

Effets de l'entraînement des élèves à des pratiques évaluatives sur leur habileté démontrée par la suite en auto-évaluation

Jeannine Lavoie-Sirois
université laval

Cet article rapporte les résultats partiels d'une étude sur l'auto-évaluation réalisée auprès de 194 élèves du secteur professionnel de la Région de Québec. Différentes modalités d'entraînement des élèves ont été appliquées au cours de deux expérimentations indépendantes dans le but d'étudier leur effet sur l'habileté des élèves à s'auto-évaluer. Avec les données et dans le contexte particulier de l'étude, aucun lien n'a été observé entre les scores obtenus par les élèves en auto-évaluation et l'entraînement préalable qu'ils avaient reçu pour apprendre à évaluer.

A study of 194 Quebec vocational education pupils found no link between capacity of self-evaluation, and previous training in methods of evaluation—even though students had been trained in various evaluation methods in two separate experimental groups.

Cet article rapporte la seconde tranche des résultats d'une étude sur l'auto-évaluation réalisée auprès d'élèves du secteur professionnel de la région de Québec. D'autres résultats de l'étude (Lavoie-Sirois, 1991) ont déjà permis d'établir la capacité d'auto-évaluation de ces élèves de même que les effets bénéfiques des pratiques auto-évaluatives sur l'apprentissage. Nous rapportons ici les résultats obtenus dans le volet des travaux qui portait sur la relation entre certaines modalités d'entraînement des élèves à l'évaluation et leur habileté démontrée par la suite dans un exercice d'auto-évaluation.

PROBLÉMATIQUE

La pratique de l'auto-évaluation à l'école est une des formes d'appropriation de l'évaluation formative faisant appel à la participation de l'élève (Scallon, 1988). À l'instar de Holec (1981), on peut définir l'auto-évaluation comme une démarche au cours de laquelle une personne observe sa propre performance, l'analyse à la lumière d'une performance témoin et interprète les résultats observés, le tout évidemment dans une situation de relative autonomie (Scallon, 1988). Pour sa part, Allal s'inspire des théories cognitives (dont celles de Miller, Galanter et Pribam, 1960) dans sa conception de

l'auto-évaluation; pour l'auteure, "l'auto-évaluation intervient à la suite d'une action du sujet lorsque celui-ci cherche à situer le résultat de son action par rapport au but visé" (Allal, Baeriswyl, Mai et Wegmuller, 1987, p. 51).

Il faut distinguer l'auto-évaluation de l'autocorrection. De façon générale, l'autocorrection sollicite la participation de l'élève pour la notation de ses travaux dans un exercice qui l'amène à comparer lui-même ses réponses ou solutions à celles que lui fournit l'enseignant, souvent par le biais d'un matériel informatif didactique. Pour Paquay, Allal et Laveault (1990), l'autocorrection pratiquée dans les situations classiques d'apprentissage scolaire apparaît comme une réduction du concept "auto-évaluation." Pour ces auteurs, l'auto-évaluation prend un sens plus large et peut signifier "l'appréciation selon plusieurs critères du degré d'adéquation d'une production complexe au moment de son élaboration" (1990, p. 7). Dans une démarche auto-évaluative, l'élève est donc mis en situation de vérifier si une production plus ou moins complexe qu'il a réalisée répond à des critères particuliers de performance. Si l'autocorrection se fonde sur la capacité de l'élève d'établir des comparaisons de réponses ou de solutions bien définies qui ne font ordinairement pas appel à l'interprétation, l'auto-évaluation repose par ailleurs sur son habileté à estimer jusqu'à quel point une performance qu'il a démontrée répond à certains indicateurs ou descripteurs. De façon générale, l'approche autocorrective est indiquée quand on veut amener l'élève à vérifier de façon objective son degré d'acquisition de connaissances ou sa maîtrise de concepts et de règles simples. Par ailleurs, si on veut plutôt que l'élève porte un jugement sur la performance qu'il a pu démontrer dans une tâche dite "complexe," c'est-à-dire dans une tâche où la réponse produite est beaucoup plus élaborée qu'une réponse brève ou un choix de réponse (Scallon, 1988), il faut alors recourir à une approche qui fait appel à l'auto-appréciation de critères de performances. Toujours selon Scallon, "par rapport aux questions à correction objective, les tâches à réaliser en matière de production complexe placent l'élève en situation de relative autonomie, puisqu'il lui faut habituellement structurer une réponse ou élaborer un produit avec un certain degré de liberté tout en respectant certaines contraintes" (1988, Tome 2, p. 150). Dans le cadre d'activités réalisées à des fins d'évaluation formative, qu'il soit en situation d'autocorrection ou d'auto-évaluation, l'élève est invité par ailleurs à utiliser un système quelconque "d'auto-notation" qui lui permet d'indiquer les points forts comme les points faibles de la tâche effectuée (indication des réponses réussies ou échouées, établissement du degré de réussite ou des points de difficultés à l'aide de codes, comparaison du score obtenu à un seuil pré-établi de réussite, expression des jugements portés en fonction de critères précis de performance, etc). L'auto-notation dont il est question ici diffère évidemment de la notation scolaire proprement dite à laquelle l'enseignant recourt pour établir la valeur d'un apprentissage en vue du bulletin.

Certains auteurs font ressortir de façon intéressante le bien-fondé de la participation de l'élève à l'évaluation de ses propres performances. Pour Schwartz (1973), l'auto-évaluation donne à l'élève la possibilité de se situer lui-même en permanence et d'analyser ses propres erreurs. Bonniol croit par ailleurs que la modalité majeure de l'évaluation formative est l'auto-évaluation formative "préparée et pilotée par l'enseignant" (1981, p. 461). Pour cet auteur, l'auto-évaluation permet à l'élève de corriger le mécanisme responsable des erreurs commises plutôt que de corriger simplement un résultat. Cardinet (1986) explique pour sa part que puisque l'école doit préparer l'enfant à continuer sa formation de façon autonome tout au long de sa vie, un de ses rôles est de lui apprendre à s'auto-évaluer.

Aux yeux de plusieurs auteurs, l'auto-évaluation apparaît comme une habileté qui se développe et qui peut même constituer un objectif important de formation. De l'avis de Dominice (1981), l'auto-évaluation se rattache en finalité aux grands objectifs d'autonomie et de sens des responsabilités de l'éducation permanente. Pour Scallon (1988), l'élève peut et devrait être amené graduellement à prendre une part active dans l'évaluation de ses apprentissages et la capacité de s'auto-évaluer devrait se poser comme un objectif de formation. Leselbaum (1982) soutient pour sa part qu'en appliquant des procédures d'auto-évaluation, l'élève apprend à porter un jugement sur la valeur de son propre travail ou du travail d'un pair en fonction de critères établis par le maître et acceptés par l'élève. De son côté, Boisvert (1985) souligne l'importance d'un apprentissage progressif de l'habileté à s'auto-évaluer, l'élève devant apprendre graduellement à passer d'un jugement plus général à une utilisation de critères différenciés. Au regard des conditions qui peuvent favoriser l'apprentissage de l'auto-évaluation, Holec (1981), Boisvert (1985) et Fontaine (1988) partagent enfin l'avis que l'élève développera d'autant mieux son habileté à auto-évaluer qu'il démontrera une attitude positive face à l'évaluation et comprendra bien le rôle de l'évaluation formative.

Si l'auto-évaluation est une habileté que l'élève peut acquérir ou développer, il faut alors accorder de l'importance aux moyens ou aux approches qui peuvent favoriser un tel apprentissage ou un tel développement. Pour Boisvert (1985), le jeu, les loisirs, la famille et l'école sont des circonstances qui permettent à l'enfant d'être confronté aux résultats de ses actions, et donc, d'apprendre à évaluer ces dernières. Et d'après Levesque (1982), si l'auto-évaluation de l'élève peut en plus être confrontée à une évaluation extérieure, l'habileté à s'auto-évaluer est alors susceptible de se développer.

La référence à des critères précis est toutefois requise pour assurer l'efficacité de l'évaluation ou de l'auto-évaluation (Bloom, Madaus, & Hastings, 1981; Boisvert, 1985; Bonniol, 1981, 1984; Broadbent, 1975; Scallon, 1988; Tulving et Pearlstone, 1966). Même, pour Leon (1977), les critères d'évaluation revêtent une telle importance qu'il faut envisager de planifier leur apprentissage spécifique avant la phase d'exécution d'une tâche. Il apparaît donc que la connaissance des critères d'évaluation est

nécessaire pour rendre opératoire une démarche d'auto-évaluation demandée aux élèves, et même, qu'un apprentissage de ces critères devrait précéder la réalisation de la tâche. On peut même penser que la progression dans l'apprentissage de l'auto-évaluation devrait éventuellement déboucher sur la capacité de l'élève d'établir lui-même les critères qui lui permettront de juger de sa performance dans une tâche donnée. D'après Paquay, Allal et Laveault, "au départ, l'apprenant se servirait d'un instrument, puis il en intérioriserait progressivement la procédure; il apprendrait ensuite seulement à se faire ses propres instruments" (1990, p. 14).

Dans cette étude, on a voulu obtenir des réponses à certaines interrogations relatives aux facteurs ou aux conditions susceptibles d'influencer l'habileté des élèves en auto-évaluation. Plus précisément, on a voulu savoir s'il existe un lien entre l'entraînement de l'élève à évaluer des tâches complexes à partir de critères précis et l'habileté qu'il démontre par la suite dans l'auto-évaluation d'une tâche d'un niveau relativement équivalent de complexité. On a également voulu savoir si certaines modalités d'entraînement à l'évaluation faite à partir de critères précis s'avéraient plus efficaces que d'autres pour amener l'élève à mieux performer dans une tâche auto-évaluative. De façon complémentaire, les données de nos travaux ont aussi permis de vérifier si l'habileté démontrée par un élève en auto-évaluation est reliée à des variables telles que la motivation aux études et les résultats scolaires. Les trois hypothèses vérifiées dans cette recherche étaient donc les suivantes:

H1: Les élèves entraînés à utiliser des critères précis pour évaluer la performance d'une autre personne qui exécute une tâche complexe démontrent par la suite une meilleure habileté en auto-évaluation que les élèves n'ayant pas reçu ce type d'entraînement.

H2: Les élèves entraînés à la pratique double de l'évaluation et de l'auto-évaluation de tâches complexes à partir de critères précis démontrent par la suite une meilleure habileté en auto-évaluation que les élèves qui n'ont été entraînés qu'à l'une ou l'autre de ces deux pratiques.

H3: La motivation aux études et les résultats scolaires sont des variables reliées à l'habileté démontrée par les élèves en auto-évaluation.

MÉTHODOLOGIE ET EXPÉRIMENTATIONS

Les sujets de l'étude et les temps d'expérimentation

Au total, 13 enseignants du secteur professionnel de la région de Québec et 194 de leurs élèves ont participé à l'une ou l'autre des deux expérimentations indépendantes de la recherche. Le niveau d'études des groupes-classes participants était de cinquième ou de sixième secondaire et l'âge moyen des

sujets était de 21 ans (moyenne relativement élevée due à la présence d'étudiants adultes dans les groupes-classes). Les métiers concernés étaient les suivants: coiffure, commerce et secrétariat, couture, plomberie et soins infirmiers. L'hypothèse H1 a été vérifiée au cours de la première expérimentation tandis que les hypothèses H2 et H3 l'ont été au cours de la seconde.

La démarche expérimentale appliquée

Au cours de chacune des deux expérimentations, après une période qui permettait à certains des sujets observés de recevoir une forme particulière d'entraînement à l'évaluation, chaque élève de l'étude devait réaliser une tâche complexe au cours de laquelle sa performance était évaluée individuellement par l'enseignant à l'aide d'un instrument développé expressément à cette fin. Après l'exécution de la tâche, l'élève procédait à son tour à l'auto-évaluation de sa performance à l'aide du même instrument. Cette auto-évaluation étant réalisée dans le cadre d'une recherche et à des fins formatives, l'élève était informé que les résultats de l'exercice ne devaient pas être comptabilisés en vue du bulletin. Pour éviter toute contamination du jugement de l'enseignant sur celui de l'élève, ce dernier n'était pas informé des résultats de l'évaluation de l'enseignant après la réalisation de la tâche.

Les tâches évaluées

Pour déterminer la tâche qui devait être évaluée à la fin de la période d'entraînement, chaque enseignant a d'abord choisi dans son programme d'études un objectif spécifique portant sur l'apprentissage d'une procédure complexe. Avec l'aide ponctuelle du personnel engagé dans la recherche ou de spécialistes de la discipline, chaque enseignant a ensuite procédé à l'analyse de tâche de l'objectif ciblé pour préciser les opérations composantes de sa procédure d'exécution. Chacune des analyses effectuées a été développée selon une approche inspirée de la démarche algorithmique de Gagné et Briggs (1979) mais ramenée à une présentation de modèle "liste" telle qu'expérimentée depuis quelques années dans des activités de perfectionnement didactique offertes à des maîtres du secteur professionnel à l'Université Laval (Lavoie-Sirois, 1983). L'analyse de tâche visait à assurer une certaine validité de contenu et une organisation plus rigoureuse des éléments qui seraient retenus par la suite dans les instruments d'évaluation.

Les instruments d'évaluation

La "liste de vérification" (check-list) est le type d'instrument qui a été retenu pour réaliser tant l'évaluation que l'auto-évaluation de la performance des élèves dans l'étude. Cet instrument consiste en une série d'énoncés-questions précis sur les opérations séquentielles d'une tâche, chaque énoncé-question portant sur un seul élément de la procédure et appelant une réponse

à caractère dichotomique (oui/non, réalisé/non réalisé). En se guidant sur l'analyse de tâche effectuée au préalable, chaque enseignant, assisté du personnel de l'étude, a construit la liste de vérification devant permettre d'évaluer la performance de ses élèves dans l'exécution de la tâche ciblée. Chacune des opérations précises énumérées dans la liste de vérification pouvait constituer soit un énoncé d'ordre factuel, soit un énoncé-jugement selon les appellations retenues par Scallon (1988). Au total, 13 listes de vérification (comportant en moyenne entre 12 et 15 énoncés-questions) furent développées à raison d'une liste par groupe-classe participant. Chaque liste de vérification a été analysée et au besoin corrigée par des spécialistes en confection d'instruments de mesure avant d'être utilisée dans l'étude. Parmi les procédures évaluées dans nos travaux, mentionnons: la pose d'un passepoil (couture), le prélèvement de sang par micro-méthode (soins infirmiers), l'application d'eau oxygénée sur une chevelure (coiffure), la création d'un panorama à partir de deux fichiers (commerce et secrétariat).

Vérification de l'hypothèse H1

L'expérimentation I a été l'occasion de vérifier si des élèves entraînés de façon intensive à recourir à des critères précis pour évaluer la performance d'une autre personne démontraient par la suite une plus grande habileté en auto-évaluation que les élèves n'ayant pas reçu cet entraînement. Lors de cette expérimentation à laquelle participèrent 94 élèves répartis dans 5 Écoles de la Région de Québec, 63 furent choisis de façon aléatoire à l'intérieur des différents groupes-classes participants pour procéder à l'auto-évaluation de leur performance suite à l'exécution de la tâche ciblée. Sur ces 63 élèves, 33 furent assignés également de façon aléatoire pour recevoir l'entraînement planifié.

Dans un premier temps, les 33 sujets soumis au traitement de l'entraînement ont visionné, sur vidéo-cassette, l'exécution d'une tâche complexe (pratique des différentes techniques du clouage) par un menuisier professionnel. La consigne alors donnée aux élèves était de regarder attentivement en essayant de bien comprendre les explications du menuisier touchant les règles à suivre ou les gestes à poser dans une tâche de clouage. Dans un deuxième temps, les élèves ont visionné à nouveau le montage sous la direction cette fois de l'enseignant qui, par le biais de questions orales, s'assurait alors de leur juste compréhension des critères de performance expliqués par le menuisier tout au long de sa démonstration. Selon le cas, les critères moins bien compris faisaient l'objet d'explications supplémentaires ou d'une reprise du visionnement sur le point concerné. Après ce second visionnement, une série de 12 diapositives furent présentées aux élèves. Ces derniers purent donc observer différentes performances d'un menuisier en train d'exécuter l'une ou l'autre des techniques précises du clouage. Certaines de ces techniques étaient correctement exécutées alors que d'autres allaient à l'encontre des critères de performance qui venaient de

leur être expliqués. La tâche demandée aux élèves était d'observer les gestes techniques posés, de les confronter aux critères de performance appris et de porter un jugement sur la qualité des performances démontrées en fonction de ces critères. Un feedback était donné aux élèves après chacune des réponses accordées. Mentionnons que pour assurer une meilleure conduite de l'exercice et uniformiser le plus possible l'entraînement donné aux élèves, les enseignants participants avaient été préalablement rencontrés en groupe pour visionner la vidéo-cassette, s'assurer de leur juste compréhension des critères de performance expliqués par le menuisier et faire le point sur la démarche commune qu'ils devaient appliquer pour mener l'exercice auprès des élèves.

Quelques jours après cette période intensive d'entraînement, (période d'une durée moyenne de 2 heures dans l'ensemble des groupes), tous les élèves de l'étude exécutèrent à tour de rôle la tâche complexe prévue pour leur groupe-classe dans le plan expérimental. La performance démontrée par chacun d'eux aux différentes composantes de la procédure fut alors notée par l'enseignant qui jouait le rôle d'évaluateur à titre de spécialiste de la discipline. Pour chaque étape de la procédure correctement exécutée au jugement de l'enseignant, l'élève se voyait accorder un point. Le nombre total de composantes exécutées selon les règles constituait le score attribué à l'élève comme résultat de la tâche. Pour des fins de comparaisons et d'analyse, les scores furent exprimés sur cent. Ainsi, un élève qui avait bien exécuté 9 procédures sur 15 se voyait attribuer un résultat de 9/15 ou de 60%.

Pour les résultats en auto-évaluation, ils furent attribués de la façon suivante: à chacun des énoncés-questions de la liste de vérification, l'élève devait avoir fourni la même réponse que l'enseignant pour se voir accorder 1 point en habileté d'auto-évaluation. Soulignons ici que le jugement de l'enseignant constituait une référence de poids puisque les critères évalués avaient été expliqués aux élèves par ce même enseignant lors de la période d'apprentissage de la procédure en cause. L'expertise de l'enseignant était donc établie au départ pour juger de la qualité de la performance démontrée par les élèves. Sur le nombre total de jugements portés dans l'exercice, c'est la proportion de jugements de l'élève identiques à ceux de l'enseignant qui constituait le score de l'élève en auto-évaluation. Comme pour les résultats accordés dans l'exécution de la tâche, ce score a ensuite été ramené sur 100 pour fins de comparaisons et de calculs. Par exemple, l'élève qui avait fourni la même réponse que l'enseignant à 12 énoncés-questions sur 15 se voyait attribuer un score de 12/15 ou de 80% en auto-évaluation.

Dans notre première hypothèse, nous avons avancé que l'entraînement des élèves à l'utilisation de critères précis pour évaluer la performance d'une autre personne qui exécute une tâche complexe influence par la suite l'habileté que ces élèves démontrent en auto-évaluation. La vérification de l'hypothèse a été faite en comparant les résultats en auto-évaluation des 33 sujets affectés au traitement de l'entraînement aux résultats des 30 autres sujets qui n'avaient pas subi ce traitement. Le tableau 1 présente les données obtenues suite aux comparaisons effectuées entre les deux groupes.

TABLEAU 1

*Résultats obtenus en auto-évaluation lors de l'expérimentation I
en fonction de la variable d'un entraînement intensif
à évaluer à partir de critères précis*

	<i>Nombre</i>	<i>Moyenne %</i>	<i>Écart-type</i>	<i>t₍₆₁₎</i>
Élèves entraînés	33	87,55	11,09	0,16
Élèves non entraînés	30	87,00	15,32	
$t_{(61; \infty, .01)} > 2,576$				

Ainsi qu'on peut le constater dans le tableau 1, les deux groupes obtiennent des résultats relativement élevés en auto-évaluation (respectivement 87,55% et 87,00%) et ne diffèrent pas de façon significative entre eux ($t=0,16$ avec 61 d.l.). D'après ces résultats, on ne peut donc conclure qu'un entraînement intensif à l'utilisation de critères précis pour observer et évaluer une tâche complexe réalisée par une autre personne influence l'habileté démontrée par la suite dans l'auto-évaluation de ses propres performances. Avec les données et dans le contexte précis de notre étude, la première hypothèse de nos travaux n'a donc pas été confirmée. Il est toutefois intéressant de noter une dispersion moins grande des résultats dans le groupe des élèves soumis à l'entraînement (un écart-type de 11,09 contre 15,32).

Vérification des hypothèses H2 et H3

Au cours de l'expérimentation II, nous avons étudié cette fois l'effet que pouvait avoir sur l'habileté en auto-évaluation un entraînement à certaines pratiques évaluatives mené sur une période plus longue et impliquant l'utilisation individualisée d'instruments par les élèves. Cette deuxième expérimentation a été menée auprès de 100 autres élèves du secteur professionnel répartis dans trois écoles de la Région de Québec. Comme lors de la première expérimentation, les élèves devaient, après la période d'entraînement, exécuter une tâche complexe dont l'apprentissage était prévu dans leur programme d'études. La performance démontrée dans cette tâche fut auto-évaluée par tous les sujets de l'étude. Pour ce qui est de l'évaluation de la tâche par l'enseignant, elle a été faite dans les mêmes conditions que celles appliquées au cours de la première expérimentation et précisées plus avant. L'étude a de plus été l'occasion de calculer des corrélations entre les scores des sujets en auto-évaluation et les deux variables des résultats scolaires et de la motivation aux études.

L'entraînement des élèves à l'évaluation a été fait selon trois modalités qui pouvaient impliquer autant la pratique évaluative que la pratique auto-évaluative. Ces trois modalités d'entraînement, qui ont constitué le traitement expérimental de la deuxième expérimentation, étaient les suivantes:

- Traitement A: entraînement limité à l'*auto-évaluation*
(*auto*) (6 moments d'auto-évaluation d'une performance complexe avec utilisation d'une liste de vérification chaque fois)
- Traitement B: entraînement limité à l'*évaluation*
(*pair*) (6 moments d'évaluation d'une performance complexe démontrée par un pair avec utilisation d'une liste de vérification chaque fois)
- Traitement C: entraînement à l'*auto-évaluation* et à l'*évaluation*
(*auto et pair*) (3 moments d'auto-évaluation et 3 moments d'évaluation d'une performance complexe démontrée par un pair avec utilisation d'une liste de vérification chaque fois)

Dans chacun des 3 traitements A, B ou C, les performances évaluées portaient sur des tâches complexes dont l'apprentissage était prévu au programme d'études des élèves. Pour chacune de ces tâches, l'enseignant, après avoir été entraîné à le faire, avait développé une liste de vérification de même type que celle qui devait être présentée à l'élève dans l'exercice d'auto-évaluation prévu à la fin de l'entraînement. Au cours de la période d'entraînement, les élèves recevaient l'aide voulue pour apprendre à utiliser correctement l'instrument. Dans chacun des groupes-classes de la recherche, la période prévue pour l'entraînement couvrait de 3 à 5 semaines et demeurait assujettie à la fois aux activités du programme et au calendrier scolaire du groupe-classe.

La deuxième hypothèse de l'étude était donc à l'effet que les élèves entraînés à la pratique double de l'évaluation et de l'auto-évaluation de tâches complexes à partir de critères précis démontreraient par la suite une meilleure habileté en auto-évaluation que les élèves entraînés soit uniquement à l'évaluation, soit uniquement à l'auto-évaluation. Pour vérifier cette hypothèse, les 100 sujets de l'expérimentation II ont été répartis au hasard en quatre groupes pour appliquer une des trois modalités d'entraînement décrites plus haut ou pour faire partie d'un groupe contrôle. Après la période d'entraînement, les sujets des quatre groupes ont été soumis à un exercice d'auto-évaluation d'une tâche complexe. Comme lors de la première expérimentation, l'outil utilisé pour procéder à l'auto-évaluation consistait en une liste de vérification décrivant avec précision les différentes opérations de la tâche demandée. Le tableau 2 présente les résultats des sujets de la deuxième expérimentation pour chacune des variables suivantes de l'étude: scores en auto-évaluation, résultats à la tâche, résultat global au bulletin et scores à un test de motivation aux études (test de 22 items développé et validé par Rousseau en 1970; ce test qui vise à mesurer l'attitude vis-à-vis les études donne indirectement un indice du degré de motivation de l'élève).

TABLEAU 2

*Résultats par groupe des sujets de l'expérimentation II
pour les différentes variables de l'étude*

Variables mesurées	Groupes avec entraînement à évaluation						Groupe contrôle	
	auto n=24		pair n=22		auto/pair n=27		n=27	
	\bar{X}	S	\bar{X}	S	\bar{X}	S	\bar{X}	S
Auto-éval.	88,8	7,8	9,8	8,9	81,1	15,5	86,8	13,0
Résultats à la tâche	74,6	12,8	83,6	8,5	62,8	19,7	71,2	15,6
Bulletin	72,4	10,3	80,6	7,4	63,5	15,3	71,6	9,3
Motivation	161,0	25,7	158,3	26,1	156,0	23,0	152,9	23,7

Les données du tableau 2 révèlent que les scores obtenus en auto-évaluation sont relativement élevés (supérieurs à 80%) dans les 4 groupes. Dans l'ensemble, on constate donc à nouveau (voir les résultats du tableau 1) que les élèves tendent à auto-évaluer leur performance de la même façon que l'enseignant dans plus de 8 occasions sur 10. Les résultats aux différentes variables mesurées sont toutefois supérieurs dans le groupe entraîné à l'évaluation d'un pair, sauf pour la variable motivation où la moyenne est moindre que celle du groupe entraîné à l'auto-évaluation. Le groupe entraîné à la fois à l'évaluation d'un pair et à l'auto-évaluation obtient par ailleurs les résultats les plus faibles pour l'ensemble des variables, à l'exception des scores en motivation qui sont supérieurs à ceux du groupe contrôle.

À prime abord, on constate donc que le groupe entraîné à l'évaluation d'un pair est celui qui a le mieux réussi en auto-évaluation tandis que le groupe entraîné à la fois à l'auto-évaluation et à l'évaluation d'un pair est celui qui a présenté la performance la plus faible pour cette même variable. Une analyse statistique de ces données a toutefois permis de tenir compte des différences des sujets au départ (performance à la tâche, résultats au bulletin et scores à un test de motivation aux études) pour mieux apprécier les résultats en auto-évaluation. Pour vérifier de façon formelle notre deuxième hypothèse, nous avons donc appliqué la technique de l'analyse de covariance entre les résultats des trois groupes expérimentaux et ceux du groupe contrôle. Nous avons retenu le groupe (traitement) comme variable indépendante et les scores en auto-évaluation comme variable dépendante. Les résultats à la tâche, les moyennes au bulletin et les scores au test de

motivation aux études ont pour leur part été retenues comme variables concomitantes. Les résultats de cette analyse de covariance sont présentés dans le tableau 3.

Les résultats de l'analyse de covariance révèlent que le traitement accordé (variable groupe) n'est pas relié aux scores obtenus en auto-évaluation ($F=0,16$ avec $\alpha=,9225$). Notre deuxième hypothèse n'a donc pas été confirmée avec les données de l'étude. Suite à nos travaux, on ne peut donc conclure qu'un entraînement portant à la fois sur l'auto-évaluation et sur l'évaluation d'un pair est davantage relié à l'habileté démontrée par la suite dans une tâche auto-évaluative qu'un entraînement limité soit à l'auto-évaluation, soit à l'évaluation d'un pair. L'analyse fait ressortir par ailleurs que les résultats à la tâche sont reliés fortement aux scores en auto-évaluation ($F=19,4$ avec $\alpha=,0001$). Dans l'ensemble, les élèves qui ont démontré une bonne performance à la tâche ont donc eu tendance à obtenir un score également bon en auto-évaluation.

Notre troisième et dernière hypothèse avançait enfin que la motivation aux études et les résultats scolaires étaient reliés à l'habileté démontrée par les élèves dans l'auto-évaluation de leur performance suite à l'exécution d'une tâche complexe. Les résultats de l'analyse de covariance du tableau 3 ont été utilisés pour vérifier cette hypothèse. Comme on peut le constater, les rapports F ne révèlent pas une relation significative entre les résultats en auto-évaluation et ceux obtenus soit au bulletin ($F=0,00$ avec $\alpha=,9651$), soit au test de motivation aux études ($F=1,14$ avec $\alpha=,2679$). Dans le contexte de notre étude, la troisième hypothèse avancée n'a donc pas été confirmée.

TABLEAU 3

Résultats de l'analyse de covariance entre les résultats des groupes expérimentaux et du groupe contrôle lors de l'expérimentation II (n=100)

<i>Source</i>	<i>DL</i>	<i>Somme des carrés</i>	<i>Carrés moyens</i>	<i>F</i>	<i>Pr>F</i>
<i>Variable indépendante:*</i>					
Groupe	3	77,38	25,79	0,16	0,9225
<i>Variables concomitantes:</i>					
Tâche	1	3114,08	3114,08	19,40	0,0001
Bulletin	1	0,31	0,31	0,00	0,9651
Motivation	1	183,45	183,45	1,14	0,2879

*Variable dépendante: auto-évaluation.

DISCUSSION ET CONCLUSIONS

Les résultats de nos travaux confirment d'abord que les élèves observés ont été capables d'utiliser correctement la liste de vérification pour auto-évaluer leur performance dans l'exécution d'une procédure. En effet, les scores moyens respectifs de 87,3% et 86,6% obtenus par les sujets des deux expérimentations indiquent que les élèves tendent à apprécier leur performance à un niveau qui se rapproche de celui de l'enseignant quand ils réalisent une tâche complexe. À notre avis, ces résultats sont intéressants et ne peuvent qu'inciter les maîtres à favoriser une pratique fréquente de l'auto-évaluation par les élèves, particulièrement au secteur professionnel où les productions complexes constituent le lot majeur des activités de formation dans les programmes d'études.

Par ailleurs, nous n'avons pas obtenu de différences significatives entre les résultats en auto-évaluation des groupes de l'étude, et ce malgré la diversité des traitements expérimentaux accordés à ces groupes au cours de deux expérimentations indépendantes. Ces résultats suscitent des interrogations. Ainsi, l'apprentissage de l'auto-évaluation exige-t-il qu'on mette en place des procédures d'entraînement systématique de l'élève avant de le mettre en situation d'auto-évaluation? Si un environnement pédagogique adéquat favorise la tâche d'auto-évaluation, des explications claires au moment d'utiliser un outil seraient-elles suffisantes pour que l'élève mène à bien la tâche demandée et estime sa performance de façon assez juste? Les élèves dont le fonctionnement est régulier ne seraient-ils pas déjà capables de porter un jugement correct sur leur performance s'ils disposent, pour guider leur appréciation, de critères précis et qu'ils comprennent bien (ce qui était le cas dans notre étude)? Chose certaine, aucune des modalités d'entraînement appliquées dans cette recherche (entraînement intensif audiovisuel réalisé en groupe dans l'expérimentation I; entraînement individuel plus long et instrumenté avec accent mis soit sur l'évaluation de ses propres tâches, soit sur l'évaluation des tâches d'un pair dans l'expérimentation II) n'a pu être reliée aux scores obtenus ultérieurement par les sujets dans un exercice d'auto-évaluation. Mentionnons toutefois que nos travaux n'ont permis d'utiliser qu'un seul type d'instrument, soit la liste de vérification. On peut se demander si un outil qui fait appel à des réponses de nature dichotomique développe autant l'habileté en auto-évaluation que les outils qui favorisent des réponses plus nuancées. Qu'on pense par exemple à la grille descriptive (voir Scallon, 1988) ou à une énumération de critères accompagnée d'une échelle d'appréciation non dichotomique (voir Auger et Laveault, 1991). Dans d'autres travaux, il serait intéressant qu'on compare l'effet de pratiques évaluatives menées à l'aide d'instruments différents sur l'habileté démontrée ensuite par les élèves en auto-évaluation.

Soulignons par ailleurs que le nombre relativement peu élevé des sujets observés n'a pas permis de tenir compte de certaines variables supplémentaires comme le métier ou l'âge dans les analyses effectuées. Il serait

intéressant de reprendre l'étude avec un plus grand nombre de sujets pour pouvoir tenir compte de telles variables dans le traitement statistique des données. À bon droit, on peut en effet se demander quelle est l'influence de l'âge des élèves sur leur habileté à s'auto-évaluer. Si nos travaux avaient été menés auprès d'élèves d'un âge moyen de 13 ans, par exemple, les résultats obtenus auraient-ils été les mêmes? Et si la tâche auto-évaluée par les élèves avait été une production écrite, un travail scolaire long, une oeuvre artisanale ou l'expression d'une création quelconque plutôt qu'une tâche de type séquentiel comme celles qu'on retrouve majoritairement au secteur professionnel, quel aurait été alors l'effet de l'entraînement expérimenté sur les scores des élèves en auto-évaluation? Comme on le constate, les questions demeurent nombreuses et on ne saurait s'appuyer sur les résultats rapportés ici pour tirer des conclusions générales ou définitives relativement à l'influence de l'entraînement des élèves pour développer leur habileté à porter un jugement correct sur leurs propres performances.

L'importance de développer l'habileté des élèves à l'auto-évaluation n'est plus à démontrer. Tout au long de sa vie d'adulte, que ce soit à titre professionnel ou dans le cadre d'activités plus personnelles, l'être humain est en effet appelé à porter un jugement sur ses actions ou sur ses productions, et dans la grande majorité des cas, la qualité des activités ou des réalisations en cause est intimement liée à la justesse de ce jugement. À l'instar des auteurs rapportés plus avant, nous croyons donc que l'apprentissage de l'auto-évaluation doit se poser comme un des objectifs prioritaires de formation à l'école. Et c'est par le biais de la recherche qu'on parviendra à identifier ou à décrire les approches ou les moyens qui semblent contribuer le plus à l'atteinte par les élèves de cet objectif supérieur de développement personnel.

RÉFÉRENCES

- Allal, L., Baeriswyl, E., Mai, T.B. et Wegmuller, E. (1987). Le jeu comme situation d'auto-évaluation. *Mesure et Évaluation en Éducation*, 10(1), 47–63.
- Auger, R. et Laveault, D. (1991). Le logiciel MicroCAT: points de vue critiques en regard de sa praticabilité pour différentes clientèles. *Mesure et Évaluation en Éducation*, 14(3), 23–41.
- Bloom, B.S., Madaus, G.F., & Hastings, J.T. (1981). *Evaluation to improve learning*. New-York: McGraw-Hill.
- Boisvert, M.T. (1985). *Étude de la justesse d'auto-notation selon diverses variables personnelles et situationnelles*. Thèse de doctorat inédite, Université de Montréal.
- Bonniol, J.J. (1981). *Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires*. Thèse de Doctorat d'état, Université de Bordeaux.
- Bonniol, J.J. (1984). Influence de l'explicitation des critères utilisés sur le fonctionnement des mécanismes d'évaluation d'une production scolaire. *Bulletin de Psychologie*, 353, 159–165.
- Broadbent, D.E. (1975). Cognitive psychology and education. *British Journal of Educational Psychology*, 45, 162–176.

- Cardinet, J. (1986). *Évaluation scolaire et mesure*. Bruxelles: De Boeck-Wesmael.
- Dominice, P. (1981). *La formation: enjeu de l'évaluation*. Berne: Peter Lang.
- Fontaine, F. (1988). *Mesure critériée des croyances des futurs maîtres à l'égard de l'évaluation des apprentissages*. Thèse de doctorat inédite, Université de Montréal.
- Gagné, R.M., & Briggs, L.J. (1979). *Principles of instructional design*. New York: Holt, Rinehart and Winston.
- Holec, H. (1981). Plaidoyer pour l'auto-évaluation. *Le français dans le monde*, 21, 15–23.
- Lavoie-Sirois, J. (1983). *Guide du stagiaire en enseignement professionnel*. Ste-Foy: Université Laval.
- Lavoie-Sirois, J. (1991). Habileté des élèves du secteur professionnel à évaluer leur performance-Effets de l'auto-évaluation sur l'apprentissage. *Mesure et Évaluation en Éducation*, 14(2), 25–36.
- Leon, A. (1977). *Manuel de psychologie expérimentale*. Paris: Presses Universitaires de France.
- Leselbaum, N. (1982). *Autonomie et auto-évaluation*. Paris: Economica.
- Levesque, M. (1982). *Relation entre la pédagogie pour la maîtrise de l'apprentissage (PMA) et le concept de soi scolaire chez des élèves de première secondaire*. Thèse de doctorat inédite, Université Laval.
- Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.
- Paquay, L., Allal, L. et Laveault, D. (1990). L'auto-évaluation en question(s): propos pour un débat. *Mesure et Évaluation en Éducation*, 13(3), 5–26.
- Scallon, G. (1988). *L'évaluation formative des apprentissages: tomes 1 et 2*. Québec: Les Presses de l'Université Laval.
- Schwartz, E. (1973). *L'Éducation demain*. Paris: Aubier Montaigne.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.

Jeannine Lavoie-Sirois est professeure au Département de mesure et évaluation, Faculté des sciences de l'éducation, Université Laval, Sainte-Foy (Québec), G1K 7P4.