proportion acceptable des étudiants et étudiantes termine le programme dans des délais acceptables, compte tenu de leur régime d'études et de leurs caractéristiques.

Or, l'ÉSP joue un rôle direct dans l'appréciation de ce critère puisque c'est avec sa réussite que prend fin la scolarité de l'élève. Mais, c'est surtout au sous-critère 5.5 qu'il est davantage question de l'ÉSP. Ce sous-critère est le suivant :

Les diplômées et diplômés satisfont aux standards convenus en ce qui regarde l'acquisition des diverses compétences établies pour le programme.

En particulier, ce sous-critère soulèvera trois questions d'appréciation auxquelles on demandera aux institutions de répondre :

Dans quelle mesure les diplômés du programme atteignent-ils effectivement les objectifs et les standards prévus au programme? Comment l'établissement procède-t-il pour conclure qu'ils ont acquis les compétences attendues?

Dans quelle mesure l'épreuve de synthèse administrée par l'établissement permet-elle d'attester que les étudiantes et les étudiants qui terminent le programme satisfont aux standards établis pour mesurer l'atteinte des compétences visées?

L'établissement a-t-il développé des moyens en vue de s'assurer que le calibre de son épreuve de synthèse est équivalent à celui d'autres collèges?

Il faut prendre garde cependant à ce que l'ÉSP n'intervienne dans l'évaluation sommative, non pas des élèves cette fois, mais des programmes et des enseignants. Pour comparer les établissements équitablement, il faudrait comparer non seulement les profils de sortie des finissants, mais aussi les profils d'entrée. Or, à quoi nous servirait-il d'apprendre que les Cégeps qui ont les meilleurs taux de succès sont ceux qui ont été en mesure de sélectionner leur clientèle en refusant un plus grand nombre d'élèves ? Même à cela, il faudrait s'assurer de l'équité dans les différents environnements dans lesquels les Cégeps sont appelés à évoluer. Enfin, et c'est sans doute la raison majeure, de telles comparaisons ne nous apprendraient rien sur ce qu'il faudrait faire localement pour améliorer les conditions de réussite.

Or, même si l'équivalence institutionnelle est impossible à atteindre au niveau des résultats, elle est réalisable au niveau des exigences. Il est important de noter que la question d'appréciation du sous-critère 5.5 de la commission d'évaluation demande si l'établissement s'est assuré que le *calibre* de son ÉSP est équivalent à celui des autres collèges : on ne parle pas des résultats, et pour cause !

Pour mieux préparer l'avenir

Considérer la fonction formative de l'ÉSP

Pour revenir à la fonction formative de l'ÉSP dans l'évaluation des programmes, mentionnons que cette considération n'est pas sans rapport avec le type d'évaluation et le système de notation choisi. Il peut être intéressant pour une même institution de s'auto-évaluer régulièrement et d'étudier comment les changements et les innovations apportées se traduisent dans les réussites des élèves. En ce sens, la façon dont les résultats sont recueillis en rapport avec l'ÉSP auront des conséquences directes sur la possibilité de tirer des conclusions sur l'évolution des programmes et de leurs prestations.

L'évaluation formative ne sert pas qu'à donner une rétroaction à l'élève sur son degré de réussite et à assurer la régulation de ses apprentissages. Dans le cas de l'élève, parvenu à la fin de son programme d'études, cette régulation arrive trop tard. Par contre, l'objet de la régulation peut aussi bien être l'enseignement et le programme d'études. Pour cela, il n'est pas trop tard en ce qui concerne l'ÉSP.

À quoi nous servirait-il en 2003 de savoir qu'au Cégep X dans le programme Z, la moyenne des résultats des finissants à l'ÉSP est passée de 74 % à 79 % plutôt que de savoir que la proportion de ceux qui ont passé a augmenté de 95 % à 98 %. Un résultat nous renseigne-t-il plus que l'autre ? Sans doute non, si l'on ne fait qu'une lecture superficielle de ce compte rendu. Pour l'interpréter, pour que ces changements aient un véritable impact sur la pratique, il faut pouvoir les situer par rapport à des critères bien précis. Or, à quoi attribuer l'amélioration perçue : aux conditions d'admission plus resserrées, à une plus grande sévérité au niveau des cours, aux améliorations du programme d'étude, à l'augmentation du financement, à de meilleurs cours ou enfin à la diminution des exigences à l'ÉSP ? Ce ne sont là que quelques-uns des facteurs qui pourraient expliquer une augmentation et ni l'une ni l'autre procédure de notation n'y changerait quoique ce soit.

Il serait dangereux d'interpréter les résultats à l'épreuve synthèse hors du contexte particulier à chaque établissement. C'est sans doute pourquoi ceux-ci sont le plus à même d'apporter une réponse aux questions soulevées. Encore faut-il que pour cela ils disposent de bons renseignements.

Rien n'empêche a priori de noter l'élève sur un profil de compétences essentielles dont la réussite à chacune serait nécessaire à la réussite totale. Des profils de résultats par groupes-critères sont souvent remis aux élèves, même si ce n'est que le résultat total qui figure au bulletin. Il faudrait peut être les conserver, car ceux-ci permettraient de comparer à chaque année quelle(s) compétence(s) sont les principales causes d'échec. Si le taux d'échec par compétence est le même (p.e. 1 à 2 %), il n'y a pas de raison de suspecter un problème particulier. Si une compétence particulière ou groupe de compétences devait faire défaut d'année en année, alors une étude plus approfondie serait justifiée. Le programme prépare-t-il bien les élèves à ces compétences ? Pourquoi ces compétences sont-elles moins bien réussies ? Y a-t-il un lien entre les exigences des cours et ce résultat ? Les conditions des élèves admis expliquent-elles ce résultat ? Le problème se situe-t-il davantage au niveau des compétences requérant la plus grande intégration ? etc... On peut se rendre compte facilement que de tels résultats qualitatifs auraient une fonction formative utile pour les programmes d'étude. Chiffrée ou pas, le résultat ici n'est pas vraiment important. Pour jouer un rôle véritablement formatif, le résultat doit être rapporté à une série de critères.

Guider l'action en fonction de principes réalistes et réalisables

Vis-à-vis l'ÉSP, il est facile de pécher par action ou par omission. On peut vouloir lui faire accomplir trop de choses ou pas assez. On peut aussi en faire trop sans que cela ne soit vraiment nécessaire et ne pas en faire assez alors que c'est réellement utile. Le tout repose sur une série de choix stratégiques, qui deviendront de plus en plus clairs au fur et à mesure de la progression. Voici quelques décisions à revoir quant à l'ÉSP.

Vous devez notamment choisir entre :

1. différencier les élèves ou les compétences ;
2. différencier les élèves à tous les niveaux ou au seuil de maîtrise seulement ;
3. utiliser ou ne pas utiliser les résultats de l'ÉSP à des fins formative pour l'évaluation des programmes d'études ;
4. noter sur des échelles dichotomiques ou graduées ;
5. trouver un équilibre entre nombre de critères et nombre de seuils suffisants.

Pour ma part, je considère que l'ÉSP doit d'abord différencier les compétences. C'est pourquoi je privilégie une approche critériée et l'évaluation au moyen de profils de scores. Je préfère évaluer plusieurs compétences *réussite* ou *échec* plutôt qu'un nombre restreint de compétences sur des échelons qui n'en finissent plus du genre : *échec, passable, satisfaisant, très satisfaisant et exceptionnel*. Dans le premier cas, les résultats sont habituellement plus fiables et plus faciles à interpréter car il n'y a qu'un seul seuil de réussite à déterminer. Dans le second cas, beaucoup de temps et d'énergie sont consacrées à établir un consensus sur ce qui est exceptionnel, alors que là n'est pas le véritable enjeu de l'ÉSP.

Il serait dommage de consacrer beaucoup de temps à la correction des travaux des élèves sans que cela ne change quoique ce soit à notre façon d'enseigner. C'est pourquoi je crois aussi à une certaine utilisation de l'ÉSP à des fins formatives à condition de bien interpréter les résultats en fonction du contexte prévalent dans chaque institution et dans chaque département. Il serait injuste de blâmer ceux qui acceptent de travailler dans des conditions difficiles à partir des résultats obtenus par les élèves à l'ÉSP.

Modération des résultats

L'analyse des résultats peut servir à la modération des notes, au cas où l'on se rendrait compte que la procédure utilisée comporte certaines lacunes. Pour des raisons d'équité, il est normal, avec un instrument que l'on utilise pour la première fois, d'apporter après coup certaines modifications de façon à s'assurer que les résultats reflètent bien les apprentissages réels des élèves. Cette analyse est l'équivalent de l'analyse d'items traditionnelle et il serait dommage de s'en passer.

Il serait utopique de croire que les ÉSP feront mouche du premier coup. Des tâtonnements et des révisions successives seront nécessaires. Il n'y a pas de mal à cela. N'oublions pas que les scores aux examens uniformes du Ministère de l'éducation sont eux-mêmes convertis et modérés. Une modération similaire devra être faite au niveau de l'ÉSP :

1. au niveau de la concordance inter-correcteurs ;
2. au niveau des épreuves elles-mêmes ;
3. au niveau des grilles de correction et des systèmes de notation ;
4. au niveau des exigences elles-mêmes : que ce soit en rapport avec les critères ou le seuil de réussite.

Cette modération sera d'autant plus nécessaire que les programmes d'études changent rapidement et qu'il faudra compter un certain temps pour qu'enseignants et élèves s'y habituent. Le résultat de l'ÉSP pouvant refléter ce ballottage, il "est important, dans un souci d'équité, d'ajuster l'ÉSP pour qu'elle soit le reflet de ce que l'élève a vraiment eu l'occasion d'apprendre durant ses études collégiales" (Allaire, 1996).

Conclusion

La raison d'être de l'ÉSP n'est pas de savoir quels élèves en savent le plus, mais plutôt d'en savoir plus sur chaque élève, en particulier, ceux dont la réussite est incertaine. L'ÉSP possède toutes les caractéristiques d'un test critérié à vocation sommative pour l'élève et formative pour les enseignants et les programmes d'études. En effet, l'ÉSP correspond aux finalités qui sont à l'origine des tests critériés en éducation. On peut en effet lire dans Legendre (1993) que :

Le test critérié a été proposé par Glasser en 1963 afin que l'éducation soit plus axée sur l'atteinte des objectifs par chacun des élèves que sur la sélection. Il se situe donc dans une optique de pédagogie de la réussite plutôt que dans celle de la discrimination ou de l'échec pour bon nombre de sujets.

On pourrait relire la même phrase en changeant "test critérié" par "ÉSP", "Glasser" par "Commission d'évaluation" et "1963" par "1994". Mais avant de pouvoir accomplir une si noble mission, l'ÉSP devra passer à nouveau par de nombreux essais empiriques, expression généreuse pour dire "tâtonnements". Mais ces tâtonnements seront nécessaires pour que l'ÉSP ne fasse pas que mesurer l'intégration des apprentissages, mais pour qu'elle s'intègre elle-même aux programmes d'étude, ce qui est loin d'être le cas présentement.

Plusieurs utiliseront même ce prétexte pour s'opposer à l'ÉSP. Mais la réforme de l'enseignement collégial aurait-elle été mieux servie par des changements encore plus grands des programmes d'études, sans avoir eu l'occasion de réfléchir aux profils de sortie des finissants et à nos propres exigences ? Ou encore, par des directives encore plus détaillées concernant l'ÉSP ? Il s'agit d'un processus itératif : les changements de programmes d'études ne pouvant intervenir avant sans que soit développée une connaissance adéquate des profils de sortie des finissants. Et, le profil de sortie de nos finissants ne pourra être évalué adéquatement, particulièrement en ce qui concerne la dimension intégration des apprentissages, tant que les programmes d'étude n'auront pas connu de véritable reformulation. C'est le problème de la poule et de l'oeuf.

L'un des mérites principal de l'ÉSP est d'avoir amené les enseignants à se concerter davantage : tant au niveau du contenu des cours que de leurs exigences respectives. Il n'est pas juste que les chances de réussite d'un élève dépendent des fluctuations des exigences des enseignants. L'ÉSP favorise l'émergence d'une "culture d'évaluation" qui suscite ces échanges, même si elle ne garantit pas les résultats. Mais quelle serait l'alternative ? Une directive supérieure imposant le contenu des ÉSP et les niveaux d'exigence cours par cours ?

Si l'on envisage le développement de l'ÉSP sur une période plus longue, il est fort possible que les réflexions qu'elle a déjà provoquées à l'intérieur des établissements se poursuivent maintenant entre eux. Cette consultation entre les établissements deviendra nécessaire lorsque la Commission évaluera les programmes d'études et demandera, à chaque institution, de démontrer comment le calibre de ses ÉSP se compare à celui d'autres ÉSP semblables dans d'autres institutions. Les enseignants devront donc démontrer en quoi leurs exigences se comparent à celles des autres institutions et leur sont équivalentes. L'ÉSP accomplirait alors la deuxième partie de sa mission : assurer l'équivalence inter-institutionnelle *au niveau des exigences*.

Il n'est pas plus équitable d'obtenir son D.E.C. parce que c'est plus facile dans un Cégep que de réussir un cours parce que les exigences d'un enseignant sont moins rigoureuses. On ne peut exiger des établissements, ni des enseignants des résultats équivalents : ceux-ci dépendent de trop nombreux facteurs hors de leur contrôle immédiat. Mais, en ce qui concerne les exigences de l'ÉSP, il est raisonnable de s'attendre à ce que chaque établissement puisse démontrer qu'il n'est ni une auberge espagnole, où tout le monde peut entrer, ni une tour d'ivoire d'où on ne peut sortir. Ainsi, l'équivalence inter-institutionnelle serait atteinte, non pas en démontrant que les résultats sont identiques partout, mais en faisant valoir que les exigences sont raisonnablement les mêmes dans chaque établissement collégial.

Mais tout ceci demande du temps : pour que les enseignants approfondissent leur compréhension du programme, qu'ils révisent leurs cours et mettent en place des nouveaux, et enfin, qu'ils tirent les leçons des résultats de l'ÉSP. Raison de plus pour se fixer des objectifs réalisables. Car, comme dit le dicton, "Qui trop embrasse, mal étirent".

RÉFÉRENCES

Allaire, H. (1996). Précisions sur le profil de sortie. *Pédagogie collégiale*, 10(2), 13.

Commission d'évaluation de l'enseignement collégial (1994). *Guide général pour les évaluations des programmes d'études réalisées par la Commission d'évaluation de l'enseignement collégial*. Gouvernement du Québec : auteur (réf. : 2410-0508).

Conseil supérieur de l'éducation (1992). *L'enseignement supérieur : pour une entrée réussie dans le XXI^e siècle*. Gouvernement du Québec : auteur.

Crocker, L.M. et Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando : Harcourt, Brace Jovanovich College Publishers.

Fortin, L. et Raymond, D. (1996). Une définition de l'épreuve synthèse de programme. *Pédagogie collégiale*, 10(2), 19-20.

Howe, R. (1997). La présentation des résultats à l'épreuve synthèse de programme : une question de forme... et de fond. *Pédagogie collégiale*, 10(3), 25-29.

Laliberté, J. (1995). L'épreuve-synthèse de programme : gage ou facteur de l'intégration des apprentissages ? *Pédagogie collégiale*, 8(3), 18-23.

Legendre, R. (1993). *Dictionnaire actuel de l'éducation*. Montréal : Guérin.

Tremblay, G. (1994). *L'épreuve-synthèse de programme : testing ou évaluation ? Actes du Colloque de la Commission des affaires pédagogiques*. Montréal : Fédération des Cégeps, 30 novembre.