

# Une évaluation de l'impact du diagnostic structuré et instrumenté sur les apprentissages réalisés dans le contexte de l'enseignement collectif\*

Djavid Ajar

Clément Dassa

université de montréal

La régulation des apprentissages fait partie intégrante du projet éducatif et constitue une piste prometteuse d'amélioration de l'efficacité de l'école. Or, sa pratique est fortement contrainte par un ensemble de facteurs dont, notamment, le manque de formation des enseignant-e-s et l'absence d'outils d'évaluation. Cet article présente les résultats d'une étude réalisée auprès de 331 élèves de Secondaire II et portant sur l'évaluation de l'impact du diagnostic structuré et instrumenté sur l'apprentissage des mathématiques. Une des modalités expérimentées portait sur l'utilisation d'un système informatisé de tests adaptatifs qui produit un diagnostic individualisé. Les conclusions portent sur: (a) la supériorité des approches structurées et encadrées par rapport à l'absence de structuration et d'encadrement, (b) l'importance de la participation des enseignants au développement de leurs propres outils d'évaluation, (c) le rôle de l'organisation et de la planification de la régulation dans la réduction des différences inter-individuelles.

The regulation of learning is an integral part of educational policy and a promising approach to improving school effectiveness. However, its practice is seriously constrained by several factors, particularly the inadequate training of teachers and the lack of assessment tools. We present the results of a study evaluating the impact of structured and instrumented diagnosis in mathematics learning carried out on 331 Secondary II students in Quebec. One method we tested uses a computerized adaptive testing system to produce individualized diagnoses. We conclude that structured approaches employed by teachers trained to use them are superior to unstructured approaches used by untrained teachers. We emphasize the importance of teachers' participation in the development of their own assessment tools, and suggest that regulation which is well organized and planned is more likely to reduce differences among individuals.

## PROBLÉMATIQUE

La problématique du diagnostic pédagogique en contexte scolaire s'inscrit au confluent des différentes tendances de l'évaluation des apprentissages qui ont

---

\*Cet article fait état de travaux réalisés dans le cadre d'une subvention du Conseil de recherches en sciences humaines du Canada (subvention no 410-89-1010).

modelé les systèmes scolaires depuis une quarantaine d'années. Dans ce qui suit nous présenterons les principaux courants qui déterminent le cadre conceptuel de cette recherche.

Fondés sur les théories d'apprentissage de Gagné (1965), les développements théoriques de Tyler (1986) et de Bloom (1971) ont profondément influencé la façon dont les systèmes scolaires ainsi que les divers agents éducatifs conçoivent le rôle et les fonctions de l'évaluation des apprentissages réalisée dans le cadre de l'enseignement collectif. S'appuyant sur un fort mouvement de démocratisation de l'éducation, Tyler (1986) a élargi les buts de l'évaluation. Celle-ci passe d'un simple système inductif centré sur l'interprétation de la note d'un examen ou du score d'un test à un système logique de gestion des apprentissages, de l'enseignement et des programmes d'études. L'évaluation cesse alors d'être exclusivement sommative pour devenir aussi formative, c'est-à-dire destinée à guider les apprentissages des élèves ainsi que les stratégies d'enseignement en cours de réalisation. Dans ce contexte, les objectifs opératoires servent à encadrer l'interprétation des comportements des apprenants en situation d'évaluation. En fait, l'évaluation formative s'appuie entièrement sur une planification de l'enseignement et une gestion de l'évaluation centrées sur les objectifs pédagogiques.

Concurremment à ce courant "d'enseignement par objectifs," on observe une recherche grandissante de l'efficacité des systèmes scolaires, des modes de gestion pédagogique ainsi que des apprentissages. La préoccupation constante des gestionnaires en ce qui a trait à la rationalisation des ressources renvoie à la poursuite systématique de la réussite scolaire. Il ne suffit plus d'augmenter les taux de diplomation mais également de s'assurer que les finissants possèdent une formation de qualité ou, à tout le moins, la "meilleure" formation possible. Lorsque cette préoccupation d'efficacité est axée sur le cheminement personnel, elle mène à des formes d'évaluation formative et critériée centrées sur l'identification des difficultés individuelles et leur remédiation (Conseil supérieur de l'Éducation [CSE], 1987). D'une certaine façon, cette recherche de la réussite s'inscrit dans la prolongation des grands principes de démocratisation de l'enseignement: les systèmes d'éducation qui prônaient l'égalité des chances d'accès visent, à présent, à optimiser les chances de réussite. Prise entre la course à la productivité collective et la nécessité de préserver, voire de promouvoir, les droits individuels de l'élève à poursuivre un itinéraire personnel garant de son développement général (Ministère de l'Éducation du Québec [MEQ], 1979), l'école cherche les moyens d'optimiser le rendement scolaire tout en visant à favoriser le "développement holistique, intégral ou global" (CSE, 1990). Dans ce contexte, si l'évaluation se veut formative dans le but de guider l'élève, l'enseignement, lui, se doit d'être différencié, c'est-à-dire d'être "différent selon les élèves par une adaptation qui se situe essentiellement au niveau des méthodes, des cheminements, des exemples, etc." (Cardinet, 1981).

Nous aborderons le problème de l'optimisation du rendement scolaire collectif par le biais du développement des apprentissages individuels et selon une

perspective psychosociale où le couple apprentissage-enseignement est étudié en tant que phénomène interactif. Nous situons l'étude de cette interaction dans son cadre naturel: le groupe-classe. La reconnaissance de l'importance de l'écologie de la classe a été à l'origine d'un ensemble de recherches appelées "Classroom Assessment" (Bateson, 1992; Stiggins, 1988; Stiggins et Bridgeford, 1985). Ce courant nord-américain rejoint les préoccupations européennes en ce qui a trait à l'élargissement de la régulation des apprentissages (Allal, 1988; Cardinet, 1988; Perrenoud, 1991). Idéalement, l'échange entre l'enseignant et l'élève devrait prendre la forme d'un dialogue permanent, d'un "acte global de communication interpersonnelle" où l'interaction entre ces deux acteurs servirait à améliorer les apprentissages. De toute évidence, cette communication ne peut se réaliser indépendamment des multiples contraintes qui modulent la recherche de l'efficacité; en fait, elle se trouve fortement structurée par ces dernières.

Le mouvement de la pédagogie de maîtrise trouve son origine dans la gestion de l'interaction professeur-élève qui vise à optimiser les apprentissages. Selon Bloom (1984), lorsque l'évaluation formative axée sur la régulation des apprentissages est bien outillée et bien gérée, elle *devrait* permettre de se rapprocher du rendement optimal obtenu par le tutorat individuel, c'est-à-dire, toujours selon Bloom, un rendement scolaire supérieur de deux écarts-types (2-sigma) au rendement obtenu par l'enseignement traditionnel collectif. Compte tenu des multiples contraintes qui s'exercent sur la classe, cette cible idéale c'est avérée illusoire. Entre la réalité "écologique" fort complexe de la classe nichée dans des systèmes (école, commission scolaire, système scolaire et société) de plus en plus éloignés de l'interaction enseignant-élève et le projet éducatif centré sur l'atteinte ultime d'un maximum d'objectifs pédagogiques, les nombreuses tentatives d'intégration de l'évaluation formative à l'écologie de la classe ont produit des effets beaucoup moins élevés que prévu. Les résultats de certaines méta-analyses ayant trait à l'efficacité de la pédagogie de maîtrise indiquent une grandeur de l'effet de l'ordre de 0,5-sigma (Kulik, Kulik et Banger-Drowns, 1990; Slavin, 1987, 1990). Au Canada, une étude en mathématiques et en français réalisée auprès de mille élèves du primaire et conçue selon l'approche LFM<sup>1</sup> instrumentée classique a confirmé la supériorité de la pédagogie de maîtrise sur les méthodes traditionnelles n'intégrant pas l'évaluation formative (Ajar, Dassa et Scallon, 1986). L'effet de l'instrumentation<sup>2</sup> est de 0,54 en français et de 0,62 en mathématiques. Cependant, l'étude a révélé la très grande difficulté pour les enseignants d'administrer, de corriger, d'interpréter et d'utiliser les résultats d'un grand nombre de tests critériés (près de 20 tests par matière répartis sur quatre semaines). Bien que convaincus de l'utilité de l'intégration de l'évaluation formative à l'enseignement, les enseignants ont clairement indiqué qu'ils ne pouvaient pas s'approprier la méthode d'enseignement sans formation adéquate, notamment en ce qui concerne le soutien à la production et à la correction des instruments de mesure. Une conclusion s'imposait alors: l'information de nature

diagnostique requise, bien qu'essentielle, ne peut être systématiquement et aisément gérée par les enseignants sans un support constant, particulièrement au chapitre de la correction et de l'interprétation des tests. Si, dans le contexte de l'enseignement collectif, l'ampleur et la complexité des informations issues de tests diagnostiques rendent inévitable le recours à l'informatique (Dassa, 1988; Dassa, Vázquez-Abad et Ajar, 1993), l'intégration de l'évaluation formative aux pratiques pédagogiques requiert, elle, un encadrement des enseignants qui tienne compte de leurs pratiques, c'est-à-dire un encadrement de type coopératif (Dassa, 1990). Un système informatisé adéquat se doit donc de reposer sur les pratiques d'évaluation des enseignants. Afin d'assurer à un tel système un maximum d'efficacité, il convient d'identifier ce qui dans les pratiques d'évaluation s'oppose à l'intégration recherchée.

Des travaux récents permettent de dégager un profil type de l'enseignant, tout acquis aux valeurs et aux principes de l'évaluation formative mais inhibé dans son action par un manque de formation et d'outillage pédagogique (Burelle, Gadbois, Parent et Séguin, 1991; Dumoulin, 1991; Gadbois, Burelle, Parent et Séguin, 1991; Perrenoud, 1988; Rogers, 1990; Séguin, Parent et Burelle, 1993; Stiggins, 1988; Stiggins et Bridgeford, 1985). "Il [s'ensuit] des conflits et des frustrations, donc une régression à des méthodes d'enseignement et d'évaluation plus conformes aux contraintes des maîtres" (Perrenoud, 1988). Les enseignants évaluent leurs élèves en s'aidant d'instruments qu'ils développent eux-mêmes, du mieux qu'ils peuvent, alors que leur formation en mesure et en évaluation est souvent inexistante et qu'ils se trouvent dénués d'outils pédagogiques. Tous soulignent enfin le problème de manque de temps pour développer ou s'approprier un outillage de qualité et pour l'intégrer à leurs prestations.

C'est dans ce contexte que se pose la question de l'intégration de l'évaluation formative à l'enseignement. Quel type d'outillage et d'encadrement des enseignants est le plus apte à améliorer les pratiques de l'évaluation formative en classe? Comment se comparent des pratiques pédagogiques comportant divers niveaux de structuration et d'instrumentation de l'évaluation formative? Dans ce qui suit, nous présenterons la méthodologie développée pour répondre à ces questions.

#### MÉTHODOLOGIE

Afin que la réponse aux questions de recherche soit la plus complète possible, il convenait d'inclure une composante informatique à l'instrumentation développée pour cette recherche (Dassa et al., 1993). C'est pourquoi un système de diagnostic informatisé, destiné à soutenir les pratiques d'évaluation en classe et conçu pour optimiser le diagnostic individuel, a été développé préalablement à la recherche qui fait l'objet de cet article. La justification et le développement du système, ses prémisses théoriques et ses caractéristiques ont fait l'objet de diverses publications (Ajar et Dassa, 1993; Auger et Dassa, 1991; Dassa et

Auger, 1991; Dassa et René de Cotret, 1994). Le lecteur trouvera dans ce qui suit une brève description du système développé ainsi que de l'effet des contraintes du milieu scolaire sur certaines de ses caractéristiques fondamentales. L'utilisation du système informatique par des enseignants encadrés fait partie d'un ensemble de modalités d'intégration du diagnostic pédagogique qui définissent autant de pratiques plus ou moins centrées sur l'individualisation de l'évaluation. L'efficacité de ces modalités a été évaluée à l'aide d'une étude comparative de nature quasi-expérimentale.

### *Objectifs de la recherche*

La recherche visait à évaluer le rôle du diagnostic structuré et instrumenté sur les apprentissages des élèves. Elle avait pour objectif de comparer le rendement scolaire des élèves de deuxième année du secondaire en algèbre et en géométrie soumis à l'une ou l'autre des quatre modalités suivantes: (a) intégration structurée du diagnostic instrumenté, interactif et informatisé, par des enseignants encadrés; (b) intégration structurée du diagnostic instrumenté traditionnel (tests critériés papier-crayon utilisés selon l'approche LFM) par des enseignants encadrés; (c) utilisation libre par des enseignants non-encadrés de tests papier-crayon élaborés ou choisis par eux-mêmes pour répondre aux besoins de leurs propres élèves; (d) utilisation prescrite par des enseignants non-encadrés des tests critériés papier-crayon élaborés par d'autres (les chercheurs et les enseignants encadrés). La troisième modalité a été complétée d'un bilan relatif à la nature et à la fréquence d'utilisation des outils d'évaluation librement utilisés.

### *Choix du contenu*

La matière retenue pour cette recherche porte sur l'introduction à l'algèbre et à la géométrie au secondaire (Secondaire II, au Québec). Ce contenu présente l'avantage de ne pas dépendre directement de préalables et donc de faciliter l'évaluation de la maîtrise de la matière. De plus, l'enseignement des notions élémentaires d'algèbre et de géométrie qui font l'objet de cette étude s'étalent sur plus de deux mois, ce qui permet de réduire les biais potentiels dus à la brièveté de l'intervention pédagogique (Bloom, Hastings et Madaus, 1971). Finalement, le choix des mathématiques repose sur l'existence d'une certaine hiérarchisation de la matière qui, à défaut de faire l'unanimité chez les didacticiens, permet un cheminement pédagogique fortement recommandé par le ministère de l'Éducation du Québec et, en pratique, carrément prescrit par les établissements scolaires.

### *Observation des pratiques*

Les pratiques d'évaluation ont été observées dans les classes de 10 professeurs appartenant à quatre écoles publiques et à une école privée situées en milieu

urbain. Les observations ont été réparties sur deux ans. La première année, quatre classes ont été suivies systématiquement tout au long de l'année académique à raison de 10 séances d'observation d'une durée de 60 minutes (un cours) chacune. La deuxième année, neuf classes ont été observées environ six fois chacune (pendant 15 à 20 minutes par visite) pour une durée d'observation totale de 90 minutes par classe. Ces observations ont permis d'établir un profil détaillé des pratiques d'évaluation des apprentissages scolaires des enseignants de ces écoles.

### *Encadrement des enseignants*

La deuxième année, six professeurs participants appartenant à deux écoles publiques et une privée se sont engagés à suivre six crédits de formation universitaire de premier cycle dans le but de parfaire leurs connaissances des principes et des techniques de mesure et d'évaluation des apprentissages. Concomitamment à cette formation, 15 séances de travail collectif réparties sur toute l'année académique avaient pour but d'analyser la matière ciblée, de définir un domaine de référence unique de façon à constituer un corpus d'outils pédagogiques (exercices, tests, examens, démarches particulières) commun. Les enseignants des deux écoles participantes ont ainsi ciblé huit contenus spécifiques (cinq en algèbre et trois en géométrie) nécessitant un enseignement d'environ 10 semaines. Ces enseignants seront identifiés dans ce qui suit par la désignation d'enseignants "encadrés." Cet encadrement s'est poursuivi au cours de l'étude comparative sur laquelle porte le présent article et qui a été réalisée la troisième année. Cette étude comptait sur la participation de quatre des enseignants encadrés (six classes) précédemment formés et de deux enseignants non-encadrés (six classes).

### *Description du milieu*

L'étude comparative s'est déroulée dans deux écoles situées en milieu urbain, l'une privée et l'autre publique. L'école privée est réservée aux filles et possède une clientèle très largement composée de Québécoises francophones de niveau socio-économique relativement élevé. L'école publique est mixte et accueille très majoritairement des membres des communautés culturelles; sa clientèle provient de milieux dont le niveau socio-économique est relativement faible. Le choix des écoles participantes a été réalisé en tenant compte de plusieurs facteurs. Tout d'abord, la proximité: la fréquence des visites (quasi-quotidiennes par moment) a restreint le choix à la région montréalaise proche de l'université d'attache des chercheurs. Ensuite, les contraintes techniques dues à l'utilisation de l'informatique: seules les écoles dont l'équipement, les installations et les possibilités d'accès pour la recherche étaient compatibles avec les objectifs de l'étude prévue pouvaient être considérées. Finalement, les écoles participantes devaient offrir un niveau de collaboration poussé avec les chercheurs de façon à assurer l'application d'un protocole de recherche de nature collaborative s'appuyant sur la

participation soutenue des divers agents pédagogiques (enseignants, conseillers pédagogiques, responsables des laboratoires informatiques, direction d'école).

### *Description du système diagnostic informatisé*

Les huit contenus spécifiques identifiés au cours de la deuxième année ont permis de créer une banque de près de 2 000 items à choix multiples. À partir de cette banque initiale, un ensemble de 16 tests diagnostiques comportant 591 items ont été sélectionnés de façon à assurer une bonne validité de contenu et ont été placés dans l'environnement MicroCAT (version 3.0). Le système informatique développé vise à produire une information diagnostique structurée et détaillée pour chaque élève ainsi qu'un résumé pour l'ensemble de la classe. Il est constitué d'un ensemble de banques d'items et de tests diagnostiques qui établissent un profil individuel des difficultés. Cette information est complétée de fiches diagnostiques informatisées qui visent à optimiser le "diagnostic pédagogique" proposé par Dassa (1988, 1990). Partie intégrante de l'évaluation formative, le diagnostic pédagogique consiste à : (a) identifier l'erreur, c'est-à-dire dans le contexte scolaire actuel, ce que l'enseignant identifie comme erroné; (b) décrire la nature de l'erreur; (c) décrire le contexte pédagogique où l'erreur se produit; (d) tenter de relier la (les) cause(s) de l'erreur à un paradigme de l'apprentissage; (e) proposer des interventions correctrices s'accordant à la didactique et à la technologie utilisées et s'inscrivant éventuellement dans un cadre théorique préalablement déterminé.

Chaque banque d'items est structurée selon trois dimensions, c'est-à-dire que chaque item d'une banque est localisé dans un espace virtuel tridimensionnel, le "cube diagnostique." Tous les items sont catégorisés en termes de *contexte diagnostique*, de *contenu notionnel* et d'*habileté cognitive*. Chaque dimension est subdivisée en strates. Le contexte diagnostique décrit les strates de mesure de l'item, c'est-à-dire ses différentes variantes. Le contenu notionnel détaille les différentes strates relatives au contenu mathématique d'un objectif pédagogique donné. Quant à l'habileté cognitive, elle désigne le type d'apprentissage visé. Cette catégorie est centrée sur une tâche et exprime des comportements observables. Pour les besoins de cette recherche, les habiletés cognitives étudiées étaient étroitement reliées aux objectifs intermédiaires et terminaux de la matière visée, objectifs prescrits par le ministère de l'Éducation du Québec. L'ensemble des items de l'étude était partitionné selon 16 cubes diagnostiques. La figure 1 présente un exemple de cube; afin de simplifier sa représentation, seule la première habileté cognitive y est représentée. Les items d'un cube donné constituaient une banque comportant, en moyenne, une quarantaine de questions. La figure 2 présente un exemple de question. Pour chaque banque, deux modalités distinctes d'administration de test ont été développées.

| Contextes diagnostiques |                                    |                                         | Contenu notionnel              | PRISME DROIT               |             |
|-------------------------|------------------------------------|-----------------------------------------|--------------------------------|----------------------------|-------------|
|                         |                                    |                                         |                                | Aire latérale              | Aire totale |
| PROBLÈMES ABSTRACTS     | Mesures identifiées sur une figure | “aire” présente nombres entiers         | 834 844 850<br>851 904         | 808 825 839<br>845 853 855 |             |
|                         |                                    | expression équivalente nombres entiers  |                                |                            |             |
|                         |                                    | “aire” présente nombres décimaux        |                                |                            |             |
|                         |                                    | expression équivalente nombres décimaux | 901                            | 836 837 902<br>909         |             |
| PROBLÈMES RÉELS         | Mesure identifiées dans un texte   | “aire” présente nombres entiers         | 807 810 811 812<br>826 832 838 | 823 841 905<br>908 922     |             |
|                         |                                    | expression équivalente nombres décimaux |                                |                            |             |
|                         |                                    | “aire” présente nombres entiers         |                                |                            |             |
|                         |                                    | expression équivalente nombres décimaux |                                |                            |             |

FIGURE 1

*Exemple de cube diagnostique (partiel)  
(seule la première habileté cognitive est représentée)*



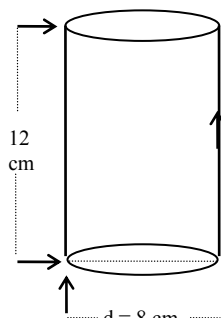
HEOMH918 — Calcule l'aire totale du solide suivant:

(a) 401,92 cm<sup>2</sup>

(b) 301,44 cm<sup>2</sup>

(c) 602,88 cm<sup>2</sup>

(d) 351,68 cm<sup>2</sup>



( $\pi = 3,14$ )

Figure 2

*Exemple de questions de la banque d'items  
(seule la première habileté cognitive est représentée)*

En mode “laboratoire,” l’ensemble des élèves de la classe se rend au laboratoire d’informatique et chaque élève passe individuellement un test adaptatif. Chaque séance dure de 12 à 20 minutes selon le profil individuel des réponses. La séance se termine par l’apparition à l’écran d’une fiche diagnostique qui décrit de façon détaillée les difficultés rencontrées; une copie imprimée destinée au professeur est ultérieurement produite. Cette information est structurée selon deux dimensions diagnostiques principales: le *diagnostic structurel*, qui identifie la nature de l’erreur observée en relation avec le cube diagnostique et le *diagnostic causal*, qui propose une explication psychopédagogique de l’erreur commise. À la fin de la séance, une fiche diagnostique synthèse résume les résultats collectifs de l’ensemble de la classe en identifiant les parties du cube qui ont été manquées, le type d’erreur le plus fréquent ainsi que les élèves présentant le plus de difficultés.

En mode “récupération,” l’enseignant rencontre quelques élèves présentant des difficultés qu’il juge importantes (typiquement entre trois et cinq élèves). La séance de récupération dure une heure et a lieu en dehors de la classe, dans un local réservé à cette fin et équipé d’un micro-ordinateur. Le système diagnostique de cette modalité vise à produire un diagnostic plus fouillé que celui produit en laboratoire. Il est ouvert à l’intervention du professeur qui définit, via une

interface de menus, le domaine de référence de la session. L'interaction enseignant-élève se vit donc via le système qui, à la fin de chaque étape, résume l'information et propose un diagnostic. De cette façon, l'enseignant peut optimiser ses interventions en décidant, à certains moments—où il ne dispose pas d'hypothèse sur la source des difficultés de l'élève—de laisser l'élève répondre aux questions du système pendant qu'il interroge oralement les autres élèves ou discute de leurs difficultés respectives, voire collectives.

Étant donné la nature stratifiée des dimensions sous-jacentes aux banques d'items, la sélection des items se fait selon une stratégie adaptative stratifiée. La stratégie est dite adaptative parce qu'elle s'exécute au fur et à mesure que l'élève répond aux questions et selon la nature de ses réponses. Elle permet d'individualiser la séance de testing en fonction de la réponse donnée par le répondant. Typiquement, le test d'un élève moyen comportait une douzaine d'items, une vingtaine d'items pour les élèves en difficulté et aussi peu que huit items pour les élèves ne présentant pas d'échec. De plus, la stratégie est dite stratifiée parce qu'elle intervient systématiquement selon un regroupement de strates délimité à partir de règles ad hoc qui tiennent compte de deux contraintes importantes: le nombre d'items relevant du contexte testé et la durée de la séance de test limitée par la grille-horaire du laboratoire. La répartition des items de chaque strate a été hiérarchisée par ordre de difficulté selon le jugement des enseignants subséquentement validé par un didacticien des mathématiques expert du contenu (voir à cet effet, Dassa et René de Cotret, 1994).

#### *Contexte de développement du système diagnostic informatisé*

En l'absence d'une taxonomie diagnostique générale et compte tenu des multiples contraintes propres au milieu scolaire en ce qui a trait à l'évaluation des apprentissages, force était d'impliquer les enseignants à l'élaboration du système diagnostic informatisé. Afin de s'assurer que les enseignants encadrés utiliseraient le système sans réticences et adéquatement pendant toute une année académique, les chercheurs ont opté pour une stratégie de recherche rétroactive. Selon cette approche, les chercheurs et les enseignants ont coopéré au développement de l'outil informatique. Les chercheurs ont assumé l'encadrement théorique et technique de la production et de l'utilisation du système informatique en étroite et constante collaboration avec les enseignants participants. Ces derniers ont notamment produits une première version de la banque d'items. Ils ont également collaboré à sa structuration en trois dimensions en fonction de leurs propres pratiques pédagogiques. Cet encadrement réciproque des chercheurs et des enseignants a créé une dynamique de rétroaction qui a permis de développer un système ayant des caractéristiques respectant les contraintes relatives au curriculum ainsi qu'à l'organisation et à la gestion de l'évaluation.

Cette façon de procéder optimise les chances que la gestion et l'interprétation des résultats soient conformes aux pratiques et aux attentes des enseignants. Elle

réduit les obstacles à l'utilisation du système informatisé en lui assurant des caractéristiques compatibles avec les diverses contraintes. Celles-ci sont organisationnelles, relatives à l'horaire des cours, à l'accessibilité du laboratoire d'informatique et à la plage horaire réservée aux séances de récupération. Elles portent aussi sur l'environnement informatique (capacité des micro-ordinateurs disponibles en laboratoire, type d'écran et système d'opération).

*Échantillon, population et schème quasi-expérimental*

Un ensemble de 12 groupes-classes comportant 331 élèves ont servi à cette étude. Sept groupes totalisant 182 élèves provenaient d'une école publique (une des quatre écoles publiques où les pratiques d'évaluation ont été observées) et 5 groupes totalisant 149 élèves provenaient de l'école privée. Bien que ces groupes aient été constitués par les administrateurs des écoles participantes et ne pouvaient être défaits au profit d'une répartition strictement aléatoire, l'examen des modalités d'allocation n'a révélé aucun biais systématique, aucun placement selon des variables cognitives ou affectives connues. De l'avis des enseignants responsables de ces groupes, ils étaient typiques des clientèles de leurs écoles respectives. Du fait que l'encadrement des enseignants était exigeant et nécessitait une collaboration soutenue, les enseignants encadrés ont été choisis sur la base de leur intérêt vis-à-vis de la recherche. Les groupes-classes constitués leur ont été assignés par les administrateurs; aucune caractéristique particulière des enseignants n'a apparemment influencé l'attribution de ces groupes. À l'exception d'un enseignant de l'école privée, qui n'avait que 3 années d'expérience, les enseignants participants avaient entre 15 et 20 années d'expérience. Il convient de souligner que la modalité non-aléatoire de la sélection des enseignants alliée à la nature des écoles participantes — l'une publique avec une très grande majorité de représentants des communautés culturelles et l'autre réservée aux filles et privée — ne permet pas de généraliser les résultats de l'échantillon à l'ensemble de la population des élèves de Secondaire II d'une région ou, *a fortiori*, à l'ensemble du réseau scolaire québécois. D'ailleurs, telle n'était pas notre intention. Notre but en sélectionnant ces écoles était de nous assurer de la collaboration, non seulement des enseignants participants, mais également des responsables pédagogiques ainsi que des responsables de l'enseignement de l'informatique qui contrôlent l'accès aux laboratoires. Le schème retenu était donc quasi-expérimental et avait pour but d'établir une base objective de comparaison des différentes modalités d'intégration du diagnostic pédagogique de façon à pouvoir les évaluer compte tenu du contexte de cette comparaison, c'est-à-dire, entre autres facteurs pertinents, le milieu de l'expérimentation.

Le schème retenu comporte deux facteurs de classification: la nature du diagnostic pédagogique et le type d'école. Le premier facteur comporte deux sous-facteurs relatifs à la présence ou à l'absence d'un encadrement des enseignants participants. Cet encadrement visait à former les enseignants à comprendre

l'approche d'évaluation formative fondée sur un diagnostic structuré et à l'utiliser à des fins correctives. L'ensemble des outils d'évaluation (tests, examens et modalités de diagnostics et de correctifs pédagogiques) nécessaires ont été élaborés en étroite collaboration avec les quatre enseignants participants qui étaient assignés aux modalités "expérimentales."

Le premier sous-facteur (enseignants encadrés) est constitué de deux niveaux (D1 et D2) qui constituent autant de modalités expérimentales de structuration du diagnostic. La modalité D1 porte sur l'utilisation de tests diagnostics interactifs (individualisés) administrés par *ordinateur* selon deux modalités pédagogiques complémentaires: (a) collectivement en mode "laboratoire" et (b) individuellement en mode "récupération." Un ensemble de 110 élèves ont suivi cette modalité: 52 élèves répartis en deux classes avec deux enseignants de l'école publique, et 58 élèves du secteur privé répartis en deux classes avec le même enseignant. La deuxième modalité (D2) consiste en l'utilisation de tests *traditionnels* ("papier-crayon") — portant sur les mêmes objectifs et contenus que les tests informatisés de D1 — administrés collectivement en classe. Un ensemble de 49 élèves répartis en deux classes avec un enseignant du secteur public ont été assignés à cette modalité.

Le deuxième sous-facteur (enseignants non-encadrés) comporte deux niveaux qui déterminent autant de modalités de contrôle (D3 et D4). La modalité D3 est définie par l'utilisation *libre* d'outils d'évaluation élaborés ou choisis (parmi les outils de D2) par l'enseignant. Ce dernier ne bénéficie d'aucun encadrement en ce qui concerne les modalités d'élaboration, d'interprétation ou d'utilisation à des fins correctives des instruments servant à l'évaluation. Cette modalité vise à reproduire le type d'évaluation réalisée par un enseignant qui, en plus de ses propres outils, aurait accès à une instrumentation adéquate mais élaborée par d'autres que lui-même. D3 a été suivie par un enseignant du secteur privé qui a enseigné à trois classes regroupant 91 élèves. La dernière modalité (D4) porte sur l'utilisation *prescrite* des mêmes tests traditionnels ("papier-crayon") que ceux de D2 — administrés collectivement en classe. Elle n'est soutenue par aucun encadrement sur la façon de poser un diagnostic ou d'utiliser les résultats à des fins correctives. Cette modalité représente une tendance récente du milieu scolaire à *prescrire* des instruments élaborés par d'autres, sans aucune participation des enseignants à leur élaboration. Trois classes regroupant 81 élèves et un enseignant de l'école publique ont suivi D4.

Le deuxième facteur contrôle les deux types d'écoles participantes: une publique et une privée.

### *Hypothèses et types d'analyses*

Les trois hypothèses de recherche<sup>3</sup> suivantes ont été testées:

H1: Les modalités portant sur le diagnostic structuré réalisées par des enseignants encadrés présentent des effets moyens supérieurs aux modalités

relatives au diagnostic non-structuré réalisées par des enseignants non-encadrés.

- H2: Parmi les modalités portant sur le diagnostic non-structuré, l'utilisation libre par l'enseignant de ses propres outils d'évaluation ou de tests traditionnels élaborés par d'autres que lui-même (modalité D3) présente des effets moyens supérieurs à l'utilisation prescrite de tests traditionnels élaborés selon les règles mais hors de toute collaboration avec l'enseignant (modalité D4).
- H3: Les modalités qui visent à optimiser le diagnostic pédagogique individuel permettent de réduire le plus les différences de rendement inter-individuelles. Selon cette hypothèse, une hiérarchie de réduction croissante doit exister pour la séquence D4, D3, D2, D1. Plus précisément, la réduction de variance entre le prétest et le post-test est d'autant plus importante que la modalité expérimentale est plus centrée sur les difficultés individuelles.

La première hypothèse vise à confirmer l'efficacité relative de la structuration du diagnostic couplée à l'encadrement des enseignants. La deuxième hypothèse pose que l'utilisation d'outils d'évaluation conçus ou librement utilisés par l'enseignant selon les besoins de ses élèves, c'est-à-dire intégrés à ses pratiques, est plus efficace que l'utilisation d'outils élaborés par d'autres et administrés selon un découpage pédagogique ne tenant pas nécessairement compte des difficultés propres à ses élèves. La dernière hypothèse concerne une conjoncture prévue par les tenants de la pédagogie de maîtrise, à savoir l'effet de plus en plus réducteur des différences entre les apprentissages des élèves lorsque le mode d'évaluation diagnostique est de plus en plus centré sur les difficultés individuelles.

Les deux premières hypothèses ont été testées à l'aide d'analyses de variance et de covariance. Un post-test sur l'algèbre et la géométrie servait de critère et, pour les analyses de covariance, un prétest servait de variable concomitante. La troisième hypothèse a été testée à l'aide de tests statistiques univariés sur les variances des prétests et des post-tests.

## RÉSULTATS

Les analyses ont été réalisées à l'aide des logiciels SPSS et BMDP. Les tableaux 1 à 4 présentent les statistiques descriptives et inférentielles issues de l'expérimentation.

L'examen des moyennes ajustées des post-tests (voir tableau 1) révèle que le gain dû à la structuration du diagnostic et à l'encadrement des enseignants est plus élevé pour les élèves de l'école publique que pour ceux de l'école privée. Cette interaction est confirmée ( $p < ,05$ ) par l'analyse des covariances à deux facteurs (cf. tableau 2) qui compare les modalités de diagnostic structuré et des enseignants encadrés (D1 et D2) aux modalités de diagnostic non-structuré et des

enseignants non-encadrés (D3 et D4). Qui plus est, l'école publique présente une différence significative ( $p < ,001$ ) entre les modalités expérimentales et contrôles alors que l'école privée n'en présente pas ( $p = ,159$ ). L'interprétation de ces résultats passe par le constat de l'existence de différences très marquées entre les deux écoles. Ainsi, les études qualitatives relatives aux pratiques d'évaluation en classe de ces deux écoles (réalisées au cours de la deuxième année de la recherche) ont accessoirement révélé des différences fondamentales entre leurs clientèles en ce qui a trait aux origines culturelles, aux niveaux socio-économiques et au rendement scolaire. De toute évidence, la clientèle de l'école privée présentait des caractéristiques permettant de prévoir un rendement scolaire plus élevé que les élèves de l'école publique: elle provient d'un milieu culturellement plus homogène et de niveau socio-économique beaucoup plus élevé. Par ailleurs, il est normal que la recherche d'une plus grande efficacité pédagogique bénéficie relativement plus aux élèves dont le rendement scolaire est plus faible; il s'agit en fait d'une prémisse de la pédagogie de maîtrise.<sup>4</sup> Puisque le but de notre recherche n'était pas de comparer les écoles mais plutôt de décrire l'effet des modalités étudiées sur une population ad hoc dont les caractéristiques étaient connues, nous procéderons dans ce qui suit à des analyses à un seul facteur relatif à la nature du diagnostic. En éliminant le facteur "école," on s'assure que les conclusions auxquelles on arrive par inférence<sup>5</sup> pragmatique ont le plus grand

TABLEAU 1

*Moyennes,\* écarts-types et tailles des échantillons aux prétests (X) et aux post-tests (Y) selon la nature du diagnostic pédagogique et le type d'école*

|                       | <i>Diagnostic structuré<br/>enseignants encadrés</i> |                  |                 | <i>Diagnostic non-structuré<br/>enseignants non-encadrés</i> |                  |                 |
|-----------------------|------------------------------------------------------|------------------|-----------------|--------------------------------------------------------------|------------------|-----------------|
|                       | <i>X</i>                                             | <i>Y observé</i> | <i>Y ajusté</i> | <i>X</i>                                                     | <i>Y observé</i> | <i>Y ajusté</i> |
| <i>École publique</i> |                                                      |                  |                 |                                                              |                  |                 |
| moyenne               | 30,0                                                 | 73,4             | 72,2            | 33,5                                                         | 64,4             | 61,4            |
| écart-type            | 19,0                                                 | 16,2             |                 | 13,1                                                         | 17,5             |                 |
| taille                |                                                      | 101              |                 |                                                              | 81               |                 |
| <i>École privée</i>   |                                                      |                  |                 |                                                              |                  |                 |
| moyenne               | 21,0                                                 | 75,6             | 78,6            | 22,8                                                         | 73,0             | 75,1            |
| écart-type            | 13,9                                                 | 15,7             |                 | 10,3                                                         | 16,3             |                 |
| taille                |                                                      | 58               |                 |                                                              | 91               |                 |

\* Les moyennes ajustées sont calculées à partir de l'ANCOVA à deux facteurs (École x Diagnostic) présentée au tableau 2.

TABLEAU 2

*Analyses de covariance selon les écoles et la nature du diagnostic*

| <i>Selon les facteurs<br/>(École x Diagnostic)</i> |           |           |          |              | <i>Selon le type d'école</i> |           |           |          |              |
|----------------------------------------------------|-----------|-----------|----------|--------------|------------------------------|-----------|-----------|----------|--------------|
| <i>Source de variation</i>                         | <i>dl</i> | <i>CM</i> | <i>F</i> | <i>prob.</i> | <i>Source de variation</i>   | <i>dl</i> | <i>CM</i> | <i>F</i> | <i>prob.</i> |
| <i>Covariable</i>                                  | 1         | 15 959,97 | 71,40    | ,000         | <i>D: école</i>              |           |           |          |              |
| <i>École (É)</i>                                   | 1         | 7 279,88  | 32,57    | ,000         | publique                     | 1         | 5 164,79  | 23,10    | ,000         |
| <i>Diag. (D)</i>                                   | 1         | 4 026,23  | 18,01    | ,000         | privée                       | 1         | 446,19    | 2,00     | ,159         |
| <i>É x D</i>                                       | 1         | 1 028,96  | 4,60     | ,033         |                              |           |           |          |              |
| <i>Résiduelle</i>                                  | 326       | 223,54    |          |              |                              |           |           |          |              |

degré de précision possible compte tenu des données disponibles, c'est-à-dire qu'on augmente la validité externe de la recherche.

Les tableaux 3 et 4 présentent les statistiques descriptives et inférentielles selon les quatre modalités retenues. En préalable à l'interprétation des résultats et à cause des différentes tailles d'échantillons, un test d'homogénéité des variances a été réalisé et s'est avéré non significatif ( $p = ,795$ ). L'analyse de variance sur les post-tests révèle une différence significative ( $p < ,001$ ) entre les modalités expérimentales. Un ensemble de sept comparaisons multiples ont été planifiées et analysées par la méthode de Dunn. Le premier contraste compare l'effet moyen des modalités D1 et D2 avec celui de D3 et D4. Le test est significatif ( $p < ,05$ ) et l'hypothèse H1 est donc confirmée. Les six autres contrastes permettent d'établir une hiérarchie d'efficacité entre les quatre modalités. D1, D2 et D3 sont chacune supérieures à D4. L'hypothèse H2 est donc confirmée: la modalité D3 présente des effets moyens supérieurs à ceux de D4.

À ce sujet, il convient de souligner la grande variété des outils d'évaluation utilisés par les groupes de la modalité D3. Le bilan prévu pour cette modalité a révélé une utilisation judicieuse: (a) de brefs tests collectifs de cinq minutes ("quiz") donnés au début du cours et visant à vérifier sur place la compréhension du cours précédent (les élèves communiquaient immédiatement les réponses à main levée), (b) d'un grand nombre d'exercices pertinents corrigés et discutés en classe sans délai, (c) d'une grande variété de format des tests et des examens utilisés, et (d) d'évaluations collectives planifiées et de retours individuels fréquents.

Finalement, l'analyse de covariance qui tient compte de l'effet implicite des préalables confirme les résultats de l'analyse de variance et des contrastes.

La comparaison des modalités du point de vue de leur effet réducteur sur les différences inter-individuelles est réalisée en testant au préalable l'hypothèse

d'homogénéité de variance des prétests et des post-tests considérés séparément (test de Bartlett-Box) puis en testant l'égalité des variances des prétests et des post-tests de chacune des modalités. Dans ce dernier cas, le test statistique tient compte du lien de dépendance qui existe entre chaque prétest et post-test. Les tests de Bartlett-Box montrent que les prétests et les post-tests ont des variances respectivement hétérogènes (11,296 avec  $p < ,001$ ) et homogènes (0,342 avec  $p = ,795$ ). Initialement hétérogènes, le rendement scolaire converge suite au traitement vers des variances homogènes. Lorsque les prétests et les post-tests sont comparés pour chaque modalité, leurs variances sont significativement différentes ( $p \leq ,005$ ) pour D3 et D4. Les statistiques de ces deux modalités sont négatives (−4,748 pour D3 et −2,764 pour D4), indiquant par là une augmentation de la variance entre le prétest et le post-test; les modalités de diagnostic non-structuré et d'enseignants non-encadrés entraînent un accroissement des différences inter-individuelles. La modalité D2 possède une probabilité élevée ( $p = ,427$  pour une statistique de 0,185), qui permet de conclure que les différences entre individus ne changent pas lorsque des enseignants encadrés structurent leur diagnostic à partir de tests traditionnels élaborés par eux-mêmes. Il en est de même pour D1 ( $p = ,079$  pour une statistique de 1,447). Les résultats indiquent que l'hypothèse H3 n'est que partiellement confirmée. Essentiellement, les différences inter-individuelles restent inchangées pour les modalités expérimentales (D1 et D2) alors qu'elles augmentent pour les modalités de contrôle (D3 et D4).

TABLEAU 3

*Moyennes, écarts-types et tailles des échantillons aux prétests (X) et aux post-tests (Y) selon la nature du diagnostic pédagogique*

|                   | <i>Diagnostic structuré<br/>enseignants encadrés</i>         |          |                                                              |          | <i>Diagnostic non-structuré<br/>enseignants non-encadrés</i> |          |                                                       |          |
|-------------------|--------------------------------------------------------------|----------|--------------------------------------------------------------|----------|--------------------------------------------------------------|----------|-------------------------------------------------------|----------|
|                   | <i>tests<br/>informatisés et<br/>individualisés<br/>(D1)</i> |          | <i>tests<br/>traditionnels en<br/>collaboration<br/>(D2)</i> |          | <i>tests<br/>traditionnels<br/>usage libre<br/>(D3)</i>      |          | <i>tests<br/>traditionnels<br/>prescrits<br/>(D4)</i> |          |
|                   | <i>X</i>                                                     | <i>Y</i> | <i>X</i>                                                     | <i>Y</i> | <i>X</i>                                                     | <i>Y</i> | <i>X</i>                                              | <i>Y</i> |
|                   |                                                              |          |                                                              |          |                                                              |          |                                                       |          |
| <i>moyenne</i>    | 25,7                                                         | 74,8     | 28,9                                                         | 72,8     | 22,8                                                         | 73,0     | 33,5                                                  | 64,4     |
| <i>écart-type</i> | 18,5                                                         | 16,3     | 15,9                                                         | 15,5     | 10,3                                                         | 16,3     | 13,1                                                  | 17,5     |
| <i>taille</i>     | 110                                                          |          | 49                                                           |          | 91                                                           |          | 81                                                    |          |



TABLEAU 4

*Analyses de variance et de covariance selon la nature du diagnostic pédagogique*

| Analyses relatives au facteur "diagnostic"                  |  |              |          |         |                                               |                     |  |                       |           |                   |       |
|-------------------------------------------------------------|--|--------------|----------|---------|-----------------------------------------------|---------------------|--|-----------------------|-----------|-------------------|-------|
| ANOVA sur post-test                                         |  |              |          |         | ANCOVA sur post-test avec contrôle du prétest |                     |  |                       |           |                   |       |
| Source de variation                                         |  | dl           | CM       | F       | prob.                                         | Source de variation |  | dl                    | CM        | F                 | prob. |
| Entre                                                       |  | 3            | 1 879,64 | 6,91    | ,000                                          | Covariable          |  | 1                     | 10 266,82 | 45,21             | ,000  |
| Résiduelle                                                  |  | 327          | 272,17   |         |                                               | Entre               |  | 3                     | 3 444,69  | 15,17             | ,000  |
|                                                             |  |              |          |         |                                               | Résiduelle          |  | 326                   | 227,11    |                   |       |
| Comparaisons multiples planifiées (Dunn)                    |  |              |          |         |                                               |                     |  |                       |           |                   |       |
| Contraste ( $\hat{C}$ )                                     |  | $t(\hat{C})$ |          | signif. |                                               | Modalité            |  | Déviation non-ajustée |           | Déviation ajustée |       |
| $\hat{C}_1 = \bar{y}_1 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4$ |  | 2,690        |          | *       |                                               | D1                  |  | 3,38                  |           | 4,11              |       |
| $\hat{C}_2 = \bar{y}_1 - \bar{y}_2$                         |  | 0,704        |          |         |                                               | D2                  |  | 1,38                  |           | 0,63              |       |
| $\hat{C}_3 = \bar{y}_1 - \bar{y}_3$                         |  | 0,804        |          |         |                                               | D3                  |  | 1,50                  |           | 3,53              |       |
| $\hat{C}_4 = \bar{y}_1 - \bar{y}_4$                         |  | 4,339        |          | **      |                                               | D4                  |  | -7,10                 |           | -9,92             |       |
| $\hat{C}_5 = \bar{y}_2 - \bar{y}_3$                         |  | 0,039        |          |         |                                               | Eta-carré           |  | 0,24                  |           | 0,34              |       |
| $\hat{C}_6 = \bar{y}_2 - \bar{y}_4$                         |  | 2,842        |          | *       |                                               |                     |  |                       |           |                   |       |
| $\hat{C}_7 = \bar{y}_3 - \bar{y}_4$                         |  | 3,413        |          | **      |                                               |                     |  |                       |           |                   |       |

\* $p < ,05$ ; \*\* $p < ,01$

## DISCUSSION ET CONCLUSION

À l'intérieur des limites de cette recherche, il a été établi que les approches structurées et encadrées sont plus efficaces que l'absence de structuration et d'encadrement. Un enseignant bien formé et qui participe pleinement à un encadrement pédagogique relatif aux pratiques d'évaluation réussira à augmenter le rendement scolaire moyen de ses élèves. Par ailleurs, l'absence de structuration et d'encadrement semble entraîner une augmentation des différences inter-individuelles. La généralisation de ces résultats est réduite du fait des multiples contraintes inhérentes au milieu scolaire, contraintes qui ont contribué à restreindre la portée du schème retenu. Malgré des efforts particuliers au chapitre de la validité externe, la nature intensive et surtout coopérative des interventions, alliée à l'utilisation de l'informatique, a également réduit la portée des conclusions. Les auteurs ne prétendent donc pas que les résultats rapportés se généralisent à l'ensemble de la population d'élèves de Secondaire II, mais simplement à la population ad hoc décrite. De ce point de vue, les conclusions se réfèrent strictement à l'étude évaluative rapportée; leur généralisation relève de l'inférence pragmatique. C'est pourquoi la validité de toute extension dépend des caractéristiques du milieu considéré.

En ce qui concerne le système de diagnostic informatisé utilisé, il convient de souligner que sa plus grande efficacité par rapport à l'instrumentation traditionnelle prescrite ne permet pas de conclure à la supériorité de l'informatisation de l'évaluation par rapport à des approches usuelles. En effet, l'utilisation d'un éventail d'outils non-informatisés mais élaborés ou sélectionnés en toute liberté par des enseignants expérimentés s'est avérée tout aussi efficace que l'usage d'un système de diagnostic informatisé. Ce qui semble être en cause ici serait la participation des enseignants au développement ou à la sélection de leur propre outillage. De ce point de vue, les résultats sont éloquentes: lorsque la base de comparaison est la modalité D4, on trouve des grandeurs des effets de 1,04, 0,74 et 1,10 respectivement en faveur de D1, D2 et D3, les modalités qui permettent à l'enseignant de participer à la détermination de son corpus d'instruments. Force est de conclure à l'importance de la participation des enseignants au développement de leurs propres outils d'évaluation.

L'absence de différences entre les modalités informatique (D1) et traditionnelle (D2) relatives à la structuration et à l'encadrement pourrait s'expliquer par le fait que la modalité D2 était fondée sur la même conception de l'intégration du diagnostic à l'évaluation formative que D1. Cette interprétation s'inscrit dans le cadre des conjectures générales de Clark (1985), Viau et Stolovitch (1987), Baird (1989) ainsi que Vázquez-Abad et LaFleur (1990), relatives à l'efficacité des applications pédagogiques de l'ordinateur. Ces suppositions ont été reprises par Dassa et al. (1993) dans le contexte de l'évaluation informatisée, à savoir que la conception systématique de l'intervention pédagogique aurait une plus grande influence sur les résultats scolaires que le fait d'utiliser un moyen ou une

technologie particulière. La présente étude souligne donc l'importance de l'organisation de l'évaluation et de la planification systématique des interventions diagnostiques et correctives indépendamment de l'utilisation de l'informatique.

Afin de décider si le système de diagnostic informatisé utilisé est suffisamment efficace pour être utilisé régulièrement, il conviendrait de comparer son utilisation *non-encadrée*—modalité conforme à la réalité scolaire—à un usage entièrement libre d'outils traditionnels. Une telle étude a été réalisée par les auteurs et fera l'objet d'une publication ultérieure. D'autres avenues de développement portent sur l'amélioration de la stratégie de sélection des items qui gagnerait à s'appuyer sur une modélisation plus approfondie des difficultés cognitives; des travaux en ce sens sont présentement en cours.

#### NOTES

- <sup>1</sup> Il s'agit de l'approche "Mastery Learning" de Bloom (1971), une variante de la pédagogie de maîtrise qui utilise des tests critériés administrés à la suite d'une période d'enseignement planifiée selon les objectifs pédagogiques.
- <sup>2</sup> La grandeur de l'effet est calculée selon la méthode de Slavin (1987, p. 184).
- <sup>3</sup> Il s'agit des hypothèses posées par les chercheurs à partir de leur cadre conceptuel. Ces énoncés ne formulent donc pas nécessairement les hypothèses nulles requises pour les tests statistiques.
- <sup>4</sup> À cet effet, voir par exemple, Bloom et al. (1971) et Slavin (1987, 1990). Concernant ce point, il convient de rappeler qu'indépendamment de l'approche pédagogique retenue, il est habituellement plus facile pour un élève dont le rendement scolaire est faible d'améliorer ses notes que pour un élève dont le rendement scolaire est initialement plus élevé. La pédagogie de maîtrise vise à réduire les différences inter-individuelles au delà de cet effet.
- <sup>5</sup> La nature quasi-expérimentale du schème retenu entraîne des restrictions formelles en ce qui a trait à la généralisation des résultats. Dans ce cas-ci, à cause de l'absence de randomisation des écoles et des enseignants, l'inférence à une population plus vaste que celle échantillonnée est de nature pragmatique. Il convient donc de ne pas réduire indûment la portée des résultats en les restreignant à un type de clientèle pour lequel le traitement est particulièrement efficace, à savoir les élèves en difficulté (plus nombreux dans une école publique à forte dominance de clientèles provenant des communautés culturelles et ayant un niveau socio-économique faible que dans une école privée, culturellement homogène et de niveau socio-économique nettement plus élevé). C'est pourquoi la population ad hoc, définie en combinant les élèves des deux écoles plutôt qu'en ne conservant que ceux de l'école publique, permet de mieux refléter l'effet que les traitements pourraient avoir sur un ensemble d'élèves relativement plus hétérogène. Lorsque les échantillons sont combinés, la grandeur de l'effet est réduite; de ce point de vue, les analyses réalisées toutes écoles confondues sont plus "conservatrices" que celles réalisées sur l'école publique. Le fait que les conclusions s'appliquent à une population mixte plus hétérogène du point de vue socio-économique et culturelle que les populations d'origines de chacune des deux écoles prises séparément permet une inférence pragmatique plus large contribuant ainsi à augmenter la validité externe de la recherche.

#### RÉFÉRENCES

- Ajar, D. et Dassa, C. (1993). *Élaboration et validation d'une banque d'items et de tests* (Guide technique). Montréal: Université de Montréal, Faculté des sciences de l'éducation.
- Ajar, D., Dassa, C. et Scallan, G. (1986, avril). *The integration of diagnostic testing to formative*

- evaluation: A Quebec experiment.* Communication présentée au congrès annuel de l'American Educational Research Association, San Francisco.
- Allal, L. (1988). Vers un élargissement de la pédagogie de maîtrise: processus de régulation interactive, rétroactive et proactive. In M. Huberman (dir.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise* (p. 86–126). Paris: Delachaux et Niestlé.
- Auger, R. et Dassa, C. (1991). L'intégration du diagnostic interactif à l'évaluation formative informatisée des apprentissages scolaires. In B. Almudever et J. Ardoine (dirs.), *Les évaluations* (p. 11–14). Toulouse: Université de Toulouse Le Mirail.
- Baird, W. E. (1989). Status of use: Microcomputers and science teaching. In J. D. Ellis (dir.), *1988 AETS yearbook: Information technology and science education* (p. 85–104). Columbus, OH: AETS and ERIC Clearinghouse.
- Bateson, D. (1992). *Classroom testing in Canada*. Vancouver: University of British Columbia, Centre for Applied Studies in Evaluation.
- Bloom, B. S. (1971). Master learning. In J. H. Block (dir.), *Mastery learning: Theory and practice* (p. 47–63). New York: Holt, Rinehart and Winston.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of instruction as effective as one-to-one tutoring. *Educational Research*, 13(6), 4–16.
- Bloom, B. S., Hastings, J. T. et Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Burelle, R., Gadbois, L., Parent C. et Séguin, S. Ph. (1991). Mesurer des croyances et des attitudes pour mieux comprendre des comportements. *Mesure et évaluation en éducation*, 14(1), 61–77.
- Cardinet, J. (1981). L'évaluation formative, problème actuel. In L. Allal, J. Cardinet et Ph. Perrenoud (dirs.), *L'évaluation formative dans un enseignement différencié* (p. 10–18). Berne: Éditions Peter Lang.
- Cardinet, J. (1988). La maîtrise, communication réussie. In M. Huberman (dir.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise* (p. 155–195). Paris: Delachaux et Niestlé.
- Clark, R. E. (1985). Confounding in educational computing research. *Journal of Educational Computing Research*, 1, 137–148.
- Conseil supérieur de l'Éducation [CSE]. (1987). *La qualité de l'éducation: un enjeu pour chaque établissement* (Rapport 1986–1987 sur l'état et les besoins de l'éducation). Québec: CSE.
- Conseil supérieur de l'Éducation [CSE]. (1990). *La pédagogie, un défi majeur de l'enseignement supérieur* (Avis au ministre de l'Enseignement supérieur et de la Science). Québec: CSE.
- Dassa, C. (1988). L'intégration du diagnostic pédagogique aux apprentissages scolaires: de la théorie à la voie informatique. *Mesure et évaluation en éducation*, 11(1), 7–26.
- Dassa, C. (1990). From a horizontal to a vertical method of integrating educational diagnosis with classroom assessment. *Alberta Journal of Educational Research*, 36, 35–44.
- Dassa, C. et Auger, R. (1991, juin). *Computerized, interactive diagnosis in a classroom setting*. Communication présentée au congrès annuel de la Société canadienne pour l'étude de l'éducation, Kingston.
- Dassa, C. et René de Cotret, S. (1994). Validation d'un système informatisé de diagnostic en mathématiques au secondaire: une approche centrée sur l'analyse didactique. *Mesure et évaluation en éducation*, 16(3), 5–26.
- Dassa, C., Vázquez-Abad, J. et Ajar, D. (1993). Formative assessment in a classroom setting: From practice to computer innovations. *Alberta Journal of Educational Research*, 39, 111–125.

- Dumoulin, J. (1991). *Étude des pratiques de l'évaluation des apprentissages réalisées par les enseignants de mathématiques et de sciences au secondaire*. Mémoire de maîtrise inédit, Université de Montréal, Montréal, Québec.
- Gadbois, L., Burelle, R., Parent C. et Séguin, S. Ph. (1991). Un instrument de mesure des croyances et des attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves. *Mesure et évaluation en éducation*, 14(2), 5–24.
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart and Winston.
- Kulik, Ch.-L., Kulik, J. A. et Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60, 265–299.
- Ministère de l'Éducation du Québec [MEQ]. (1979). *L'École québécoise, énoncé de politique et plan d'action*. Québec: MEQ.
- Perrenoud, Ph. (1988). Formation à l'évaluation: entre réalisme conservateur et idéalisme béat. In M. Gather Thurler et Ph. Perrenoud (dirs.), *Savoir évaluer pour mieux enseigner. Quelle formation des maîtres?* (Cahier no 26) (p. 49–71). Genève: Service de la recherche sociologique.
- Perrenoud, Ph. (1991). Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation*, 13(4), 49–81.
- Rogers, W. T. (1990). *Educational measurement in Canada: Evolution or extinction?* Texte de la Communication de l'invité du Président (ACCÉ), 18e Congrès annuel de la Société canadienne pour l'étude de l'éducation, Victoria, Colombie-Britannique.
- Séguin, S. Ph., Parent, C. et Burelle, R. (1993). Croyances saillantes et conditions de praticabilité de l'évaluation formative des apprentissages telles que perçues par les enseignants. In R. Hivon (dir.), *L'évaluation des apprentissages: réflexions, nouvelles tendances et formation* (p. 137–159). Sherbrooke: Éditions du CRP.
- Slavin, R. E. (1987). Mastery Learning reconsidered. *Review of Educational Research*, 57, 175–213.
- Slavin, R. E. (1990). Mastery Learning re-reconsidered. *Review of Educational Research*, 60, 300–302.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 65, 363–368.
- Stiggins, R. J. et Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22, 271–286.
- Tyler, R. W. (1986). *Changing concepts of educational evaluation*. Toronto: Pergamon Press.
- Vázquez-Abad, J. et LaFleur, M. (1990). Design of a performance-responsive drill and practice algorithm for computer-based training. *Computers and Education*, 14(1), 43–52.
- Viau, R. et Stolovitch, H. D. (1987). *L'impact de l'ordinateur sur l'apprentissage: réalité ou illusion?* Montréal: APO-Québec.

---

Djavid Ajar et Clément Dassa sont professeurs à la Faculté des sciences de l'éducation, Université de Montréal, C.P. 6128, succursale A, Montréal (Québec), H3C 3J7.

