

Quelques considérations sur l'évaluation de l'efficacité d'une mesure d'aide à la réussite



Denis Monaghan

*Professeurs au département de psychologie
Cégep de Sainte-Foy*



Natalie Chaloux

Après avoir incité les institutions collégiales à se doter de mesures visant à accroître la réussite scolaire, le ministère de l'Éducation s'attend maintenant à ce que les collèges rendent compte de l'efficacité des interventions mises en place¹. Il est possible d'y voir une occasion de flagorner les mérites d'un collègue en divulguant quelques données percutantes sélectionnées avec soin. Cependant, pour répondre judicieusement à cette requête du ministère, une forme d'évaluation plus crédible et productive doit être entreprise. Cet article a donc pour but de sensibiliser la lectrice et le lecteur à l'importance de quelques principes méthodologiques fondamentaux intimement liés à une démarche d'évaluation rigoureuse, objective et, surtout, valide. Ces principes connus sont trop souvent omis puisque les évaluations doivent s'effectuer avec un minimum de ressources et à l'intérieur de délais très courts. Pour bien exposer et illustrer toute l'importance de ces principes, nous utilisons principalement des données issues d'une recherche PAREA menée au cégep de Sainte-Foy, de l'automne 2001 à l'hiver 2003 et intitulée *Évaluation d'un programme d'aide à la réussite*².

1. Lettre de la Commission d'évaluation de l'enseignement collégial adressée aux directeurs généraux en mai 2003.
2. MONAGHAN, Denis et NATALIE CHALOUX, *Évaluation d'un programme d'aide à la réussite*, Recherche PAREA, Sainte-Foy, Cégep de Sainte-Foy, 2004.

La mesure d'aide évaluée dans cette recherche porte le nom de *Cheminement Zénith*. Cette intervention est conçue pour les élèves qui réussissent moins de 50 % des unités rattachées aux cours suivis à l'une ou l'autre de leurs deux premières sessions au cégep. Tous les élèves invités ont cependant obtenu une cote finale de 70 % ou plus pour leurs études secondaires. Puisqu'il semble que ces élèves possèdent déjà les aptitudes intellectuelles nécessaires pour réussir, l'intervention ne porte pas sur des facteurs d'ordre cognitif, mais plutôt d'ordre affectif. Ainsi, le *Cheminement Zénith* vise à développer la connaissance de soi et la valorisation personnelle pour faire en sorte que l'élève s'engage dans ses études et les réussisse. L'intervention, qui s'étend sur une session entière, intègre des cours adaptés, un stage et une supervision en tutorat.

L'évaluation de l'efficacité de ce genre de mesure d'aide pourrait se faire rapidement, à peu de frais, tout en donnant des résultats encourageants mais pourtant futiles. Quelques paradoxes issus de données bien concrètes démontreront la nécessité de procéder à des évaluations plus systématiques, souvent plus coûteuses mais combien plus valides et fécondes.

La réussite scolaire augmente chez les élèves aidés, mais l'intervention est inefficace

Pour plusieurs mesures d'aide mises en place dans les collèges, l'objectif principal est de promouvoir la réussite scolaire des élèves qui en bénéficient. Pour cette raison, l'évaluation de l'efficacité de l'intervention repose donc en bonne partie sur une analyse des résultats scolaires des élèves aidés.

À ce propos, les données issues de notre recherche indiquent qu'avant l'intervention, les élèves inscrits au *Cheminement Zénith* réussissent en moyenne seulement 41 % des unités rattachées aux cours qu'ils ont suivis. Lors de la session d'intervention, ces mêmes élèves réussissent 78 % de leurs unités et, un an après l'encadrement, ce taux de réussite grimpe à 85 %. De même, le *Programme d'Encadrement Spécialisé* (PEP), offert au cégep de Rimouski à des élèves subissant aussi des échecs multiples, semble avoir un impact similaire sur la réussite scolaire (Désilets et Dubé, 2002). En effet, chez les élèves ayant bénéficié du programme, le pourcentage de cours réussis sur le nombre de cours suivis se situe entre 25 % et 35 % avant

l'encadrement et il oscille entre 71 % et 85 % lors des sessions qui suivent l'intervention.

Malgré des améliorations aussi impressionnantes de la réussite scolaire, nul ne peut affirmer que les programmes sont efficaces. En effet, rien n'indique que celles-ci soient imputables aux encadrements offerts. La hausse du taux de la réussite scolaire pourrait être observable chez tout élève à qui l'on offre la chance de se reprendre après avoir subi des échecs multiples. Pour vérifier l'existence d'un lien de cause à effet entre une intervention et la réussite scolaire, le chercheur doit absolument constituer un groupe de comparaison, appelé groupe témoin ou contrôle, selon que la recherche s'inscrit dans un plan corrélationnel ou expérimental (Lamoureux, 2000). Ce groupe de comparaison est composé d'élèves possédant les mêmes caractéristiques que les élèves encadrés constituant le groupe expérimental. Dans notre recherche, les élèves du groupe de comparaison et du groupe expérimental sont inscrits dans les mêmes programmes d'études préuniversitaires ; de plus, ils sont équivalents en ce qui concerne leur cote finale au secondaire et leur taux de réussite au collège avant l'intervention. Le groupe de comparaison ne bénéficie évidemment pas d'une aide à la réussite aussi spécifique et élaborée que celle qui est offerte au groupe expérimental, mais peut normalement profiter d'autres formes d'aide disponibles au collège.

C'est ainsi que, selon notre analyse, les élèves du groupe de comparaison, aussi bien que les élèves du groupe expérimental, réussissent en moyenne 41 % de leurs unités de cours avant l'intervention. Le taux de réussite du groupe de comparaison se situe à 79 % durant la session qui correspond à l'encadrement, bien que ces élèves ne bénéficient d'aucun encadrement particulier et il se hausse à 86 %, un an plus tard. Il n'y a donc aucune différence significative sur le plan de la réussite scolaire entre les élèves aidés et ceux qui ne le sont pas. Ces données laissent croire que la réussite scolaire augmente de toute façon chez les élèves qui persévèrent dans les études collégiales, malgré un taux de réussite inférieur à 50 % lors d'une de leurs deux premières sessions au collège. Par le fait même, il semble que l'intervention n'ait pas d'effet particulier sur le plan de la réussite scolaire.

À la lumière de ces résultats, il nous apparaît essentiel d'insister sur **la nécessité de former un groupe de comparaison lors de l'évaluation de mesures d'aide**. Sans groupe de comparaison, les résultats sont susceptibles d'être mal interprétés et attribués par erreur aux mesures d'aide. Dans ce cas, l'évaluation est jugée invalide.

*[Certaines données obtenues]
laissent croire que la réussite scolaire
augmente de toute façon chez les élèves qui
persévèrent dans les études collégiales,
malgré un taux de réussite inférieur à 50 %
lors d'une de leurs deux premières sessions
au collège.*

La satisfaction de la clientèle à l'égard d'une mesure d'aide est indépendante de son efficacité

Lorsqu'un même élève bénéficie simultanément de plusieurs mesures d'aide ou que l'encadrement est constitué de plusieurs interventions, il devient extrêmement difficile d'évaluer l'impact de chacune d'elles séparément. Il faudrait créer un groupe de comparaison pour chacune des différentes interventions, ce qui mènerait à des plans expérimentaux beaucoup trop complexes et irréalisables. Pour tenter d'estimer l'impact d'une intervention particulière, l'évaluateur peut prendre une mesure d'appréciation de chacune des interventions auxquelles un élève a été soumis. Ainsi, les moyennes des appréciations fournies par les élèves aidés peuvent être comparées et le chercheur peut déterminer que telle ou telle intervention a été plus appréciée ou a été perçue comme plus satisfaisante qu'une autre.

C'est ce que nous avons fait dans notre recherche. Les élèves qui ont participé au *Cheminement Zénith* ont indiqué leur appréciation pour chacune des interventions liées à ce cheminement : le cours de psychologie (gestion et connaissance de soi), le cours d'éducation physique (défi-expédition), le stage, le service d'une conseillère en orientation et la supervision en tutorat. L'instrument de mesure de l'appréciation contient des questions portant sur les objectifs, les contenus de cours, les moyens pédagogiques utilisés, etc. Sur une échelle d'accord de 1 à 5 (5 signifiant la plus grande satisfaction possible), les moyennes observées des différentes appréciations s'étalent de 3,6 à 4,0. Très peu d'élèves ont coté négativement les énoncés proposés (cotes 1 et 2). La qualité de la relation de tutorat a été mesurée à l'aide du IAT³. L'échelle du IAT s'étend de 1 à 7. La moyenne d'estimation de la qualité de la relation de tutorat par les élèves est alors de 5,4. Les commentaires obtenus de la part des

3. L'inventaire d'alliance de travail (IAT) est une version française du *Working Alliance Inventory*. Cet instrument développé par HORVATH(1986) a été traduit et validé par BACHELOR et SALAMÉ (1991).

élèves sont très positifs et même flatteurs pour les intervenantes et les intervenants. On constate donc que chacune des interventions est particulièrement appréciée des élèves. Évidemment, un tel taux de satisfaction est en soi très encourageant. C'est une perception de l'élève reflétant un sentiment d'avoir été aidé. Sur ce point, le *Cheminement Zénith* a eu un impact positif.

Toutefois, notre recherche montre qu'il n'y a aucune corrélation positive significative entre le taux de satisfaction de chacune des interventions spécifiques et la réussite scolaire⁴ de l'élève durant ou après la session d'encadrement. Les coefficients de corrélation de Pearson⁵ pertinents oscillent entre -0,150 et +0,199 sans qu'aucun d'eux ne soit significatif au seuil de 0,05. Voici donc un fait empirique démontrant que l'appréciation d'une mesure d'aide ne prédit pas nécessairement son impact réel sur la performance d'un individu. Il est possible que le haut taux de satisfaction résulte du fait que l'élève attribue à la mesure d'encadrement la hausse de ses performances scolaires.

[...] notre recherche montre qu'il n'y a aucune corrélation positive significative entre le taux de satisfaction de chacune des interventions spécifiques et la réussite scolaire de l'élève durant ou après la session d'encadrement.

[...] Il est possible que le haut taux de satisfaction résulte du fait que l'élève attribue à la mesure d'encadrement la hausse de ses performances scolaires.

D'autre part, une forte appréciation de la mesure d'aide peut être le reflet d'un sentiment de bien-être chez l'élève. Ce bien-être de l'élève pourrait correspondre à un objectif clairement identifié de l'intervention et, dans ce cas seulement, il constitue un élément de l'efficacité de cette dernière. Cependant, l'évaluateur doit être conscient du fait qu'il ne s'agit pas là d'un indicateur ou d'un prédicteur de la réussite scolaire. Ainsi, **il est nécessaire que l'évaluation soit plus qu'une simple mesure d'appréciation**. Malheureusement, certaines études rapportent que plusieurs évaluations de mesures d'aide ne tiennent compte que de l'appréciation des participants et des

intervenants (Fédération des cégeps, 1999 ; Soucy, Duchesne, Larose, 2000).

L'atteinte des objectifs ne garantit pas leur validité

Les mesures d'aide développées dans les collèges sont multiples et s'adressent à des clientèles souvent bien ciblées mais diversifiées. Les objectifs sont donc variés et habituellement adaptés aux caractéristiques et besoins particuliers de ces dernières. Lorsque ces objectifs sont formulés clairement et de façon opérationnelle, il est possible de les évaluer. Idéalement, des modèles théoriques ou des constats empiriques guident le choix des objectifs et des moyens à prendre pour les atteindre à court, à moyen ou à long terme.

Ainsi, il est raisonnable de croire que les élèves du *Cheminement Zénith*, peu habitués à l'échec scolaire au secondaire, puissent se sentir dévalorisés suite aux nombreux échecs subis au début de leurs études collégiales. Selon Denise Barbeau (1993), à la suite d'échecs multiples, l'élève développe un sentiment d'impuissance s'il perçoit ces échecs comme hors de son contrôle. Il devient alors passif. Cette vision est aussi partagée par Bernard Rivière (1995) qui remarque que l'élève qui subit des échecs scolaires éprouve un sentiment d'aliénation et adopte, conséquemment, des stratégies d'évitement qui l'amènent à une inhibition de l'action. C'est pour pallier ce problème que le *Cheminement Zénith* s'est donné comme objectif de permettre à l'élève de vivre des expériences scolaires et parascolaires positives et valorisantes. D'une part, celles-ci peuvent contrebalancer les effets pervers sur le sentiment d'impuissance déclenché par les échecs et, d'autre part, elles favorisent le passage à l'action, une façon de contrer les forces inhibitrices auxquelles sont soumis ces élèves. Le fait de vivre des expériences valorisantes devrait donc ultérieurement faciliter la réussite scolaire.

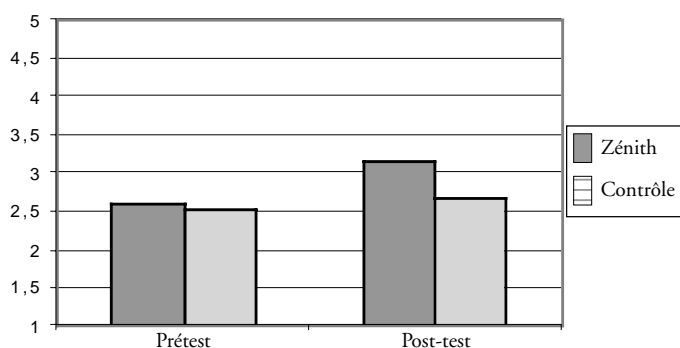
Conséquemment, afin de vérifier que la mesure d'aide est bien adaptée, l'évaluation doit vérifier si les élèves se sentent dévalorisés suite à leurs nombreux échecs. Pour le savoir, il faut prendre **une mesure** des expériences valorisantes **en prétest** c'est-à-dire avant l'intervention. Cette mesure de prétest s'avère également essentielle pour déterminer s'il y a eu ou non un changement quelconque de perception chez l'élève sur le plan des expériences valorisantes après l'intervention. L'instrument de mesure utilisé dans notre recherche est un questionnaire qui permet à l'élève d'indiquer sur une échelle de 1 à 5 s'il considère que sa dernière session lui a permis de vivre des expériences valorisantes dans 10 domaines de sa vie.

Le graphique suivant présente les résultats du prétest et du post-test des élèves du groupe expérimental (*Zénith*) et du groupe de comparaison en ce qui concerne cette variable nommée « expériences valorisantes ».

4. La réussite scolaire est mesurée à l'aide du pourcentage d'unités de cours échouées et aussi par la cote R.

5. Le coefficient de corrélation de Pearson est une mesure d'association linéaire entre deux variables.

Moyenne des résultats au test des expériences valorisantes pour les deux groupes de sujets (Zénith et contrôle) au prétest et au post-test



Les élèves perçoivent leur session passée dans le *Cheminement Zénith* (post-test) comme plus valorisante que la précédente (prétest). Chez les élèves du groupe contrôle, il n'y a pratiquement pas de différence entre le prétest et le post-test⁶. L'intervention a donc eu un impact positif sur le sentiment d'avoir vécu des expériences valorisantes. Le *Cheminement Zénith* atteint ainsi un de ses objectifs immédiats. Il semble donc efficace sur ce point.

Cependant, il faut se rappeler que cet objectif est issu de la croyance que les élèves ayant subi des échecs multiples ont vécu une expérience dévalorisante. Pour savoir si cette perception est bien réelle, nous avons fait remplir le questionnaire en prétest par un groupe d'élèves des mêmes programmes et de même niveau qui n'ont pas subi d'échecs au collège à leur première session. Étonnamment, les élèves qui ont subi plusieurs échecs (groupe expérimental et contrôle) disent avoir vécu autant d'expériences valorisantes que les élèves qui n'ont pas subi d'échecs. Il n'y a pas de différence significative entre les moyennes des expériences valorisantes des élèves ciblés par la mesure d'aide et celles des élèves qui ont réussi leurs cours dans les mêmes programmes préuniversitaires⁷. Il semble que l'objectif de faire vivre des expériences valorisantes soit justifié théoriquement, mais qu'il n'obtienne pas de support empirique. Ainsi, l'évaluation révèle que la mesure d'aide atteint un objectif jugé important, mais que la validité de cet objectif est remise en question.

Il appert donc que **les objectifs d'une mesure d'aide doivent être fondés non seulement sur des modèles théoriques mais aussi sur des constats empiriques**. De plus, la recherche peut vérifier la valeur et l'atteinte d'un objectif en autant que ce dernier soit clair, précis et mesurable.

6. Ces résultats sont confirmés par une analyse de variance selon un plan factoriel combiné. Il y a un effet d'interaction entre le groupe et le temps de mesure (Prétest et post-test), $F = 10,64$ et $p = 0,001$.

7. Résultat de l'analyse de variance : $F = 1,541$ et $p = 0,217$.

La planification de l'évaluation avant l'implantation de la mesure d'aide

Pour en arriver à une véritable évaluation de l'efficacité des mesures d'aide, il n'est pas suffisant de prendre le pouls des participantes et des participants. Comme nous l'avons démontré, l'appréciation d'une mesure d'aide n'est pas forcément garante de son efficacité, puisqu'il n'y a pas toujours de corrélation entre l'appréciation de la mesure d'aide et la réussite scolaire. De même, la seule collecte de données relatives à la réussite scolaire auprès des élèves aidés ne suffit pas. En effet, sans utilisation d'un protocole expérimental ou quasi expérimental, l'évaluatrice ou l'évaluateur ne peut prétendre attribuer la réussite des élèves à la mesure d'aide mise en place. L'évaluation de l'efficacité d'une mesure d'aide peut aussi impliquer la vérification de l'atteinte et de la valeur d'objectifs autres que la réussite scolaire. Nous avons illustré la possibilité d'évaluer des objectifs portant sur des variables d'ordre affectif. L'évaluation nécessite alors des mesures de prétest prises avant l'intervention.

Somme toute, l'évaluation de l'efficacité implique le respect de certaines conditions fondamentales telles la comparaison des résultats avec un groupe témoin ou contrôle et la collecte de données avant l'intervention (en prétest). Or, le respect de ces deux seules conditions exige une bonne planification méthodologique engendrant un investissement important de temps et de ressources du milieu. En effet, la constitution d'un groupe de comparaison peut s'avérer très complexe. Comment est-il possible de former un groupe contrôle si tous les participants potentiels bénéficient déjà de la mesure d'aide ? De même, les mesures de prétest ne peuvent se prendre *a posteriori* et, dans certains cas, nécessitent l'élaboration d'instruments de mesure.

À notre avis, la planification conjointe de l'intervention et de son évaluation peut éluder le problème de la composition du groupe contrôle. En effet, une fois que la population ciblée a été identifiée, la répartition aléatoire des élèves en deux échantillons se fait aisément. Ainsi, certains élèves choisis au hasard dans cette population sont invités à participer à la mesure d'aide et les autres constituent le groupe contrôle. C'est ce qu'Archambault et Aubé (1996) ont fait lors de leur recherche évaluative sur le suivi personnalisé des élèves à risque. Ce type de plan de recherche est qualifié d'expérimental. La randomisation soulève cependant un problème d'éthique lorsque la mesure vise à aider une population qui est en difficulté. On ne peut moralement priver d'aide une personne en difficulté lorsqu'on sait que cette aide est disponible. Dans une intervention préventive qui s'adresse à tous, le problème est moins criant. Bouchard et Cyr (1998) présentent divers moyens afin de contrer ce problème d'éthique. Afin de ne pas pénaliser une partie de la clientèle dite à risque, ils proposent, entre autres, de comparer la nouvelle intervention à une intervention standard ou à une intervention non spécifique,

ce que nous avons fait dans notre recherche. Par exemple, la chercheuse ou le chercheur compare la nouvelle intervention à une autre sans guide, ou sans formateur ou, encore, à une autre sans modèle de base.

La formation de groupes aléatoires n'est toutefois pas possible si toute la population ciblée participe à la mesure d'aide ou est même seulement invitée à participer. À partir du moment où un élève refuse de participer à une mesure d'aide, il démontre qu'il possède une caractéristique différente de celui qui accepte de participer volontairement. Dans un tel cas, l'évaluation de la mesure d'aide comprend la composition d'un groupe de comparaison à partir d'individus possédant le plus de caractéristiques similaires possibles à celles des participantes et des participants du groupe expérimental. La recherche de Larose et Roy (1993) sur le programme d'intégration aux études collégiales et celle de Cyrenne et Lacombe (1997) sur une intervention de tutorat maître-élève sont des exemples de ce type de recherche quasi expérimentale.

L'autre difficulté est liée aux mesures de prétest à effectuer pour toute variable liée à l'efficacité qui n'est pas inscrite dans le dossier scolaire de l'élève. C'est le cas pour l'ensemble des facteurs cognitifs et affectifs pour lesquels la seule mesure en post-test est insuffisante. La passation du prétest doit être prévue et exécutée avant que l'intervention débute, sinon il y a déjà contamination des résultats. De plus, dans certains cas, il faudra choisir ou même élaborer des instruments de mesure qui soient valides en regard des objectifs spécifiques poursuivis par la mesure d'aide. Tout cela prend du temps, nécessite des ressources, mais se trouve facilité lorsque l'évaluation est planifiée avant même l'implantation de la mesure d'aide.

Selon toute évidence, les mesures rétroactives ou prises en post-test seulement sont nettement moins fiables et valides que celles issues de recherches expérimentales avec mesures prétest et post-test. En fait, nous croyons que plusieurs évaluations fragmentaires ne produiront jamais autant de fruits qu'une seule évaluation rigoureuse et valide.

En conclusion, pour bien répondre à la requête du Ministère concernant l'évaluation de l'efficacité des mesures d'aide à la réussite, il nous apparaît plus rentable de consacrer les énergies et ressources disponibles à quelques évaluations de qualité plutôt qu'à des bilans exhaustifs qui ne permettent pas de juger de l'efficacité réelle des mesures d'aide mises en place. ▣

dmonaghan@cegep-ste-foy.qc.ca
nchaloux@cegep-ste-foy.qc.ca

RÉFÉRENCES BIBLIOGRAPHIQUES

- ARCHAMBAULT, Guy et Rachel AUBÉ, *Le suivi personnalisé des élèves à risque dans leur transition du secondaire au collégial*, Rapport de recherche, St-Georges, Cégep Beauce-Appalaches, 1996.
- BARBEAU, Denise, « La motivation scolaire », *Pédagogie collégiale*, vol. 7, n° 1, 1993.
- BOUCHARD, Stéphane et Carolyne CYR, *Recherche psychosociale : pour harmoniser recherche et pratique*, Sainte-Foy, Presses Universitaires du Québec, 1998.
- CYRENNE, Diane et Henriette LACOMBE, *Le tutorat maître-élève : mesure d'aide à la réussite au collégial*, Rapport de recherche, Québec, Collège Mérici, 1997.
- DÉSILETS, Jean et Lucie DUBÉ, *Programme d'Encadrement Personnalisé : rapport d'évaluation du trimestre 013 remis au comité d'aide à la réussite*, Rimouski, Cégep de Rimouski, 2002.
- FÉDÉRATIONS DES CÉGEPs, *La réussite et la diplomation au collégial : des chiffres et des engagements*, 1999.
- HORVATH, Adam O. et Leslie GREENBERG, « The Development of the Working Alliance Inventory », in GREENBERG, L. et W. PINSOFF (éditeurs), *The Psychotherapeutic Process : A Research Handbook*, New York, The Guilford Press, 1986, p. 529-556.
- LAMOUREUX, Andrée, *Recherche et méthodologie en sciences humaines*, Laval, Éditions Études Vivantes, 2000, 352 p.
- LAROSE, Simon et Rolland ROY, *L'intégration aux études collégiales : problématique, dépistage, intervention et évaluation*, Rapport de recherche, Sainte-Foy, Cégep de Sainte-Foy, 1993.
- MONAGHAN, Denis et Natalie CHALOUX, *Évaluation d'un programme d'aide à la réussite*, Rapport de recherche PAREA, Sainte-Foy, Cégep de Sainte-Foy, 2004.
- RIVIÈRE, Bernard, « Comprendre les décrocheurs afin de mieux les aider », *Pédagogie collégiale*, vol. 9, n° 2, 1995.
- SOUCY, Nathalie, Stéphane DUCHESNE et Simon LAROSE, « Examen des programmes de tutorat maître-élève dans les collèges du réseau québécois », *Pédagogie collégiale*, vol. 13, n° 4, 2000.

Denis MONAGHAN est détenteur d'une maîtrise en psychologie et enseigne cette discipline au cégep de Sainte-Foy depuis 1979. Pendant quatre ans, il a agi comme professeur et tuteur dans un programme d'intégration aux études collégiales. Il a aussi contribué à une autre recherche PAREA portant sur le lien entre les habiletés spatiales et la réussite des cours de physique au collégial.

Natalie CHALOUX enseigne la psychologie depuis 1992. Elle a débuté au cégep de Limoilou et est au cégep de Sainte-Foy depuis 1994. Elle est détentrice d'une maîtrise en psychologie. Elle a œuvré dans divers programmes d'accueil et d'intégration depuis plus de dix ans.