

L'évaluation pour améliorer la qualité de l'enseignement

Thomas Kellaghan et Vincent Greaney

Paris 2002

UNESCO : Institut international de planification de l'éducation

Dans cette collection* :

1. Qu'est-ce que la planification de l'éducation ? *P.H. Coombs*
2. Les plans de développement de l'éducation et la planification économique et sociale, *R. Poignant*
3. Planification de l'éducation et développement des ressources humaines, *F. Harbison*
4. L'administrateur de l'éducation face à la planification, *C.E. Beeby*
5. Le contexte social de la planification de l'éducation, *C.A. Anderson*
6. La planification de l'enseignement : évaluation des coûts, *J. Valzey, J.D. Chesswas*
7. Les problèmes de l'enseignement en milieu rural, *V.L. Griffiths*
8. Le rôle du conseiller en planification de l'enseignement, *A. Curle*
9. Les aspects démographiques de la planification de l'enseignement, *T.N. Châu*
10. Coûts et dépenses en éducation, *J. Hallak*
11. L'identité professionnelle du planificateur de l'éducation, *A. Curle*
12. Planification de l'éducation : les conditions de réussite, *G.C. Ruscoe*
13. L'analyse coût-bénéfice dans la planification de l'éducation, *M. Woodhall*
14. Planification de l'éducation et chômage des jeunes, *A. Callaway*
15. Planification de l'éducation pour une société pluraliste, *C. Hon-chan*
17. La planification des programmes d'enseignement primaire dans les pays en voie de développement, *H.W.R. Hawes*
18. Planification de l'aide à l'éducation pour la deuxième décennie du développement, *H.M. Phillips*
19. Les études à l'étranger et le développement de l'enseignement, *W.D. Carter*
20. Pour une conception réaliste de la planification de l'éducation, *K.R. McKinnon*
21. La planification de l'éducation en relation avec le développement rural, *G.M. Coverdale*
22. La planification de l'éducation : options et décisions, *J.D. Montgomery*
23. La planification du programme scolaire, *A. Lewy*
24. Les facteurs de coûts dans la planification des systèmes de technologies éducatives, *D.T. Jamison*
25. Le planificateur et l'éducation permanente, *P. Furter*
26. L'éducation et l'emploi : une étude critique, *M. Carnoy*
27. Planification de l'offre et de la demande d'enseignants, *P. Williams*
28. Planification de l'éducation préscolaire dans les pays en développement, *A. Heron*
29. Moyens de communication de masse et éducation dans les pays à faible revenu : répercussions sur la planification, *E.G. McNany, J.K. Mayo*
30. La planification de l'éducation non formelle, *D.R. Evans*
31. Education, formation et secteur traditionnel, *J. Hallak, F. Caillods*
32. Enseignement supérieur et emploi : l'expérience de l'IHPE dans cinq pays en développement, *G. Psacharopoulos, B.G. Sanyal*
33. La planification de l'éducation comme processus social, *T. Malan*
34. Enseignement supérieur et stratification sociale : une comparaison internationale, *T. Husén*
35. Un cadre conceptuel pour le développement de l'éducation permanente en URSS, *A. Vladislavlev*
36. Education et austerité : quelles options pour le planificateur ? *K.M. Levin*
37. La planification de l'éducation en Asie, *R. Roy-Singh*
38. Les projets d'éducation : préparation, financement et gestion, *A. Magnen*
39. Accroître l'efficacité des enseignants, *L. Anderson*
40. L'élaboration des programmes scolaires à l'échelon central et à l'échelon des écoles, *A. Lewy*
41. Planification des ressources humaines : méthodes, expériences, pratiques, *O. Bertrand*
42. Redéfinition de l'éducation de base en Amérique latine : les enseignements de l'Ecole Nouvelle colombienne, *E. Schiefelbein*
43. La gestion des systèmes d'enseignement à distance, *G. Rumble*
44. Stratégies éducatives pour les petits Etats insulaires, *D. Atchoarena*
45. Evaluation de la recherche en éducation fondée sur l'expérimentation et sur les enquêtes, *R.M. Wolf*
46. Droit et planification de l'éducation, *J. Birch*
47. Utilisation de l'analyse sectorielle de l'éducation et des ressources humaines, *F. Kemmerer*
48. Analyse du coût de l'insertion scolaire des populations marginalisées, *M.C. Tsang*
49. Un système d'information pour la gestion fondé sur l'efficacité, *W.W. McMahon*
50. Examens nationaux : conception, procédures et diffusion des résultats, *J.P. Keeves*
51. Le processus de planification et de formulation des politiques d'éducation : théorie et pratiques, *W.D. Haddad, assisté par T. Demsky*
52. A la recherche d'un enseignement adapté : l'orientation vers le travail dans l'éducation, *W. Hoppers*
53. Planifier pour l'innovation en matière d'éducation, *D.E. Inbar*
54. Analyse fonctionnelle de l'organisation des ministères d'éducation, *R. Sack, M. Saïdi*
55. Réduire les redoublements : problèmes et stratégies, *T. Eisonon*
56. Faire davantage participer les filles et les femmes à l'éducation, *N. P. Stromquist*
57. Installations et bâtiments éducatifs : ce que les planificateurs doivent savoir, *J. Beynon*
58. La planification de programmes d'alphabétisation des adultes centrés sur les élèves, *S.E. Malone, R.F. Arnove*
59. Former les enseignants à travailler dans des établissements et/ou des classes réputés difficiles, *J.-L. Auduc*
60. L'évaluation de l'enseignement supérieur, *J.L. Rontopoulou*
61. A l'ombre du système éducatif. Le développement des cours particuliers : conséquences pour la planification de l'éducation, *M. Bray*
62. Une gestion plus autonome des écoles, *I. Abu-Duhou*
63. Mondialisation et réforme de l'éducation : ce que les planificateurs doivent savoir, *M. Carnoy*
64. La décentralisation dans l'éducation : pourquoi, quand, quoi et comment? *T. Welsh, N.F. McGinn*
65. L'éducation préscolaire : besoins et possibilités, *D. Weikart*
66. La planification de l'éducation dans le contexte du VIH/sida, *M.J. Kelly*
67. Aspects légaux de la planification et de l'administration de l'éducation, *C. Durand-Prinborgne*
68. Améliorer l'efficacité de l'école, *J. Scheerens*
69. La recherche quantitative au service des politiques éducatives : le rôle de l'analyse de la littérature, *S.J.Hite*
70. La cyberformation dans l'enseignement supérieur : développement de stratégies nationales, *Tony Bates*

* Série publiée également en anglais. Autres titres à paraître.

L'Agence suédoise d'aide au développement international (Asdi) a fourni une aide financière pour la publication de cette brochure.

Publié en 2002 par l'Organisation des Nations Unies
pour l'éducation, la science et la culture
7, place de Fontenoy, 75007 Paris

Maquette de couverture : Pierre Finot
Composition : Linéale Production
Imprimé en France par l'Imprimerie Alençonnaise

ISBN 92-803-2215-X
© UNESCO 2002

Principes de la planification de l'éducation

Les ouvrages de cette collection sont destinées principalement à deux catégories de lecteurs : ceux qui occupent déjà des fonctions dans l'administration et la planification de l'éducation, dans les pays en développement comme dans les pays industrialisés ; et d'autres, moins spécialisés – hauts fonctionnaires et hommes politiques, par exemple – qui cherchent à connaître de façon plus générale le mécanisme de la planification de l'éducation et les liens qui la rattachent au développement national dans son ensemble. Ces ouvrages sont, de ce fait, destinées soit à l'étude individuelle, soit à des cours de formation.

Depuis le lancement de cette collection en 1967, les pratiques et les concepts de la planification de l'éducation ont subi d'importants changements. Plusieurs des hypothèses qui étaient sous-jacentes aux tentatives antérieures de rationaliser le processus du développement de l'éducation ont été critiquées ou abandonnées. Toutefois, si la planification centralisée, rigide et obligatoire, s'est manifestement révélée inadéquate, toutes les formes de planification n'ont pas été abandonnées. La nécessité de rassembler des données, d'évaluer l'efficacité des programmes en vigueur, d'entreprendre des études sectorielles et thématiques, d'explorer l'avenir et de favoriser un large débat sur ces bases s'avère au contraire plus vive que jamais pour orienter la prise de décision et l'élaboration des politiques éducatives. Pour faire des choix de politique générale judicieux, il est capital de dresser un bilan précis de la situation, de déterminer les objectifs visés, de coordonner les moyens propres à les atteindre et de contrôler les résultats obtenus. La planification est donc aussi un moyen d'organiser l'acquisition des connaissances : par la définition de plans, de cibles, d'actions et de mesures correctives.

La planification de l'éducation a pris une envergure nouvelle. Outre les formes institutionnelles de l'éducation, elle porte à présent sur toutes les autres prestations éducatives importantes, dispensées

hors de l'école. L'intérêt consacré à l'expansion et au développement des systèmes éducatifs est complété, voire parfois remplacé, par le souci croissant d'améliorer la qualité du processus éducatif dans son ensemble et d'évaluer les résultats obtenus. Enfin, planificateurs et administrateurs sont de plus en plus conscients de l'importance des stratégies de mise en œuvre et du rôle joué à cet égard par les divers mécanismes de régulation : choix des méthodes de financement, d'examen et de délivrance des certificats et diplômes, ou d'autres structures de régulation et d'incitation. La démarche des planificateurs répond à une double préoccupation : mieux comprendre la valeur et le rôle de l'éducation par l'observation empirique des dimensions particulières qui sont les siennes, et contribuer à définir des stratégies propres à amener le changement.

Ces ouvrages ont pour objet de refléter l'évolution et les changements des politiques éducatives et de mesurer leurs effets sur la planification de l'éducation ; de mettre en lumière les questions qui se posent actuellement en la matière et de les analyser dans leur contexte historique et social ; et de diffuser des méthodes de planification pouvant s'appliquer aussi bien aux pays en développement qu'aux pays industrialisés.

Pour les décideurs et les planificateurs, l'expérience d'autrui est extrêmement riche d'enseignements : les problèmes auxquels d'autres sont confrontés, les objectifs qu'ils recherchent, les méthodes qu'ils expérimentent, les résultats auxquels ils parviennent et les résultats involontaires qu'ils obtiennent méritent d'être analysés.

Afin d'aider l'Institut à bien identifier les préoccupations actuelles dans les domaines de la planification et de l'élaboration des politiques de l'éducation dans diverses parties du monde, un Comité de rédaction a été mis en place. Il comprend deux rédacteurs en chef et des rédacteurs associés, venus de différentes régions, tous éminents spécialistes dans leurs domaines respectifs. Lors de la première réunion de ce nouveau Comité de rédaction en janvier 1990, ses membres ont défini les sujets les plus importants à traiter dans les numéros ultérieurs sous les rubriques suivantes :

1. L'éducation et le développement.
2. L'équité.
3. La qualité de l'éducation.
4. Structure, administration et gestion de l'éducation.
5. Les programmes d'enseignement.
6. Coût et financement de l'éducation.
7. Techniques et approches de la planification.
8. Systèmes d'information, suivi et évaluation.

Chaque rubrique est confiée à un ou deux rédacteurs.

La collection correspond à un plan d'ensemble soigneusement établi, mais aucune tentative n'a été faite pour éliminer les divergences, voire les contradictions, entre les points de vue exposés par les auteurs. L'Institut, pour sa part, ne souhaite imposer aucune doctrine officielle. S'il reste entendu que les auteurs sont responsables des opinions qu'ils expriment – et qui ne sont pas nécessairement partagées par l'UNESCO et l'IIPE – elles n'en sont pas moins dignes de faire l'objet d'un vaste débat d'idées. Cette collection s'est d'ailleurs fixé comme objectif de refléter la diversité des expériences et des opinions en donnant à des auteurs venus d'horizons et de disciplines très variés la possibilité d'exprimer leurs idées sur l'évolution des aspects théoriques et pratiques de la planification de l'éducation.

L'évaluation est devenue, au cours des années, une clé importante de l'amélioration de la qualité de l'enseignement. Elle est l'un des moyens les plus fiables d'identifier les problèmes, qu'ils se situent au niveau du système ou de l'école ou qu'ils concernent tel ou tel élève.

Cependant, l'évaluation soulève un certain nombre de questions. Les opinions ne convergent pas toujours sur ce qu'il convient d'évaluer, ni sur la manière de procéder. De plus, il est souvent difficile, pour des raisons diverses, d'apprécier l'évolution de la qualité dans le temps.

Une autre question cruciale est de savoir que faire des résultats d'une évaluation. De nombreux débats sont en cours sur la nécessité et la manière d'utiliser ceux-ci à des fins de responsabilisation.

Thomas Kellaghan et Vincent Greaney abordent ces questions, ainsi que beaucoup d'autres, et montrent qu'il n'est jamais facile d'y répondre. Les décideurs trouveront en cet ouvrage stimulant un guide utile pour la définition et la mise en œuvre de stratégies d'évaluation.

Gudmund Hernes
Directeur, IIEP

Composition du Comité de rédaction

- Président :* Gudmund Hernes
Directeur, IPE
- Rédacteurs en chef :* Françoise Caillods
IPE
- T. Neville Postlethwaite
(Professeur émérite)
Université de Hambourg
Allemagne
- Rédacteurs associés :* Jean-Claude Eicher
Université de Bourgogne
France
- Claudio de Moura Castro
Consultant international
Brésil
- Kenneth N. Ross
IPE
France
- Richard Sack
Consultant international
France
- Rosa Maria Torres
Consultante internationale
Argentine

Préface

Il y a maintenant de nombreuses années que les planificateurs de l'éducation se préoccupent des moyens d'y introduire des changements substantiels et d'en améliorer sa qualité. La focalisation de ce souci de qualité est progressivement passée des intrants aux résultats en termes d'acquis scolaires.

Les réformes de l'éducation de naguère, englobant les structures, les programmes d'études et la formation des enseignants, ont dans une large mesure cédé la place à des réformes plus modestes, orientées vers l'amélioration des acquis scolaires, de l'efficacité de l'école, de la gestion et de la responsabilité. La décentralisation, la gestion au niveau de l'école et les acquis scolaires sont devenus des concepts clés dans les réformes de l'éducation des années 1990. La concurrence économique mondiale a mis en évidence l'importance critique de la qualité des ressources humaines et la demande de compétences nouvelles dans la société de l'information d'aujourd'hui. Le système éducatif, les écoles et chacun des élèves sont tous soumis à des pressions croissantes en faveur des résultats.

L'évaluation, et notamment l'évaluation de l'acquis scolaire des élèves, est devenue l'objet de beaucoup d'attention et d'activités à travers le monde, que ce soit dans les pays industrialisés ou dans les pays en développement.

Les études internationales comparant les résultats scolaires obtenus en lecture et écriture, en mathématiques et en sciences dans différents pays sont en plein essor. La publication des résultats de ces études est dans la plupart des pays participants un événement majeur, souvent au détriment d'une analyse plus approfondie des différences de résultats entre écoles et entre groupes, et des facteurs qui expliquent ces différences. La plupart des pays entreprennent également des évaluations régulières de leurs élèves. Ces études ne sont pas seulement destinées à aider les autorités nationales à contrôler le système : elles sont considérées comme un levier du changement en soi, susceptible de provoquer des modifications dans les écoles et dans les salles de classe.

Préface

Étant donné l'importance qui est aujourd'hui attachée à l'évaluation, un problème se pose souvent aux décideurs en matière d'éducation : quel type d'évaluation faut-il introduire, et à quelles fins ? Le présent ouvrage, rédigé par Thomas Kellaghan, directeur du Centre de recherche sur l'éducation du St Patrick's College, à Dublin, et Vincent Greaney, spécialiste principal de l'éducation à la Banque mondiale, répondra à nombre de leurs questions. Les auteurs clarifient des concepts de base tels que l'évaluation, la qualité, les normes et la responsabilité, mais ils présentent aussi les caractéristiques de l'évaluation qui ressortent de plusieurs propositions de réforme, examinent différents modèles d'évaluation nationale et évoquent des points cruciaux de la conception d'une évaluation.

Tout en reconnaissant que les évaluations nationales ne sont pas la seule manière de contrôler les performances et de renforcer la qualité de l'éducation, les auteurs présentent et examinent de façon extrêmement claire et concise les facteurs les plus importants à prendre en compte lorsqu'on envisage le recours à l'évaluation dans un contexte national spécifique. Le Comité de rédaction est très reconnaissant à Thomas Kellaghan et à Vincent Greaney pour cette contribution actuelle et stimulante.

Françoise Caillods
Corédacteur en chef

Remerciements

Cet ouvrage est issu d'un document présenté par le premier cité de ses auteurs lors d'une réunion du Groupe international de travail sur l'éducation tenue à Florence du 14 au 16 juin 2000. Ce document a été réalisé sur la suggestion de Liz Higgins, de Ireland Aid ; sa préparation et la participation de l'auteur à la réunion ont été financées par Ireland Aid.

Les auteurs tiennent à exprimer leur reconnaissance à Mme Françoise Caillods, de l'Institut international de planification de l'éducation, pour ses conseils concernant la préparation de cet ouvrage et pour son appui et sa patience durant une gestation plus longue que prévu ; au Professeur T. Neville Postlethwaite pour son apport et son orientation à un premier stade de ce travail ; à leurs collègues de la Banque mondiale, et notamment à Luis Benveniste, pour leur aide ; et à Hilary Walsh et Regina Klöditz, qui ont saisi le manuscrit.

Table des matières

Préface	11
Remerciements	13
I. Introduction	17
II. Évaluation, qualité, normes et responsabilité	21
L'évaluation dans l'enseignement	21
La qualité dans l'enseignement	25
Les normes	26
La responsabilité	29
Conclusion	31
III. Caractéristiques de l'évaluation dans les propositions de réformes actuelles	33
La focalisation sur les résultats	33
La focalisation sur l'évaluation externe	35
L'évaluation comme levier de la réforme	36
Conclusion	37
IV. L'évaluation nationale des acquis scolaires	39
Qu'est-ce qu'une évaluation nationale ?	39
Modèles d'évaluation nationale	42
Pourquoi effectuer une évaluation nationale ?	46
Les besoins d'information auxquels répond une évaluation nationale	48
L'extension de l'activité d'évaluation nationale	49
Conclusion	51
V. Décrire et utiliser l'information issue d'une évaluation nationale	55
Décrire les résultats scolaires	55
La communication de l'information issue d'une évaluation nationale	64
L'utilisation des données de l'évaluation nationale	68
Conclusion	71
	15

Table des matières

VI.	Les évaluations internationales des acquis scolaires	73
	Qu'est-ce qu'une évaluation internationale ?	73
	Pourquoi participer à une évaluation internationale ?	75
	Les problèmes des évaluations internationales	76
	L'utilisation des données de l'évaluation internationale	79
	Conclusion	81
VII.	Utiliser les données des résultats scolaires pour évaluer les écoles	83
	Évaluer la qualité des écoles	83
	Actions consécutives à l'évaluation des écoles	84
	Les classements des établissements scolaires	87
	L'impact des données de l'évaluation nationale sur les écoles	92
	Conclusion	94
VIII.	Les problèmes de la conception d'une évaluation	97
	Les considérations qui interviennent dans la conception d'une évaluation	99
	La participation à une évaluation internationale	103
	Les coûts	103
	L'utilisation de l'information issue d'une évaluation	104
	Conclusion	106
	Bibliographie	109

I. Introduction

Le besoin d'améliorer la qualité de l'enseignement est fréquemment formulé dans les écrits contemporains, dans les débats politiques et dans les médias. C'est un besoin qui a tout autant de chances de s'exprimer dans un pays du monde industrialisé possédant un système éducatif développé que dans un pays en développement n'ayant pas encore atteint le but de l'enseignement primaire pour tous. Bien entendu, les préoccupations relatives à la qualité de l'enseignement ne sont pas nouvelles, et différentes réformes ont été proposées, au cours des années, pour leur répondre, qui s'étendent des changements dans l'organisation des écoles et des programmes d'études jusqu'à l'amélioration des structures de gestion des écoles, la formation pédagogique, le tutorat entre pairs et une plus grande implication des parents. Les réformes se poursuivent aujourd'hui dans ces secteurs ; cependant, celle qui a dominé la scène de l'éducation dans les années 1980 et 1990, et jusque dans les premières années 2000, propose le recours à l'évaluation pour améliorer la qualité de l'enseignement.

Lorsqu'ils envisagent cette réforme, ainsi que les propositions portant sur leur mise en œuvre, les décideurs, les planificateurs et les gestionnaires de l'éducation doivent assumer deux décisions majeures. La première porte sur l'alternative suivante : faut-il investir des ressources dans les activités d'évaluation (que ce soit pour améliorer celles qui existent ou pour en introduire de nouvelles), ou pourrait-on obtenir de meilleurs résultats si ces ressources étaient investies dans quelque autre aspect du système éducatif, comme les bâtiments scolaires, les manuels, la formation pédagogique ou le nombre des enseignants ? Étant donné l'importance attachée actuellement, dans le monde entier, à l'évaluation, il est peu probable que ce secteur soit écarté comme ne méritant ni approfondissement ni investissement, auquel cas une autre question se pose. Quel type d'évaluation et quelle utilisation de l'information issue de l'évaluation ont-ils des chances d'exercer un impact sur la qualité de l'enseignement dans les écoles ? Il n'existe évidemment pas de réponses simples à ces questions. On ne parviendra à établir des réponses que lorsque tous les facteurs pertinents (incluant les besoins d'information, la conception des valeurs,

les considérations financières, politiques et logistiques et les traditions en matière d'évaluation) auront été pris en compte. L'analyse conceptuelle des questions pertinentes et la disponibilité d'informations portant sur la pratique d'autres systèmes éducatifs peuvent également aider à évaluer les options et à anticiper les effets des actions envisagées.

Le but de cet ouvrage, qui s'inspire d'une expérience acquise à travers le monde, consiste à fournir aux planificateurs, aux administrateurs et aux gestionnaires de l'éducation cette analyse et ces informations. Il décrit les options en matière d'évaluation, ainsi que leurs avantages et leurs inconvénients, tels qu'ils ressortent de la recherche et d'autres types d'informations. De plus, l'évaluation n'étant pas envisagée seulement pour décrire les résultats scolaires, mais aussi comme un levier de changement des systèmes éducatifs conduisant à une amélioration de la qualité et à une élévation du niveau des acquis, on prendra en considération les incidences de l'évaluation sur les programmes scolaires, l'enseignement et l'apprentissage.

Cet ouvrage est axé sur les évaluations nationale et internationale des élèves, dans la mesure où celles-ci (et notamment cette dernière) représentent des sources d'information importantes aux yeux des décideurs et des gestionnaires de l'éducation. Focaliser l'attention sur ces secteurs n'est pas dévaloriser les autres secteurs de l'évaluation. Par exemple, l'évaluation par leurs propres enseignants des connaissances acquises en classe par les élèves mérite une attention toute particulière dans un contexte d'amélioration de la qualité de l'enseignement. Cependant, l'évaluation effectuée en classe ne bénéficie pas d'une place importante dans les projets de réforme et ne sera donc pas envisagée dans cet ouvrage comme un thème distinct. Un autre type d'évaluation, qui ne sera pas non plus étudié, est l'examen extérieur ou public, bien que ses incidences sur la pratique scolaire soient largement documentées et que certains commentateurs aient même soutenu que des examens bien conçus ont aussi un rôle important à jouer dans l'amélioration des acquis scolaires.

Après l'introduction, des définitions des concepts clés des projets de réforme actuels (évaluation, qualité, normes et responsabilité) seront présentées dans le *Chapitre II*. Dans le *Chapitre III*, on esquissera les

caractéristiques du type d'évaluation qui peut être envisagé pour améliorer la qualité de l'enseignement. Dans le *Chapitre IV* seront décrites les évaluations nationales (et par État) des acquis scolaires. Bien que relativement récentes sur la scène de l'éducation, ces évaluations ont connu un développement remarquable dans les années 1990. Elles sont le centre d'un grand nombre de discussions et d'activités et constituent le principal type d'évaluation qui soit associé à la qualité et aux normes dans le débat actuel. Les moyens divers qui ont été utilisés pour décrire les acquis scolaires, ainsi que la manière dont les informations issues des évaluations nationales ont été utilisées, seront décrits dans le *Chapitre V*. Les évaluations internationales des acquis scolaires, qui feront l'objet du *Chapitre VI*, ont aussi connu une extension au cours des années 1990. Elles utilisent des procédures qui sont très semblables à celles auxquelles on a recours dans les évaluations nationales. Cependant, comme elles doivent répondre à des besoins communs reconnus entre pays participants, des compromis seront nécessaires dans leur exécution (en ce qui concerne, par exemple, la conception des instruments d'évaluation) et les évaluations s'harmoniseront moins bien avec les besoins spécifiques de chaque pays que les évaluations nationales. Cependant, elles autorisent les comparaisons entre les performances des élèves dans des systèmes éducatifs différents et ont donc apporté une dimension internationale à la notion de normes, suscitant ainsi un vif intérêt de la part des médias et des milieux politiques.

Bien que l'évaluation des établissements scolaires soit traditionnellement effectuée, dans de nombreux pays, par des inspecteurs ou des superviseurs scolaires, un certain nombre de pays ont commencé, dans les années 1990, à utiliser à cette fin les performances réalisées par des élèves dans les tests de niveau des évaluations nationales ou par État. Cette utilisation, ainsi que les problèmes connexes, fera l'objet du *Chapitre VII*. Le dernier chapitre décrira les décisions qu'implique la conception d'une évaluation et la manière dont pourraient être utilisées les informations provenant d'une évaluation.

II. Évaluation, qualité, normes et responsabilité

Pour situer dans leur contexte les propositions visant à utiliser l'évaluation pour améliorer la qualité de l'enseignement, un certain nombre de concepts clés seront décrits dans ce chapitre. Les significations, les formes et les buts de l'évaluation y seront soulignés, après quoi trois concepts associés seront examinés : la qualité, les normes et la responsabilité.

L'évaluation dans l'enseignement

Définition de l'évaluation

Le terme d'« évaluation » peut être utilisé, en matière d'enseignement, par référence à toute procédure ou activité conçue pour collecter des informations sur les connaissances, les attitudes ou les compétences d'un apprenant ou d'un groupe d'apprenants. Selon une définition plus détaillée, l'évaluation est :

le processus consistant à obtenir des informations qui seront utilisées pour prendre des décisions sur les élèves en matière d'enseignement, pour donner à l'élève des informations en retour sur ses progrès, ses points forts et ses faiblesses, pour juger de l'efficacité de l'enseignement et de l'adéquation des programmes d'études, et pour informer les politiques générales (AFT, NCME, NEA, 1990, p. 1).

Dans le type d'évaluation que reconnaîtront tous ceux qui ont fréquenté une école, un enseignant ou un examinateur est impliqué. Une évaluation pourrait aussi être effectuée par l'apprenant lui-même ou par une autre partie. Quelles que soient les circonstances, l'évaluation, telle qu'elle ressort de la définition ci-dessus, a traditionnellement été associée à l'appréciation portée sur des individus. Cependant, le terme est utilisé de plus en plus fréquemment depuis les années 1980 pour décrire les performances des écoles ou des systèmes scolaires.

Les formes de l'évaluation

Il existe de multiples manières d'effectuer une évaluation. Elle peut être orale, auditive, écrite, pratique ou fondée sur l'observation. Elle peut être effectuée dans le cadre d'un tête-à-tête, d'un petit groupe ou d'un groupe important, ou sur ordinateur. Si l'on adopte une conception holistique des acquis scolaires, l'évaluation sera fondée sur l'ensemble des performances des élèves dans l'exécution d'une tâche complexe (par exemple construire un objet en bois, défendre un point de vue dans une rédaction argumentée, préparer un repas). Par contre, si l'on adopte un point de vue analytique, dans lequel les acquis scolaires seront décomposés en une série de segments de dimension réduite, l'évaluation comportera une diversité de brèves réponses à des questions. Le degré de formalisation peut également varier, allant de celui dans lequel l'individu répond à une série de questions ou de suggestions dans des conditions contrôlées et dans un laps de temps limité à une situation dans laquelle le comportement de l'individu est observé sans intrusion sur une période prolongée.

Les buts de l'évaluation

L'information issue de l'évaluation est utilisée à des fins diverses. Au niveau individuel des élèves, elle est utilisée (a) pour décrire leurs acquis, pour discerner et diagnostiquer les problèmes d'apprentissage et pour prévoir l'enseignement ou l'apprentissage à venir ; (b) pour apporter aux élèves une orientation dans le choix des cours qu'ils suivront ultérieurement ou dans leurs décisions relatives à leurs options professionnelles ; (c) pour motiver les élèves en leur proposant des buts ou des cibles, en clarifiant la nature des tâches d'apprentissage et en faisant savoir aux élèves et à leurs enseignants comment ils progressent ; (d) pour certifier que les intéressés ont atteint un certain niveau de compétence ; et (e) pour effectuer une sélection individuelle en vue de l'accès au niveau suivant du système éducatif ou à un emploi.

Les buts principaux de l'évaluation d'une institution ou d'un système sont de former un jugement sur l'efficacité d'une école, et de former un jugement sur l'adéquation des performances d'un système éducatif ou d'une partie de celui-ci (voir *Encadré 2.1*). Même dans

ces cas-là, l'évaluation est basée sur les performances individuelles des élèves, qui sont agrégées au niveau de l'institution ou du système, généralement pour fournir des informations aux décideurs et à d'autres personnes.

Encadré 2.1 L'utilisation des informations générées par l'évaluation pour évaluer les institutions et les systèmes scolaires

Évaluation des institutions scolaires. Les résultats d'une évaluation sont parfois utilisés pour apprécier l'efficacité d'un établissement scolaire. La « réussite » en matière de résultats peut être mise en équation avec le pourcentage d'élèves qui atteignent un niveau de « compétence » donné, ou avec le pourcentage de ceux qui passent au niveau immédiatement supérieur du système éducatif. Dans certains pays, ce processus est rendu public par la diffusion des résultats sous la forme de classements par établissement, ou par des écoles qui annoncent dans la presse les résultats obtenus par leurs élèves.

Évaluation des systèmes scolaires. L'information générée par une évaluation peut être utilisée pour former un jugement sur l'adéquation des résultats d'un système éducatif ou d'une partie de celui-ci. Les normes impliquées par cette appréciation peuvent être prédéterminées (exemple : 80 % des élèves atteignent un niveau de compétence spécifié) ou comparatives (exemples : résultats obtenus à une date antérieure, ou encore résultats obtenus par les élèves dans un système similaire).

Si l'on peut établir une distinction entre les buts des évaluations individuelles, institutionnelles et systémiques, il arrive dans la pratique que cette distinction soit estompée. Certaines évaluations sont conçues avec plusieurs objectifs et sont utilisées pour fournir des informations non seulement sur les systèmes éducatifs, mais aussi sur les écoles, et même pour remplir des fonctions qui étaient traditionnellement associées à l'évaluation au niveau de l'élève (en relation, par exemple, avec le passage en classe supérieure, la motivation ou l'obtention d'un diplôme).

Considérations qui interviennent dans l'utilisation des informations générées par l'évaluation

Il est facile, lorsqu'on décrit les résultats d'une évaluation, de donner l'impression que la définition et l'appréciation des acquis humains n'ont rien de problématique. En particulier, les comptes rendus médiatiques semblent supposer que les évaluations fournissent une information qui est aussi dénuée d'ambiguïté, objective et fiable que l'information portant sur des caractéristiques physiques comme la taille ou le poids.

Il importe que les personnes qui participent à la prise de décisions sur la base des informations générées par l'évaluation n'adoptent pas cette conception et soient sensibles aux difficultés de définition et d'appréciation des acquis scolaires. L'une de ces difficultés est qu'une tâche d'évaluation ne peut inclure qu'un échantillon des connaissances et compétences appropriées et un échantillon de réponses des élèves. Dans une telle situation, on est fondé à se demander dans quelle mesure les performances réalisées pour une tâche donnée – surtout si elles se limitent à des réponses écrites dans des conditions contrôlées – peuvent raisonnablement être généralisées à d'autres tâches et à d'autres contextes. De plus, si l'on considère le type de résultats (connaissances et compétences) dont les réformateurs affirment l'importance dans la société contemporaine eu égard à la compétitivité économique qu'ils assurent, on peut se demander comment devront être évalués, par exemple, les capacités d'investigation ou l'engagement dans un apprentissage. Pour prendre un exemple concret dans le domaine des mathématiques, comment évaluer l'aptitude à construire et à appliquer des principes (au lieu de faire simplement une démonstration des procédures de calcul), ou la capacité d'utiliser des méthodes diverses pour résoudre des problèmes qui ne relèvent pas de la simple routine ?

Outre les problèmes techniques que rencontre l'appréciation de résultats, il faut aussi reconnaître que l'évaluation est une activité sociopolitique. Il en va nécessairement ainsi lorsque les procédures d'évaluation sont définies et prescrites dans le cadre de décisions de politique générale. De plus, l'information dérivée d'une évaluation peut avoir des conséquences personnelles et sociales d'une portée

considérable, intentionnelles et involontaires, positives et négatives, notamment quand elles sont utilisées pour prendre des décisions sur l'orientation des élèves ou pour rendre des individus ou des institutions responsables de leurs actes.

La qualité dans l'enseignement

Dans une grande partie des discours contemporains sur l'enseignement, le terme de « qualité » apparaît fréquemment, bien qu'il soit rarement défini. En l'absence d'une définition formelle généralement reconnue, on peut inférer de l'usage habituel que le terme est lié à l'adéquation ou à la pertinence d'objets ou de processus par rapport aux buts auxquels ils étaient destinés. De plus, la qualité implique une échelle, et souvent la référence à des normes. Un objet peut être de « bonne » ou de « médiocre » qualité, ou bien satisfaire ou ne pas satisfaire à telle ou telle norme.

Lorsqu'on choisit l'aspect de l'éducation sur lequel se concentrera l'attention, un problème de définition de la qualité dans l'enseignement se pose. L'enseignement ayant des buts et des composantes multiples, on peut raisonnablement poser des questions en rapport avec la qualité sur tout aspect important d'un système : infrastructure, bâtiments scolaires, administration, formation pédagogique, matériels didactiques, activité enseignante ou acquis scolaires. On notera que tous ces éléments sont corrélés, et qu'une carence sérieuse de l'un d'entre eux aura sans doute des implications en matière de qualité pour les autres.

Si l'on se concentre sur les résultats de l'enseignement (comme c'est le cas dans les propositions de réforme actuelles), un autre problème surgit, puisqu'il n'existe pas de consensus sur ce que devraient être les buts de la scolarité. Les buts ou les objectifs sont affaire de choix et doivent donc être fondés sur des jugements de valeur. Pour certains, le rôle de l'enseignement est de favoriser le développement cognitif, moral et social des élèves ; pour d'autres, l'enseignement est un moyen de promouvoir la cohésion sociale et le renforcement de la nation ; pour d'autres encore, une préparation au monde du travail. Peut-être est-ce à cause de cette divergence des points de vue que beaucoup d'évaluations du système éducatif – mais non leur totalité –

se concentrent sur des connaissances et des compétences dont l'importance est universellement reconnue. Les élèves savent-ils lire ? Savent-ils compter ? Comprennent-ils les concepts scientifiques élémentaires ? Cette focalisation ne doit cependant pas conduire à négliger les compétences cognitives d'un ordre supérieur que de nombreuses réformes aspirent à promouvoir, ni de nombreuses autres acquisitions précieuses de la scolarité, qui ont trait aux attitudes, aux valeurs, aux aspirations, à la motivation, à la conception de soi et à la socialisation.

Les normes

L'idée de normes est étroitement liée à l'idée de qualité, et elle a souvent joué un rôle dans les débats sur l'enseignement depuis les années 1980. En fait, les préoccupations en matière de normes ont été au cœur du débat politique aux États-Unis qui a débouché sur des propositions de réformes impliquant l'évaluation. Les critiques du système éducatif ont conclu, souvent en s'appuyant sur des informations insuffisantes données par des évaluations nationales et internationales, que les élèves ne recevaient pas, et ne pouvaient donc pas acquérir, les connaissances et les compétences dont ils auraient besoin pour affronter la concurrence économique et pour soutenir une démocratie vivante.

Quand on considère la signification du terme « norme », un certain nombre de définitions du dictionnaire semblent appropriées à l'enseignement : degré d'excellence requis pour atteindre un but particulier ; indicateur de ce qui est adéquat ; et niveau de performance décrit sur le plan social ou pratique. Parmi ces trois définitions, on n'a guère prêté d'attention à la première pour définir des normes éducatives. La plupart des initiatives ont été d'une ampleur plus limitée et se sont concentrées sur les connaissances et les compétences dont l'acquisition par les élèves est considérée comme appropriée et adéquate à des niveaux particuliers du système éducatif. À la suite de l'introduction en Grande-Bretagne d'un code des écoles élémentaires révisé en 1862, une signification d'ordre distinctement éducatif associée à l'évaluation a été attribuée au terme « norme » quand il a été défini dans l'*Oxford English Dictionary* comme « chacun des degrés de

compétence reconnus d'après les épreuves d'un examen, en fonction duquel est établi un classement des écoliers ». Cette acception est devenue obsolète quand les examens organisés au niveau central ont été abolis, bien que le terme ait continué d'être utilisé pendant un certain temps par référence à un degré de classement dans les écoles primaires.

Aujourd'hui, la question des normes dans l'enseignement revêt autant d'importance qu'au XIX^e siècle. Le Royaume-Uni a adopté en 1998 une loi qui détermine les normes et le cadre de la scolarité, et il existe au sein du Département de l'éducation et de l'emploi une Unité des normes et de l'efficacité. Aux États-Unis, les critiques adressées à la qualité de l'enseignement ont conduit à des propositions d'adoption de normes applicables au contenu ou au cadre du programme d'études approprié aux différentes classes (*normes de contenu*) et aux résultats attendus des élèves pour chaque classe (*normes de performance*). Sous-jacente à ces propositions était la perception du besoin d'une plus grande uniformité dans ce qui était enseigné dans les écoles à travers un pays qui ne possédait pas de programme d'études national et dans lequel l'enseignement était contrôlé par le gouvernement de chaque État, et non par les autorités fédérales. De nombreux pays n'auront pas à aborder ce problème, car ils possèdent de longue date des programmes d'études nationaux, renforcés dans bien des cas par un système d'examens extérieurs.

Cependant, d'autres aspects des propositions américaines méritent l'attention, même s'ils semblent souvent être méconnus lorsque des réformes fondées sur l'utilisation de l'évaluation sont introduites dans d'autres pays. Premièrement, les normes spécifient des définitions stimulantes des compétences, qui ont été décrites de manière variable comme comprenant des capacités de réflexion d'un ordre supérieur, des aptitudes à la résolution de problèmes, des capacités d'investigation et de raisonnement, une amélioration des moyens de communiquer et un engagement dans l'apprentissage permanent. Deuxièmement, les normes sont destinées à *tous* les élèves. Certains groupes d'élèves ne seront pas écartés, et on ne leur proposera pas des programmes d'études moins stimulants. Troisièmement, les normes ne constituent qu'un élément des réformes, et leur adoption n'aura sans doute que peu

d'impact si tous les grands mécanismes d'orientation de l'enseignement au sein du système éducatif ne sont pas alignés sur celles-ci : évaluations des élèves, matériels didactiques, préparation puis perfectionnement professionnel des enseignants, manuels, développement et amélioration des capacités scolaires, services d'appui aux écoles et dispositifs ciblés sur l'équité. Il est reconnu, en particulier, que les élèves n'acquerront pas les compétences requises sans l'intervention de changements dans les programmes d'études et dans les salles de classe. Des approches nouvelles seront nécessaires, qui impliqueront davantage d'engagement de la part des élèves ; un apprentissage qui soit heuristique et fondé sur l'expérience de situations concrètes, et non axiomatique et abstrait ; et la possibilité donnée aux élèves d'élaborer des stratégies de résolution des problèmes, de réfléchir de manière efficace et d'apprendre par eux-mêmes. Le rôle particulier des normes consistera à rendre plus concrets pour les enseignants les buts d'un système éducatif et à les aider, ainsi que d'autres, à se concentrer sur ce que les élèves, à des niveaux d'acquis variables, doivent savoir et être capables de faire.

Fixer des normes applicables aux performances scolaires implique de déterminer, sur une échelle de notation, un ou plusieurs seuils définissant des niveaux ou des classes correspondant à la gamme de notations qui interviendra. Cependant, ce processus ne va pas sans problèmes. Premièrement, les résultats d'un groupe d'élèves ne se répartissent pas en catégories nettement distinctes : ils forment plutôt un continuum dans lequel il existe très peu de différences entre les notes obtenues par chacun des élèves. Il n'existe pas de base évidente pour décider du point auquel on peut dire qu'une norme est atteinte, ni pour affirmer que les élèves obtenant une note immédiatement supérieure à un seuil donné diffèrent de manière substantielle de ceux qui obtiennent des résultats situés immédiatement en dessous de ce seuil. Deuxièmement, une note ou un classement récapitulatifs peuvent être obtenus avec des réponses de structures différentes. Les connaissances et compétences qui méritent un classement identique peuvent donc différer d'un élève à l'autre. Troisièmement, il n'existe pas de critères absolus par rapport auxquels puissent être appréciés les résultats d'une procédure d'établissement de normes. Si l'on peut proposer des normes en s'appuyant sur le fait qu'il faut apporter aux élèves les connaissances et les compétences dont ils auront besoin

dans la suite de leur vie, il est impossible de spécifier ces normes avec précision ; de plus, elles différeront de toute manière d'un élève à l'autre en fonction du type d'activité vers lequel chacun s'orientera.

En quatrième lieu, l'établissement effectif de normes est généralement fondé sur des appréciations individuelles. Pour cette raison, elles sont considérées comme arbitraires, non pas en ce sens qu'elles relèveraient du caprice ou de l'irrationnel, mais parce qu'elles impliquent l'exercice d'un choix et l'application de valeurs. Bien que les individus qui émettent ces appréciations soient choisis en fonction de leur connaissance du contenu des programmes et/ou des élèves pour lesquels une évaluation est conçue, et que l'on dispose de diverses procédures pour parvenir à un consensus entre ces individus, leurs appréciations seront toujours sujettes à l'erreur humaine. Cinquièmement, la détermination de normes est toujours basée, explicitement ou implicitement, sur des décisions de politique générale, selon que l'accent est mis, par exemple, sur la « compétence minimale » ou sur l'« excellence ». Comme toutes les décisions de nature sociopolitique, celles-ci sont affaire d'appréciation, sont sujettes au changement et impliquent des compromis entre des valeurs et des buts concurrents (Kane, 2001).

Même si des normes sont parfaitement prescrites et détaillées, et sont accompagnées d'une évaluation centralisée, les propositions visant à les utiliser pour spécifier le comportement des enseignants doivent prendre en compte les réalités de l'école. On ne peut supposer que tous les enseignants de toutes les écoles enseigneront le même contenu et les mêmes compétences. Le travail de chaque enseignant est toujours influencé par d'autres facteurs divers : leur compréhension des propositions de réforme, et notamment de ce que représentent les normes proposées ; leurs connaissances et leur expérience antérieures ; les conditions d'exercice de leur profession ; et enfin les caractéristiques des élèves.

La responsabilité

Les propositions d'utilisation de l'évaluation pour améliorer les normes en milieu scolaire doivent beaucoup aux mouvements en faveur

de la responsabilisation qui, en réaction aux pressions politiques, sociales et économiques, sont maintenant une des caractéristiques de l'administration gouvernementale dans de nombreux pays (Kirst, 2000). Si la qualité n'atteint pas un niveau acceptable, il paraît en effet raisonnable de poser la question : qui est responsable ?

La responsabilité peut être définie comme l'obligation de rendre compte de l'accomplissement de certaines tâches. Quand on examine la manière dont ce concept a été appliqué dans le domaine de l'enseignement et est utilisé dans les débats actuels, deux points apparaissent clairement. Premièrement, la responsabilité est centrée, pour l'essentiel, sur des résultats ou des conséquences spécifiés. Il en est ainsi en dépit du fait que l'on pourrait s'attendre à une prise en compte de tout ce qui se passe dans les écoles, et non pas des seuls résultats scolaires : cela exigerait, par exemple, que les écoles s'expliquent sur leur mission, leur organisation, leurs activités et les priorités concernant les élèves (Haney et Raczek, 1994).

En second lieu, les systèmes de responsabilisation sont souvent muets sur les responsabilités d'individus ou d'organismes autres que les enseignants et les écoles. S'il est évident que les enseignants et les élèves doivent assumer une certaine responsabilité dans l'apprentissage, il paraît raisonnable qu'ils soient exclusivement tenus responsables des questions dont ils ont le contrôle, et que la responsabilité des autres institutions, organismes et individus qui contrôlent les ressources et les activités des écoles s'inscrive également dans un système de responsabilisation. Ainsi, le gouvernement (central ou local, selon le mode d'organisation du système éducatif), les planificateurs et les gestionnaires de l'éducation semblent responsables de certains aspects de l'enseignement. D'autres qui ne sont pas directement impliqués dans l'administration du système ou dans le fonctionnement des écoles peuvent aussi être considérés comme responsables dans une certaine mesure (par exemple, les institutions de formation pédagogique, les parents et même les contribuables). Ainsi, dans le cas d'une évaluation de l'adéquation des ressources existantes qui indiquerait des déficiences dans le programme d'études ou dans la préparation pédagogique, la responsabilité de faire face à ces déficiences ne pourrait guère être imputée aux enseignants considérés individuellement. De même, si

l'on discerne des inégalités dans la répartition des ressources (par exemple, dans l'accès des filles aux cours scientifiques) ou le besoin de cibler certaines catégories d'élèves ou certaines écoles en vue d'un traitement particulier, la responsabilité de ceux qui sont chargés de la fourniture des ressources se posera. Les systèmes d'évaluation diffèrent dans la mesure où les enseignants et autres participants au système éducatif sont tenus pour responsables des résultats scolaires.

Conclusion

Dans ce chapitre ont été définis les concepts clés des propositions d'utilisation de l'évaluation pour améliorer la qualité de l'enseignement : évaluation, qualité, normes et responsabilité. Lorsque ces termes sont utilisés dans le langage de la réforme, ils sont souvent présentés de manière assez simpliste et comme posant relativement peu de problèmes. Le bref examen effectué dans ce chapitre soulève des questions sur cette utilisation et devrait contribuer à souligner la complexité des problèmes qui se posent. Il devrait aussi indiquer la nécessité d'envisager l'évaluation dans le contexte des autres éléments du système éducatif lorsque l'on évalue sa capacité d'influer sur la qualité de l'apprentissage scolaire.

III. Caractéristiques de l'évaluation dans les propositions de réformes actuelles

Les préoccupations concernant la qualité soulèvent immédiatement le problème de l'évaluation, puisqu'il est nécessaire d'émettre une forme quelconque d'appréciation des éléments d'un système éducatif avant de pouvoir porter un jugement sur sa qualité. Les mouvements de réforme actuels vont plus loin et spécifient trois caractéristiques du type d'évaluation qui sera à même d'exercer un impact sur la qualité. Premièrement, l'accent est mis sur les résultats. Deuxièmement, l'évaluation qui est menée (et parfois mandatée) fait intervenir un organisme extérieur à l'école. Troisièmement, on attend de l'évaluation qu'elle fasse office de levier de la réforme.

La focalisation sur les résultats

Dans le passé, malgré certaines exceptions, l'évaluation de la qualité de l'enseignement s'est focalisée principalement sur les intrants (exemples : taux de participation des élèves, installations physiques, matériels d'études, livres et formation pédagogique). Ce n'est plus le cas, si déficient que puisse être tel ou tel de ces éléments dans de nombreux pays. Aujourd'hui, la question posée par de nombreux commentateurs et décideurs est centrée sur les *conséquences* de l'enseignement : leur contact avec la scolarité fait-il acquérir aux élèves les connaissances, compétences, comportements et attitudes appropriés ? Cette préoccupation est compréhensible, étant donné qu'une information objective et systématique sur les acquis scolaires dans les systèmes éducatifs est assez rare, bien qu'il soit manifestement nécessaire d'en disposer pour des raisons diverses. Les décideurs ou les gestionnaires, par exemple, peuvent avoir besoin d'informations qui les aideront à former un jugement sur l'adéquation des acquis scolaires générés dans le système ; ils peuvent aussi avoir besoin de bases de données des résultats scolaires, en fonction desquelles ils apprécieront la progression d'un programme de réforme de l'enseignement, ou encore de données qui leur permettront de documenter les différences démographiques. Les enseignants peuvent

éprouver le besoin de disposer d'informations sur les résultats scolaires dans le système, auxquels ils pourront comparer ceux de leurs propres élèves ou grâce auxquels ils pourront évaluer l'efficacité de leur propre enseignement.

Parmi les déclarations qui expriment des préoccupations en matière de résultats, l'une des plus influentes se trouve dans le document adopté par la Conférence mondiale sur l'éducation pour tous organisée à Jomtien (Thaïlande) en mars 1990, selon lequel l'offre d'une éducation de base pour tous n'a de signification que si les enfants acquièrent réellement « des connaissances, de la capacité de raisonnement, des savoir-faire et des valeurs utiles ». En conséquence, l'article IV de la *Déclaration mondiale sur l'éducation pour tous* (1990) affirme que l'éducation fondamentale doit être axée sur « l'acquisition effective et les résultats de l'apprentissage, et non pas sur le seul fait de s'inscrire à une formation, de la suivre jusqu'à son terme et d'obtenir le certificat qui la sanctionne ». Le Cadre d'action de Dakar (2000), rédigé au terme de la période décennale de suivi de Jomtien, a souligné à nouveau qu'il était important de disposer d'« une définition claire et d'une évaluation précise des résultats de l'apprentissage (comportant les connaissances, les savoir-faire, les attitudes et les valeurs) » alors que les gouvernements travaillent pour assurer à tous une éducation fondamentale de qualité (UNESCO, 2000).

On peut discerner quatre éléments qui sont générateurs de préoccupations quant aux résultats. En premier lieu, de nombreux élèves (notamment ceux dont le cursus scolaire a été court) semblent tirer peu de profit de leurs expériences éducationnelles. Dans une étude effectuée au Bangladesh, par exemple, on a constaté qu'ils n'étaient guère plus d'un tiers, parmi ceux qui avaient seulement suivi l'enseignement primaire, à avoir atteint un niveau de compétence minimal dans l'apprentissage des savoir-faire fondamentaux (lecture, écriture et mathématiques) (Greaney, Khandker et Alam, 1998). Deuxièmement, on perçoit dans de nombreux pays une dégradation de la qualité, bien que les preuves à l'appui de cette perception ne soient généralement pas disponibles. On estime pourtant que la dégradation est inévitable dans une situation où les ressources

disponibles par élève sont réduites, où les apports financiers diminuent et où le nombre des élèves augmente. Troisièmement, compte tenu de l'importance de la scolarité dans la réforme économique et de la nécessité de préparer les élèves à répondre aux demandes de l'entreprise, on se préoccupe du fait que les compétences acquises même par les élèves qui demeurent longtemps dans le système éducatif puissent ne pas être adéquates pour répondre aux besoins de l'économie concurrentielle mondiale du XXI^e siècle, fondée sur l'information. Enfin, les ministères de l'Éducation recherchent de plus en plus, comme d'autres ministères, des éléments d'appréciation relatifs aux résultats de leurs investissements substantiels.

La focalisation sur l'évaluation externe

Une deuxième caractéristique des réformes actuelles impliquant l'évaluation est que l'essentiel des ressources a été investi dans des évaluations qui sont effectuées par un agent extérieur à l'école, et cela pour un certain nombre de raisons. Premièrement, ce type d'évaluation peut être mandaté à l'extérieur ; il n'est pas nécessaire de s'en remettre entièrement aux écoles pour le mettre en œuvre. Deuxièmement, un programme d'évaluation peut être établi et exécuté dans un laps de temps relativement court. Troisièmement, un tel programme est peu coûteux, comparé, par exemple, à la réduction de la taille des classes ou à la mise en œuvre de changements dans les programmes scolaires qui pourraient impliquer un perfectionnement professionnel substantiel des enseignants. Quatrièmement, on peut exercer des pressions en faveur de l'amélioration des écoles en rendant visibles les résultats d'une évaluation, voire en les rapportant dans la presse. Cinquièmement, les décideurs peuvent raisonnablement attendre, avec le temps, une certaine amélioration des notes obtenues aux tests du fait de l'attention accrue portée dans les écoles aux domaines couverts dans l'évaluation. Cependant, cette amélioration ne sera pas nécessairement accompagnée d'une amélioration dans les constructions d'acquis de plus grande ampleur (par exemple, la capacité de résolution de problèmes, les aptitudes à l'investigation) que les évaluations étaient censées appréhender (Linn, 1983, 2000 ; Torrance, 1997).

Enfin, dans de nombreux pays, l'administration de l'éducation a accordé à chaque établissement scolaire davantage de liberté dans la gestion de ses propres affaires, ou bien, étant constamment contrainte de négocier avec les syndicats des changements, même mineurs, d'ordre pratique, elle a perdu le contrôle de ce qui se passe dans les écoles. L'érosion du rôle des inspectorats scolaires illustre cette tendance. Dans de telles situations, les progrès technologiques (notamment ceux qui sont associés à l'évaluation et au traitement rapide et peu onéreux de grandes quantités de données) ont fourni le moyen d'établir un système alternatif de responsabilisation fondé sur les résultats obtenus par les élèves.

L'évaluation comme levier de la réforme

Une troisième caractéristique des réformes actuelles est que l'évaluation est considérée comme n'étant pas seulement un moyen d'obtenir des informations sur les systèmes éducatifs, mais aussi un levier de la réforme (Madaus et Kellaghan, 1992 ; Popham, 1987). Un rôle actif lui est ainsi dévolu ; elle est délibérément utilisée dans le processus éducatif pour introduire le changement dans le comportement des enseignants et les études scolaires, ce qui, par voie de conséquence, est censé élever le niveau de l'apprentissage des élèves.

L'utilisation de l'évaluation comme levier du changement soulève un certain nombre de problèmes. Premièrement, pour être efficace, elle doit être liée à – et compatible avec – la politique relative aux autres mécanismes d'orientation des études dans le système éducatif (par exemple, les programmes scolaires, les autres procédures d'évaluation, la formation pédagogique, les manuels). Deuxièmement, l'information fondée sur une évaluation doit être rendue disponible sans retard excessif et sous une forme aisément compréhensible, assurément aux planificateurs et aux administrateurs de l'éducation, probablement aux écoles et peut-être au grand public. Troisièmement, la question de la responsabilisation (présentée dans le *Chapitre II*) se pose dès lors que l'on envisage d'agir sur la base de cette information. Deux options se présentent alors. Un ministère de l'Éducation peut accepter la responsabilité principale des normes et utiliser l'information issue de l'évaluation pour prendre des décisions sur la répartition des

ressources. À l'inverse, on peut aussi attribuer la responsabilité principale aux écoles, en attendant d'elles qu'elles prennent toutes les mesures requises.

Conclusion

Les significations et les utilisations de l'évaluation sont multiples. Cependant, les mouvements de réforme actuels spécifient les caractéristiques des évaluations qui sont susceptibles d'affecter la qualité de l'enseignement : focalisation sur les résultats scolaires ; mise en œuvre de l'évaluation par un organisme extérieur à l'école ; attribution à l'évaluation du rôle de levier de la réforme. Ces trois caractéristiques s'appliquent toutes aux évaluations nationales et internationales, qui feront l'objet des chapitres suivants.

IV. L'évaluation nationale des acquis scolaires

L'évaluation nationale (parfois appelée évaluation du système éducatif, évaluation de l'apprentissage ou évaluation des résultats de l'apprentissage) est le secteur de l'évaluation qui a bénéficié de la plus grande attention dans le contexte de l'amélioration de la qualité de l'enseignement depuis les années 1990. Ce chapitre traitera d'un certain nombre d'aspects de cette forme d'évaluation : la nature de l'évaluation nationale ; les modèles de l'évaluation nationale ; les raisons d'effectuer une évaluation ; les besoins d'information auxquels elle répond ; et l'extension des activités d'évaluation (voir Greaney et Kellaghan, 1996 ; Kellaghan, 1997 ; Murphy, Greaney, Lockheed et Rojas, 1996). Les évaluations à l'échelle de la région ou de l'État qui sont effectuées dans un certain nombre de pays (comme l'Argentine, l'Australie, le Brésil, l'Inde et les États-Unis d'Amérique) sont examinées dans le cadre général de l'évaluation nationale, car elles constituent également une forme d'évaluation du système éducatif et partagent beaucoup de caractéristiques communes avec les évaluations nationales.

Qu'est-ce qu'une évaluation nationale ?

Une évaluation nationale peut être définie comme une opération conçue pour décrire le niveau des acquis, non pas d'élèves considérés individuellement, mais de la totalité d'un système éducatif ou d'une partie clairement définie de ce système (par exemple, les élèves de quatrième année d'études, ou les élèves âgés de 11 ans). Un certain nombre de pays n'ont jusqu'à maintenant connu qu'une seule évaluation nationale. Dans d'autres pays, on peut dire qu'un système d'évaluations nationales est établi, car des évaluations y sont effectuées depuis un certain temps sur une base plus ou moins régulière.

Les ressemblances et les différences entre les évaluations nationales à travers le monde sont évidentes. Toutes comportent l'évaluation de la première langue des élèves et des mathématiques. Les sciences y sont parfois incluses, ainsi qu'une deuxième langue, l'art, la musique et les sciences sociales. Dans la quasi-totalité des systèmes, les élèves sont soumis à une évaluation au niveau du primaire. Dans la plupart

des systèmes, ils sont également évalués dans le secondaire, généralement pendant la période d'enseignement obligatoire. Les systèmes diffèrent par la fréquence des évaluations. Dans certains, on effectue une évaluation chaque année, bien que le secteur du programme scolaire qui est évalué puisse varier d'une année à l'autre. Dans d'autres systèmes, les évaluations sont moins fréquentes. La participation d'une école peut être volontaire, ou bien être mandatée. Si elle est volontaire, la non-participation peut biaiser les résultats de l'évaluation de telle sorte qu'ils ne refléteront pas avec exactitude le niveau des acquis scolaires dans le système.

Les évaluations nationales étant relativement nouvelles sur la scène de l'éducation, de nombreux planificateurs et administrateurs de ce secteur peuvent n'être que vaguement familiarisés avec ce qu'elles impliquent. Ils peuvent être avertis du fait que ces évaluations entrent généralement dans la catégorie des enquêtes, et que des données sont recueillies dans un grand nombre d'écoles en utilisant des instruments d'évaluation et des questionnaires. Cependant, ils peuvent ne pas connaître les nombreuses étapes que comporte la mise en œuvre d'une évaluation.

Si les évaluations présentent des variations, elles sont nombreuses à posséder les traits caractéristiques présentés dans l'*encadré 4.1*. L'élément central de l'évaluation est la collecte de données dans les écoles. Les élèves répondent généralement à des instruments d'évaluation ou à des tests et à des questionnaires en groupe, et la durée totale consacrée par un élève à cette tâche dépasse rarement quelques heures. On peut demander aux enseignants de remplir des questionnaires dans lesquels ils fournissent des informations considérées comme appropriées à l'interprétation des résultats de leurs élèves. Dans certains cas, ils subissent également le test auquel sont soumis les élèves. L'administration des instruments d'évaluation peut être confiée aux enseignants d'une école, à moins que des examinateurs ne soient délégués de l'extérieur. Dans certaines évaluations nationales, dans l'un et l'autre cas, des superviseurs visitent un échantillon d'écoles afin de s'assurer que les procédures sont suivies comme il convient.

Encadré 4.1 Les principaux éléments d'une évaluation nationale

- Le ministère de l'Éducation désigne un agent d'exécution au sein du ministère, ou bien un organisme extérieur indépendant (par exemple un département universitaire ou une organisation de recherche) et fournit un financement.
- Les besoins de politique générale auxquels doit répondre l'évaluation sont déterminés par le ministère, parfois en consultation avec des partenaires clés en matière d'enseignement (représentants des enseignants, spécialistes des programmes scolaires, entrepreneurs, parents...).
- Le ministère de l'Éducation, ou un comité de pilotage qu'il a désigné, détermine la population qui sera soumise à l'évaluation (élèves de quatrième année, par exemple).
- On détermine le secteur des acquis scolaires qui sera évalué.
- L'agent d'exécution prépare des tests d'évaluation des acquis, ainsi que les questionnaires complémentaires et les manuels relatifs à leur administration.
- Les tests et les documents complémentaires sont l'objet d'un essai-pilote, puis ils sont révisés pour déterminer leur adéquation programmatique et technique.
- L'agent d'exécution choisit l'échantillon cible (ou la population) d'écoles ou d'élèves, prévoit l'impression de la documentation et établit des contacts avec les écoles qui ont été choisies.
- Les personnes qui administrent les tests (enseignants, inspecteurs scolaires ou étudiants diplômés) sont formées par l'agent d'exécution.
- Les instruments d'enquête (tests et questionnaires) sont administrés dans les écoles à une date spécifiée.
- Les instruments d'enquête sont collectés, renvoyés à l'agent d'exécution, apurés et préparés pour l'analyse.

Un laps de temps considérable sera nécessaire avant que le travail sur le terrain pour une évaluation soit effectué dans les écoles en vue

de la préparation à la collecte des données (par exemple, la construction d'instruments d'évaluation) et, après le travail sur le terrain, pour l'analyse des données et la préparation de rapports. Il importe que les acquis scolaires qui sont évalués soient des résultats importants de la scolarité ; que la méthode d'échantillonnage assure que les données collectées représentent de manière adéquate le niveau du système éducatif dans son ensemble (ou d'une partie clairement définie de celui-ci) ; et que les analyses cernent et décrivent les caractéristiques principales des données collectées, y compris les relations entre les variables représentatives. Toutes ces activités demanderont des ressources et un soutien politique considérables.

Modèles d'évaluation nationale

Il existe ce que l'on pourrait décrire librement comme deux « modèles » fondamentaux d'évaluation nationale ou systémique, dont l'un et l'autre ont changé de plus d'une façon au cours des années. L'un est dérivé de l'Évaluation nationale de l'état de l'éducation (NAEP), qui a commencé aux États-Unis en 1969 (voir Johnson, 1992) ; l'autre de la forme la plus récente de l'évaluation au Royaume-Uni, qui a été introduite avec le Programme d'études national en 1988 (voir Daugherty, 1997). Ces deux approches visent à obtenir des données empiriques qui puissent être interprétées comme indicatives de l'apprentissage dans des secteurs spécifiques du programme d'études et à des niveaux clairement définis du système éducatif. Elles diffèrent cependant sous certains aspects importants, et ces différences ont des implications en matière d'interprétation, d'impact et de coût.

La première différence réside dans le fait que le modèle américain utilise un échantillon d'écoles ou d'élèves, alors qu'au Royaume-Uni tous les élèves de la classe d'âge concernée participent à une opération de recensement de la population scolaire. Le Chili possède également une longue tradition d'évaluations fondées sur le recensement. Deuxièmement, le modèle américain adopte une conception analytique des acquis scolaires et se fonde sur une évaluation administrée dans des conditions normalisées au cours de quelques séances. Dans le modèle britannique, l'évaluation est holistique et fondée sur les performances, intégrée dans la pratique scolaire quotidienne et

administrée sur plusieurs journées par les enseignants, ce qui évite la séparation artificielle entre l'évaluation et l'enseignement-apprentissage (bien que cet aspect ait été modifié et que l'on se soit orienté vers l'utilisation de procédures plus standardisées). En Nouvelle-Zélande, le Programme national de suivi de l'éducation (NEMP) comporte lui aussi une série d'épreuves de performance complexes, bien que les évaluations soient effectuées, dans ce cas, par des paires d'enseignants spécialement formés provenant d'autres écoles (voir *encadré 4.2*).

Troisièmement, dans le modèle d'évaluation nationale américain, chaque élève ne subit qu'une partie d'un grand nombre d'épreuves d'évaluation (échantillonnage matriciel). Cela permet de couvrir largement le programme d'études sans demander aux élèves de consacrer un temps considérable à répondre à l'ensemble des épreuves. Conformément à sa conception holistique des acquis scolaires, le modèle britannique ne prévoit pas d'attribuer systématiquement des épreuves différentes à des élèves différents. Quatrièmement, le modèle américain ne fournit pas d'informations sur les écoles ou les circonscriptions scolaires, contrairement au modèle britannique.

Cinquièmement, une des fonctions importantes du système américain consiste à suivre dans le temps les acquis scolaires. Bien qu'il y ait eu des études sur le changement fondées sur les données obtenues dans le système britannique, le manque de normalisation des procédures d'évaluation confère aux comparaisons dans le temps une validité incertaine. Enfin, le système américain n'a pas été conçu pour avoir une incidence directe sur les écoles et sur les enseignants, bien que les informations dérivées d'une évaluation puissent être utilisées pour parvenir à des décisions de politique générale qui auront en définitive un impact sur les écoles et sur les enseignants. Par contre, le système britannique a été conçu pour avoir cet impact, d'une part en donnant aux enseignants des exemples de bonne pratique de l'évaluation, dans l'espoir d'influencer leurs priorités programmatiques et leurs méthodes d'évaluation, d'autre part par la publication d'informations sur les performances des élèves de chaque école.

Encadré 4.2 Le Programme national de suivi de la Nouvelle-Zélande

Dans le Programme national de suivi de l'éducation de la Nouvelle-Zélande, les élèves de 4^e et 8^e année subissent une large série d'épreuves, qui demandent jusqu'à quatre heures sur une durée d'une semaine. Le programme diffère de celui du Royaume-Uni dans la mesure où les épreuves sont administrées par des enseignants extérieurs à l'école dans laquelle les élèves sont soumis à l'évaluation.

Dans le cas de l'évaluation des acquis en sciences sociales, quatre approches sont utilisées : (i) onze épreuves sont administrées dans le cadre de tête-à-tête où les élèves utilisent des informations matérielles et visuelles et répondent oralement ; (ii) dix épreuves sont présentées à de petits groupes d'élèves qui travaillent ensemble ; (iii) pour cinq épreuves, les élèves travaillent de manière indépendante et consignent leurs réponses par écrit ; (iv) pour trois autres épreuves écrites, quatre élèves travaillent indépendamment.

Source : Flockton et Crooks, 1998.

Les deux modèles sont utilisés en France (voir Michel, 1995). Une des opérations consiste à prélever, environ tous les cinq ans, des échantillons d'élèves en fin de 7^e, 9^e et 10^e année pour obtenir des informations sur les acquis scolaires au niveau du système et en relation avec le programme d'études national. Dans une autre opération distincte, l'ensemble de la population des élèves subit des évaluations « de diagnostic » en début de 3^e, 6^e et 10^e année. Ces dernières évaluations sont conçues pour fournir des informations sur chacune des écoles, et des informations en retour sont envoyées aux écoles, aux élèves et à leurs parents, l'objectif étant d'aider les enseignants à adapter leur pédagogie aux besoins de leurs élèves.

Bien que la NAEP utilise encore, aux États-Unis, des échantillons d'élèves, la plupart des États de l'Union ont maintenant adopté certains aspects du modèle britannique et ont mis en place des programmes destinés à évaluer les acquis des élèves et à rendre les enseignants et/ou les élèves responsables des résultats de l'apprentissage. Quand les évaluations permettent l'identification des circonscriptions scolaires, des écoles, des enseignants et même des élèves, elles peuvent être

considérées comme des substituts des examens publics. Le développement de ce type d'évaluation a été très marqué dans des pays qui ne possèdent pas de système d'examens publics (États-Unis et pays d'Amérique latine) et dans des pays qui, s'ils possèdent ce type de système, n'organisent pas d'examens aux niveaux d'études où s'effectuent les évaluations (Royaume-Uni, France).

Les deux modèles sont également utilisés au Brésil (voir Crespo, Soares et de Mello e Souza, 2000). Comme aux États-Unis, on prélève un échantillon d'écoles et d'élèves en vue de l'évaluation nationale (*Sistema Nacional de Avaliação do Ensino Básico*). Dans certains États, l'évaluation s'applique à toutes les écoles.

Bien que les évaluations nationales ou les évaluations par État appliquées à toute une population scolaire soient aussi une caractéristique d'un certain nombre d'autres pays (comme l'Argentine ou le Chili), on en trouve l'élaboration la plus vaste et la plus systématique dans les évaluations mandatées aux États-Unis par les différents États. Ces évaluations sont complexes et diverses, et elles ont subi des variations annuelles dans les années 1990 sur la base de leurs effets perçus et de l'information en retour émanant des éducateurs, des décideurs et du public. Tous les aspects des programmes ne sont pas encore en place dans tous les États. Il existe des variations dans les procédures d'évaluation, dans les classes et les secteurs du programme d'études qui sont évalués, et dans les conséquences, pour les individus et les institutions peu performants, des performances réalisées dans les épreuves d'évaluation. Les États diffèrent aussi par l'ampleur de la diffusion des résultats des évaluations et par le choix de présenter ou non des résultats par circonscription ou par école.

De manière générale, les États attachent plus vraisemblablement un prix élevé aux performances dans le cas des enseignants, des écoles et des circonscriptions que dans celui des élèves. Cependant, dans 22 États, les conséquences sont importantes aussi bien pour les élèves que pour les enseignants ou les institutions. Pour les élèves, elles peuvent impliquer l'impossibilité de passer dans la classe supérieure ou d'obtenir un diplôme du second degré. Pour les enseignants, elle

peuvent signifier des augmentations salariales ou la décision de mettre un terme à la nomination d'un principal. Dans un nombre d'États plus réduit, les sanctions sont limitées à la publication des résultats (Shore, Pedulla et Clarke, 2001).

Pourquoi effectuer une évaluation nationale ?

Les motivations qui président à la mise en œuvre d'une évaluation nationale ne sont pas toujours clairement exprimées. Cependant, un examen des évaluations nationales à travers le monde révèle un certain nombre de motivations.

- Les évaluations nationales sont effectuées dans l'intention d'élever le niveau scolaire. Ce besoin peut résulter d'une perception de sa dégradation ou de la perception du fait que des niveaux de connaissances et de compétences plus élevés sont nécessaires dans les entreprises pour rendre le pays économiquement compétitif.
- Les évaluations nationales sont effectuées pour maintenir le niveau scolaire. En Thaïlande, par exemple, une évaluation nationale a été introduite à la fin de l'enseignement secondaire (12^e année) parce que l'on craignait que l'abolition des examens nationaux extérieurs n'ait un impact négatif sur les acquis scolaires. En fait, la « pression » exercée par l'évaluation nationale devait remplacer la « pression » exercée par l'examen national.
- Les évaluations nationales sont effectuées pour obtenir des informations qui pourront être utilisées pour faciliter les décisions concernant la répartition des ressources. Les décisions peuvent impliquer de mettre des ressources à la disposition :
 - du système éducatif en général (par exemple, pour la réforme des programmes d'études ou pour le perfectionnement des enseignants) ;
 - des catégories d'écoles qui possèdent des caractéristiques particulières, afin de redresser les injustices (par exemple, les écoles des zones rurales ou celles qui reçoivent des élèves

- de zones défavorisées sur le plan socio-économique) ; ou encore
- de certaines écoles (par exemple, pour récompenser des écoles dans lesquelles les élèves réalisent de bonnes performances lors de l'évaluation).
 - Les évaluations nationales sont effectuées afin d'obtenir des informations susceptibles de justifier les performances des élèves.
 - Les évaluations nationales sont effectuées dans le cadre d'un mouvement de modernisation, peut-être sous l'influence des agences de bailleurs de fonds, parfois en prêtant peu d'attention à la durabilité ou à la manière dont sera utilisée l'information qui est générée.
 - Les évaluations nationales sont effectuées pour modifier l'équilibre des pouvoirs dans un système éducatif. Une évaluation peut être utilisée pour faire en sorte que le contenu de l'enseignement scolaire dépende moins de l'appréciation professionnelle des enseignants et davantage des autorités centrales qui ont mandaté l'évaluation. Elle peut aussi être utilisée pour modifier l'équilibre des pouvoirs entre une autorité centrale et une autorité locale. Ainsi, une évaluation nationale effectuée en Égypte a été perçue par les autorités locales comme un moyen d'accroître le contrôle exercé par le pouvoir central sur le système éducatif. Au contraire, les évaluations conduites dans certains États de l'Argentine ont été en partie inspirées par une volonté de marquer leur indépendance par rapport au pouvoir central.
 - Les évaluations nationales sont effectuées pour compenser la médiocrité des pratiques d'évaluation des enseignants. En Égypte, l'introduction d'une évaluation nationale concernant toutes les écoles a été suggérée par des problèmes d'inefficacité (taux élevé d'abandon scolaire et de redoublement) qui ont été attribués à des déficiences dans les pratiques d'évaluation des enseignants. On a considéré que les enseignants manquaient de clarté quant aux objectifs ou aux procédures de l'évaluation des élèves, que le niveau des acquis qu'ils attendaient des élèves n'était pas informé par la connaissance des normes nationales ou régionales, que les

examens locaux étaient de qualité médiocre et que les pratiques d'évaluation n'étaient pas liées aux principes directeurs d'une action corrective. On a envisagé que l'évaluation nationale puisse fournir aux enseignants et aux parents une information de meilleure qualité sur les acquis scolaires, qu'elle discerne les élèves exposés à l'échec scolaire et, enfin, qu'elle fournisse la base d'une action corrective (Carroll, 1996).

Les besoins d'information auxquels répond une évaluation nationale

Les besoins d'information auxquels répond une évaluation nationale ainsi que les procédures qui sont choisies pour cette opération dépendent de facteurs aussi divers que les intérêts des principaux partenaires, le niveau de complexité des décisions, les opinions sur la nature de l'éducation et de la responsabilité, et la capacité de l'agent d'exécution. Les besoins des décideurs et des gestionnaires sont primordiaux. Cependant, l'implication d'autres partenaires dans la détermination des besoins peut aussi être bénéfique, dans la mesure où elle assure la représentation de leurs préoccupations. Les structures de pouvoir éducationnel-politique différant selon les pays, les intérêts des partenaires qui sont représentés dans une évaluation nationale varient nécessairement d'un pays à l'autre.

Malgré la diversité des besoins d'information auxquels répond une évaluation nationale, toute évaluation cherche à répondre à une ou plusieurs des questions suivantes :

- Quelle est la qualité de l'apprentissage scolaire dans le système éducatif (par référence aux attentes générales, aux buts du programme d'études ou à la préparation à la vie) ?
- Existe-t-il des indices de l'existence de points forts et de points faibles particuliers dans les connaissances et les compétences des élèves ?
- Certains sous-groupes de la population réalisent-ils des performances médiocres ? Existe-t-il, par exemple, des disparités

entre les acquis des garçons et des filles, des élèves des zones urbaines et de ceux des zones rurales, des élèves de langues ou de groupes ethniques différents, des élèves de régions différentes du pays ?

- Quels sont les facteurs associés aux résultats scolaires ? Dans quelle mesure ces résultats varient-ils avec les caractéristiques de l'environnement d'apprentissage (par exemple, les ressources scolaires, la préparation et la compétence des enseignants, le type d'école) ou avec la situation familiale et communautaire des élèves ?
- Les résultats scolaires varient-ils avec le temps ? La question peut présenter un intérêt particulier si des réformes du système éducatif sont en cours. Pour y répondre, il sera nécessaire d'effectuer des évaluations produisant à des moments différents des données comparables.

L'extension de l'activité d'évaluation nationale

Si la plupart des pays industrialisés possèdent depuis un certain temps des systèmes d'évaluation nationale, dont certains sont établis sur une base légale (France, Royaume-Uni), c'est seulement au cours de la dernière décennie que des efforts sérieux ont été déployés pour développer la capacité d'administrer des évaluations nationales dans d'autres parties du monde. La plupart des pays ont suivi l'approche des États-Unis, fondée sur l'utilisation d'échantillons, et ont eu recours à des instruments d'évaluation normalisés. Cependant, un certain nombre d'entre eux ont effectué des évaluations dans toutes les écoles.

Le Projet de suivi de l'UNESCO-UNICEF, en conjonction avec l'initiative de l'Éducation pour tous, a eu la responsabilité d'engager dans près de cinquante pays des évaluations par sondage axées sur l'acquisition des compétences de base en alphabétisation et en calcul ainsi que sur les savoir-faire requis dans la vie quotidienne (en rapport avec la conscience et la connaissance des questions de santé, de nutrition, d'assainissement et d'hygiène) après quatre années de scolarité (Chinapah, 2000). Le but de ce projet était de fournir aux

décideurs des données très récentes qui leur permettraient d'apprécier les résultats et de cerner les insuffisances les plus sérieuses dans les efforts déployés pour atteindre les objectifs de l'éducation de base. Vingt-quatre des pays participants sont en Afrique, douze en Asie, cinq au Moyen-Orient, trois en Amérique latine et dans les Caraïbes et trois en Europe (Kellaghan et Greaney, 2001).

Au cours des années 1990, un certain nombre de ministères de l'Éducation de l'Afrique australe ont travaillé en association avec l'Institut international de planification de l'éducation (IIEP) pour fournir aux planificateurs de l'éducation une formation aux compétences techniques requises pour le suivi et l'évaluation des systèmes d'éducation de base. Une proposition préparée par un groupe de planificateurs de l'éducation visant à créer une association connue sous le nom de Consortium d'Afrique australe pour le pilotage de la qualité de l'enseignement (SACMEQ) a reçu un accueil positif de la part de certains ministères de l'Éducation et a été officiellement lancée en 1995. Entre 1995 et 1998, sept ministères de l'Éducation ont collecté des informations, dans le cadre de SACMEQ I, sur les indicateurs de base applicables aux intrants de l'éducation, aux conditions générales de la scolarité, aux évaluations de l'équité dans la répartition des ressources humaines et matérielles et aux niveaux d'alphabétisation des élèves de 6^e année. Quinze ministères ont participé à SACMEQ II entre 1998 et 2001 (Ross *et al.*, 2000).

Le ministère vietnamien de l'Éducation et de la Formation a fait appel aux compétences techniques de l'IIEP en matière d'échantillonnage, de fusion et d'apurement de fichiers pour son enquête sur les acquis scolaires en lecture et en mathématiques en 5^e année. L'échantillon comportait 3 660 écoles, dans chacune desquelles ont été désignés au hasard 20 élèves et deux enseignants de 5^e année.

Si le SACMEQ est évoqué ici à propos des évaluations nationales, il possède également une dimension internationale, car les questions qui font l'objet des recherches, les instruments, les procédures d'échantillonnage, les populations cibles et les analyses sont communs à tous les pays participants. Ainsi, bien que les données transnationales

comparatives ne soient pas présentées dans les sept rapports nationaux de SACMEQ I qui ont été préparés (par exemple, Machingaidze, Pfukani et Shumba, 1998), les comparaisons entre pays sont possibles. Des politiques de présentation des données comparatives sont établies par les pays participants, et il a été convenu, dans l'étude SACMEQ I, que seuls les rapports nationaux, et non transnationaux, seraient publiés. Selon certaines indications, cette décision pourrait être révisée pour l'étude SACMEQ II.

En outre, l'établissement d'évaluations nationales a connu un développement rapide en Amérique latine et dans les Caraïbes au cours des années 1990. Les évaluations ont souvent été associées à la mise à disposition de bases de données pour les réformes de l'éducation. Pratiquement tous les ministères de l'Éducation de la région ont maintenant intégré des évaluations nationales dans leur agenda (Rojas et Esquivel, 1998). Il existe une similitude considérable, mais non une uniformité, entre les secteurs du programme scolaire qui font l'objet des évaluations. La langue ou la communication et les mathématiques figurent dans chacune de celles-ci, et les sciences et les sciences sociales ou l'histoire et la géographie dans près de la moitié. La plupart des évaluations ont été effectuées dans des écoles primaires. Les deux procédures, censitaire et par sondage, ont été utilisées.

Outre les évaluations nationales appuyées par l'UNESCO-UNICEF et l'IIPE, plusieurs pays ont procédé à des évaluations avec l'appui de la Banque mondiale et d'autres organismes (par exemple la Colombie, l'Inde, la Namibie, le Népal, l'Ouganda et la Zambie) (voir, par exemple, Kelly et Kanyika, 2000).

Conclusion

Les évaluations nationales représentent une forme d'activité relativement nouvelle, conçue pour fournir des informations sur le niveau des acquis scolaires dans un système éducatif, ou dans une partie clairement définie de celui-ci. Bien qu'elles diffèrent des formes d'évaluation plus traditionnelles qui étaient axées sur les élèves considérés isolément, elles se fondent sur ces évaluations pour orienter la construction d'instruments d'évaluation : en fait, les estimations

portant sur le niveau d'un système scolaire sont fondées sur l'agrégation des performances de certains élèves. Cependant, il se pose, dans une évaluation nationale, un certain nombre de problèmes qui n'apparaissent pas dans le cas d'une évaluation effectuée pour prendre une décision sur tel ou tel élève. Certains sont d'ordre technique et se rapportent à l'élaboration d'instruments d'évaluation, à l'échantillonnage et à l'analyse des données, et d'autres sont politiques et se rapportent à la détermination des besoins de politique générale et à l'obtention de l'appui des principaux partenaires.

Dans l'idéal, la décision de participer à une évaluation nationale devrait suivre une analyse des besoins d'information des décideurs, des planificateurs et des administrateurs de l'éducation, qui peuvent être aidés dans cette tâche par d'autres partenaires à l'intérieur du système éducatif. Si le fait de s'assurer la participation des principaux partenaires peut contribuer à renforcer l'appui apporté à une évaluation, on peut cependant s'attendre à quelques problèmes, à des conflits et à des résistances. En Égypte, une évaluation nationale a ainsi conduit à un conflit entre le gouvernement central et les autorités locales ; en Angleterre, les enseignants ont boycotté, en une occasion, l'organisation des évaluations.

Dans la plupart des évaluations nationales, les informations recherchées portent sur les acquis des élèves, sur les points forts et les faiblesses de leurs connaissances et de leurs savoir-faire, sur les performances de sous-groupes de la population (par région, par genre ...) et sur les corrélats du niveau scolaire. Il existe cependant des variations considérables entre les évaluations dans les secteurs des programmes scolaires que l'on évalue, dans l'année d'études ou la classe d'âge auxquelles s'applique l'évaluation, dans les éléments de la population qui en sont exclus (par exemple ceux qu'il serait trop difficile d'atteindre ou qui ne seraient pas en mesure de répondre à l'instrument d'évaluation), dans la nature de l'instrument ou du processus d'évaluation, dans l'application de l'évaluation à tous les élèves d'une année d'études ou d'un groupe d'âge déterminé, ou bien à un échantillon de la population scolaire, dans la manière de présenter les résultats, dans la manière de communiquer et d'utiliser l'information obtenue et, enfin, dans le choix de celui qui est considéré comme responsable des performances des élèves (voir Ross et Mählick, 1990).

La nouveauté et la complexité de l'évaluation nationale doivent nous rappeler la nécessité de veiller à ce que des ressources humaines et financières adéquates soient disponibles pour accomplir cette tâche de manière compétente et opportune. Malheureusement, nombreuses sont les évaluations nationales qui ne répondent pas aux normes techniques requises pour la conception des instruments, l'échantillonnage, l'administration des tests, la préparation des données, leur analyse ou leur interprétation. Par conséquent, les données obtenues peuvent fournir des informations inexactes, rendant problématique leur utilisation à des fins immédiates de politique générale ou pour un suivi dans le temps.

V. Décrire et utiliser l'information issue d'une évaluation nationale

Il importe que l'information issue d'une évaluation nationale soit présentée d'une manière significative pour les décideurs, les planificateurs et les gestionnaires de l'éducation, le personnel des écoles et tous ceux pour qui elle peut avoir des implications. Dans la première partie de ce chapitre seront décrites diverses procédures qui ont été utilisées pour rendre compte des performances scolaires. Bien que certaines évaluations comportent l'appréciation de caractéristiques non cognitives des élèves (par exemple, leur attitude envers un domaine du programme scolaire et l'intérêt qu'ils lui portent), les procédures décrites ici se limitent à l'évaluation de l'acquis cognitif des élèves. Certaines présentent une analyse détaillée des résultats obtenus par les élèves pour une tâche d'évaluation. D'autres sont axées sur des critères et examinent les performances réelles à la lumière des performances attendues, ou encore décrivent les performances en termes de niveau de compétence. D'autres examinent les performances scolaires dans un contexte comparatif, alors que certaines procédures comportent les résultats d'analyses dans lesquelles sont décrites les relations entre les acquis scolaires et d'autres aspects du système éducatif. Chaque approche possède ses avantages et ses inconvénients et peut être plus appréciée aux besoins de certains utilisateurs.

La communication et l'utilisation des conclusions de l'évaluation nationale seront examinées dans les deuxième et troisième parties de ce chapitre.

Décrire les résultats scolaires

La présentation des résultats d'une évaluation nationale diffère par un certain nombre d'aspects de leur présentation dans le cas d'évaluations aux formes plus traditionnelles. La différence essentielle résulte du fait que l'intérêt primordial d'une évaluation nationale réside dans les performances agrégées des élèves, et non dans les performances de chacun d'eux, comme c'est le cas, par exemple,

dans les examens extérieurs (publics). Ainsi, la description des performances dans une évaluation nationale peut être considérée comme un *indicateur*, qui est une mesure quantifiable utilisée pour décrire et contrôler le processus éducatif. Les indicateurs doivent contribuer à améliorer la compréhension du fonctionnement de ce processus et peuvent fournir une base pour les procédures de diagnostic et de responsabilisation (voir Burnstein, Oakes et Guiton, 1992).

Dès lors que l'on envisage un indicateur représentatif des acquis scolaires, il convient de répondre à une question fondamentale : comment représenter de manière significative les performances d'un groupe d'élèves ? Plusieurs approches ont été adoptées. Dans ce qui est peut-être la plus simple, on calcule la moyenne des scores bruts ou le pourcentage des points sur lesquels les élèves ont répondu correctement. Des méthodes plus élaborées impliquent d'échelonner les réponses des élèves. L'une de ces méthodes, dont l'utilisation est de plus en plus fréquente, comporte la modélisation des données selon la théorie de l'« item response » (IRT). Elle a été utilisée pour traiter des problèmes résultant de la conception des instruments d'évaluation, et lorsqu'il est nécessaire de mettre en équation les performances réalisées sur deux instruments ou plus. C'est ainsi qu'elle a servi à placer sur une échelle commune les résultats de l'évaluation nationale américaine obtenus pour des années différentes dans le domaine de la lecture. L'IRT est parfois controversée, notamment pour l'hypothèse qu'elle implique selon laquelle une seule aptitude ou un seul trait de caractère sous-jacents représenteraient de manière adéquate la complexité des résultats scolaires.

Rapporter les résultats d'une évaluation sur la base d'une moyenne de notations brutes, en pourcentage ou selon une échelle ne communique pas, en soi, beaucoup d'informations. Un certain nombre d'autres stratégies prenant en compte cette limitation ont été adoptées pour l'analyse des données et l'établissement de rapports. Elles peuvent être classées dans quatre rubriques : analyse des résultats scolaires ; comparaisons fondées sur des normes ; comparaisons fondées sur des critères ; et enfin examen des relations existant entre les acquis et d'autres indicateurs.

Analyse des résultats scolaires

Performances réalisées sur des items distincts. Si sommairement que soient présentés les résultats d'une évaluation, une description détaillée des performances des élèves peut révéler des informations relatives aux points faibles de l'enseignement et de l'apprentissage. Par exemple, des points faibles peuvent devenir apparents par l'affichage du pourcentage d'élèves qui répondent correctement à chaque item d'une tâche d'évaluation. Si cette information peut être trop développée pour la plupart des individus, sa lecture attentive peut, pour ceux qui sont disposés à en prendre le temps ou qui sont mus par un intérêt particulier (par exemple les spécialistes des programmes scolaires), aider à mieux appréhender les secteurs du programme dans lesquels les élèves réussissent et ceux dans lesquels ils échouent.

Performances réalisées dans certains domaines du programme. Une méthode de remplacement de la présentation détaillée des résultats (telle que le pourcentage de réponses correctes aux différents items d'une évaluation) consiste à stratifier le programme scolaire en un certain nombre de domaines et à rapporter les scores obtenus par les élèves dans chacun de ces domaines. Par exemple, les performances réalisées en mathématiques peuvent être décrites en fonction des scores moyens obtenus par les élèves dans les domaines des nombres entiers, des fractions et de la proportionnalité, des mesures, de l'estimation et du sens des nombres, de la représentation de données et des structures, relations et fonctions. Les performances réalisées en lecture peuvent être décrites en fonction des scores moyens obtenus par les élèves pour des items présentant différents types de textes : prose narrative (texte continu dans lequel le but principal de l'auteur est de raconter une histoire – réelle ou fictive) ; prose explicative (texte continu conçu pour décrire ou expliquer une information factuelle ou une opinion) ; et documents (présentations d'informations structurées sous la forme de graphiques, de tableaux, de cartes ou de notices). De nombreuses évaluations nationales fournissent des informations sur les performances des élèves par domaine. Cependant, à défaut d'analyses plus approfondies, ces informations ne révèlent pas avec précision les connaissances ou les compétences acquises par les élèves dans chaque domaine.

Performances définies en termes de processus cognitifs. Certains instruments d'évaluation sont conçus pour fournir des éléments d'appréciation sur l'utilisation par les élèves, dans les tâches d'évaluation, de processus cognitifs hiérarchiquement définis. Ainsi, dans le cas de la lecture, les niveaux de traitement d'un texte peuvent être classés par catégories de complexité croissante : réponse littérale (correspondance textuelle), paraphrase, détermination de l'idée principale d'un texte, déductions à partir de l'information reçue et, enfin, localisation et traitement de l'information. Dans une évaluation nationale effectuée en Palestine, on a constaté que les élèves de 6^e année avaient des difficultés à déceler l'idée principale d'un texte et à faire face à des tâches qui demandaient des niveaux de compréhension plus élevés (ministère palestinien de l'Éducation, 1998). L'*encadré 5.1* présente les conclusions d'une évaluation portant sur la lecture en Irlande et concernant des problèmes révélés par les élèves de 5^e année dans l'interprétation d'un texte, ainsi que dans leur aptitude à utiliser différents types de documentation.

Encadré 5.1 Problèmes de lecture des élèves révélés dans une évaluation nationale

De manière générale, les élèves ont obtenu de meilleurs résultats dans les domaines de la narration et des documents sur les questions les concernant de près et sur les questions dont la résolution demandait moins de démarches que sur les questions en TextPlus et sur celles qui exigeaient un niveau plus élevé de complexité cognitive et davantage de démarches pour leur résolution. L'analyse des réponses des élèves suggère les conclusions suivantes. Les élèves ont eu des difficultés à distinguer les voix passive et active, à discerner le thème d'un texte et à le différencier du protagoniste et de la cause de l'action, à discerner les motifs d'une action et à les différencier des conséquences d'une action, à identifier et à comprendre la relation de cause à effet, à déduire la signification d'un terme spécifique dans un texte en utilisant des indices contextuels, à comprendre le concept de fonction et à déduire la fonction du texte, à utiliser des index, des tables des matières et des dictionnaires, à suivre des renvois internes sur un horaire ou un graphique à deux dimensions ou plus et à intégrer des éléments distincts provenant de textes continus, comme des récits, et des textes discontinus, comme des cartes et des graphiques, afin de fournir une réponse qui ne soit pas apparente dans le stimulus.

Source : Cosgrove, Kellaghan, Forde et Morgan. 2000, pp. 43-44.

Comparaisons fondées sur des normes

Une comparaison fondée sur des normes indique la position d'un groupe par rapport à un ou plusieurs autres groupes qui ont répondu à un instrument d'évaluation.

Performances d'un groupe comparable. Le score moyen obtenu lors d'une évaluation nationale peut être comparé au score moyen des élèves lors d'une évaluation antérieure. La comparaison peut être effectuée dans le contexte d'un programme de réforme de l'éducation ; on devrait s'attendre à ce que le niveau de performance moyen s'élève à la suite d'une réforme, ou du moins qu'il ne diminue pas après une réforme qui aurait augmenté la proportion d'enfants scolarisés dans une classe d'âge donnée.

La possibilité de suivre dans le temps les résultats d'un système éducatif est une caractéristique intéressante du système de l'évaluation nationale. Il faut cependant admettre l'existence de trois difficultés associées à cette procédure. Premièrement, si l'on utilise différents instruments d'évaluation, il faut élaborer un barème commun pour associer les performances réalisées avec ces instruments. Il est cependant très difficile, dans ce contexte, de s'assurer que les différences de performances observées ne sont pas dues à des différences de difficulté des instruments. Deuxièmement, si le changement est suivi sur une période au cours de laquelle des modifications ont été introduites dans le programme scolaire, il peut arriver qu'un instrument d'évaluation qui était approprié avant ces modifications ne le soit plus après celles-ci. Il en ira ainsi si le programme révisé diffère de celui qu'il remplace dans ses objectifs, dans son contenu ou dans les procédures recommandées pour sa mise en œuvre. Troisièmement, l'interprétation du changement – ou de l'absence de changement – dans le temps devra prendre en compte toutes les variations des taux de participation au système éducatif qui seront intervenues durant la période considérée. Ainsi, par exemple, si l'on interprétait les résultats scolaires obtenus dans le système éducatif à la suite de l'introduction de l'enseignement primaire universel en Ouganda, il serait nécessaire de garder présentes à l'esprit les implications des changements intervenus dans les caractéristiques

des élèves du système à la suite de l'augmentation du taux de participation dans les zones rurales défavorisées, ainsi que de l'impact exercé sur les écoles par la nécessité d'accueillir un plus grand nombre d'élèves sans que les ressources aient augmenté dans les mêmes proportions.

Performances des sous-groupes dans une évaluation. On peut établir des comparaisons entre les performances réalisées par des sous-groupes dans une évaluation, ainsi qu'avec les performances d'autres groupes. Si l'on dispose depuis longtemps, dans de nombreux pays, d'indications relatives à l'accès et à la participation à l'enseignement de différents sous-groupes, les données résultant des évaluations nationales peuvent permettre de faire un autre pas en avant en fournissant des indications sur les acquis des sous-groupes. Si la taille des échantillons est suffisante, les données peuvent fournir des indications sur les résultats scolaires par sexe, par région, par secteur (urbain/rural), selon l'appartenance à un groupe ethnique ou linguistique, selon le type d'institution fréquentée par les élèves (publique/privée), ou en fait selon toute autre variable pouvant présenter un intérêt.

Comparaisons fondées sur des critères

Une comparaison fondée sur des critères implique l'interprétation des scores obtenus, non par référence aux performances d'autres élèves (ce qui est le cas des comparaisons fondées sur des normes), mais par référence à des connaissances et à des savoir-faire spécifiques, qui sont souvent basés sur des objectifs spécifiés dans un programme scolaire. Les comparaisons fondées sur des critères ont été largement utilisées pour situer les individus par rapport à des niveaux de performances spécifiques, tels ceux qui sont associés à la « compétence minimale » ou à la « maîtrise ». Dans les évaluations nationales, trois formes de comparaison fondée sur des critères ont été utilisées.

Maîtrise des objectifs du programme scolaire. Dans certaines évaluations, les performances des élèves sont notées selon que les élèves ont ou non atteint la « maîtrise » d'une série d'objectifs du programme. On supposera qu'un élève a atteint cette maîtrise lorsqu'il répondra correctement à un nombre prédéterminé d'items évaluant

l'atteinte de certains objectifs ou, dans des approches plus techniques, sur la base de la probabilité de réponses correctes à certains items au niveau de la maîtrise. Les résultats de l'évaluation seront présentés sur la base du pourcentage d'élèves qui maîtrisent chaque objectif. Certains secteurs du programme scolaire (comme les mathématiques ou les sciences) se prêtent plus aisément à cette approche que d'autres. Citons quelques exemples d'objectifs en mathématiques pour les élèves de la fin de la scolarité primaire : « Les élèves savent compléter des phrases comportant des nombres simples » ; « Les élèves savent identifier la propriété distributive » ; « Les élèves savent déterminer le plus petit commun multiple de deux nombres ».

Dans la mesure où un grand nombre d'objectifs sont probablement inclus dans la spécification d'un programme, une évaluation nationale qui décrira les performances en termes de maîtrise d'objectifs donnera un nombre de scores tout aussi important. Un aspect positif de cette conséquence est que l'information obtenue pourra fournir des indications sur le différentiel existant en matière de réussite dans la maîtrise des objectifs, indications qui pourront à leur tour constituer la base d'une exploration approfondie des problèmes d'exécution du programme ou de conception des programmes de formation pédagogique en cours d'emploi. Un aspect négatif en est que, comme dans le cas des informations relatives aux performances réalisées sur des items distincts, les décideurs et le public risquent de trouver la quantité importante de données ainsi générées difficile à absorber et à interpréter.

Performances réalisées sur des cibles de niveau. Une autre possibilité de décrire les performances réalisées dans une évaluation nationale consiste à indiquer la proportion d'élèves qui atteignent des cibles de niveau dans le programme scolaire. Cela est réalisable si le programme est structuré par niveaux de performances attendus des élèves à un âge donné. Par exemple, le programme national d'Angleterre et du pays de Galles définit huit niveaux, jusqu'à l'âge de 14 ans, dans les principales matières scolaires. Les rapports d'évaluations nationales indiquent le pourcentage d'élèves qui atteignent chaque niveau de performances.

Performances réalisées à différents « niveaux de performances ». Dans certaines évaluations, le continuum des résultats scolaires est divisé en deux niveaux ou plus (par exemple « satisfaisant » et « insuffisant » ; « élémentaire », « compétent » et « avancé »), et le pourcentage des élèves ayant atteint chaque niveau est rapporté. Il y a manifestement là une manière très attrayante de présenter des résultats, non seulement pour les planificateurs et les administrateurs de l'éducation, mais aussi pour le grand public. Elle semble particulièrement utile lorsque des évaluations nationales sont effectuées régulièrement, car elle paraît fournir une catégorisation aisément compréhensible pour des comparaisons entre niveaux dans le temps.

L'attractivité de cette catégorisation des données issues des évaluations nationales a conduit dans certains pays à l'adoption de critères utilisés dans d'autres systèmes d'évaluation pour représenter les performances. Ainsi, les élèves qui obtiennent un score de 40 % (la mention passable pour un examen extérieur) peuvent être classés comme étant « admis » ou « compétents ». Cette approche est dans une large mesure arbitraire, car elle n'est pas basée sur l'analyse des performances de l'élève qui serait nécessaire avant de le juger « compétent ».

Si l'on a conçu plusieurs procédures plus élaborées pour déterminer, sur le continuum des résultats scolaires, le ou les points auxquels on peut dire qu'un individu répond ou satisfait à un niveau de performances donné, celles-ci comportent toute une part d'arbitraire. Comme on l'a noté dans le *chapitre II*, la plupart des procédures sont fondées sur des appréciations sur la manière dont se comporteraient, lors de l'évaluation, des élèves de niveaux de « compétence » variables, et sont donc sujettes à l'erreur humaine. Il n'est pas surprenant que des méthodes différentes aboutissent à des pourcentages fort différents d'élèves considérés comme ayant atteint un niveau donné.

Les tentatives de description des connaissances et des savoir-faire qui représentent les performances à chaque niveau de compétence (« descriptifs des niveaux de résultats ») ont été importantes, que ce soit pour l'établissement de niveaux de performances valables ou pour

la communication d'informations sur la signification de ces niveaux. Dans ces procédures, des jurys dans lesquels sont représentés les spécialistes des programmes scolaires, le public, les parents d'élèves, les décideurs, le monde des affaires et les enseignants attribuent des niveaux de performances aux scores réalisés pour des tâches d'évaluation ; autrement dit, des types particuliers de connaissances et de savoir-faire sont associés avec les scores obtenus, afin de dresser un tableau de ce que savent et peuvent faire les élèves classés aux différents niveaux de compétence (Baker et Lynn, 1995). Ces tentatives se sont révélées plus difficiles qu'on ne le prévoyait à l'origine et ont donné lieu à des controverses considérables.

Relations entre les résultats et les autres indicateurs

Dans de nombreuses évaluations nationales, les données sont collectées dans des questionnaires provenant des élèves, des enseignants, des chefs d'établissement et parfois des parents, et portant sur des facteurs tels que les caractéristiques du milieu familial et de la communauté des élèves, les attitudes et les aspirations des élèves, les différents aspects de l'enseignement et de l'apprentissage (tels que le temps consacré à l'enseignement d'un thème particulier), et les ressources et l'organisation scolaires. Cette information est intéressante en elle-même et peut être fort révélatrice de la qualité de l'offre éducative. Elle a été utilisée pour décrire le contexte dans lequel fonctionnent les écoles (par exemple, en prenant en compte les caractéristiques socio-économiques de l'ensemble des élèves) lorsque l'on établit des comparaisons entre les écoles, et pour parvenir à des décisions sur l'attribution de ressources complémentaires aux écoles (voir *chapitre VII*).

Des données générales peuvent aussi être collectées pour être corrélées avec les résultats obtenus par les élèves, dans l'attente que les conclusions des analyses fassent la lumière sur les facteurs généraux concernant l'école et l'élève qui contribuent aux résultats scolaires et méritent donc d'être pris en compte dans la politique générale et la prise de décisions en matière d'éducation. Cependant, si séduisante que puisse être cette perspective, l'utilité de l'information dérivée d'un examen des relations existant entre les résultats scolaires et d'autres

informations obtenues lors d'une évaluation nationale se trouve limitée par un certain nombre de facteurs. Les données générales fournies par les enseignants et les élèves dans les questionnaires peuvent être inexactes, cependant que des caractéristiques personnelles telles que les attitudes, les valeurs et les aspirations sont extrêmement difficiles à apprécier de manière fiable. De plus, l'analyse des données elle-même pose de sérieux problèmes. Il est difficile de démêler la seule contribution aux résultats scolaires parmi des facteurs qui sont eux-mêmes corrélés, comme de distinguer entre les effets des facteurs intervenant au niveau de la communauté, de la famille, de l'école, et de l'élève considéré individuellement. Le fait que les analyses soient basées sur des données transversales limite encore les déductions qui peuvent être faites sur les facteurs dont on constate qu'ils sont associés aux résultats scolaires. L'association n'implique pas nécessairement une relation de causalité. Des déductions relatives aux « causes » des résultats ne peuvent être établies que si des éléments d'appréciation complémentaires sont disponibles à d'autres sources.

La prise en compte de ces limitations ne doit pas conduire à la conclusion que la collecte d'informations générales à l'occasion d'une évaluation nationale serait sans valeur. Ces informations peuvent aider à l'interprétation des résultats et être utiles, par exemple, pour déterminer les types d'écoles qui, à la suite d'investigations approfondies, pourront exiger des ressources complémentaires.

La communication de l'information issue d'une évaluation nationale

L'un des arguments formulés à l'appui des évaluations nationales est que les indicateurs qui décrivent les résultats peuvent fournir des informations qui seront largement utiles à des publics divers ayant des besoins et des intérêts tout aussi divers. Il est donc opportun, lors de la planification d'une évaluation, d'examiner comment l'information qu'elle produira s'inscrira dans les structures de gestion, d'administration et de responsabilisation du système éducatif, et comment elle sera communiquée et utilisée. Il semblerait qu'il y ait des variations considérables dans la pratique. Dans les études du SACMEQ, les ministères de l'Éducation ont été impliqués dès le début

dans l'établissement des questions auxquelles les études devaient apporter une réponse. Beaucoup d'efforts ont également été déployés pour s'assurer que les suggestions finales de politique générale étaient approuvées dans les sphères supérieures des ministères avant publication. Dans d'autres études, la question de savoir comment pourrait être utilisée l'information n'a guère retenu l'attention. Dans certains cas, les données n'ont pas été mises à la disposition des planificateurs et des administrateurs de l'éducation, et dans un pays au moins les données collectées n'ont pas été analysées.

Les données d'évaluation devraient intéresser particulièrement les décideurs et les administrateurs de l'éducation, comme les hauts fonctionnaires des ministères de l'Éducation (planificateurs et gestionnaires) et les responsables politiques. Sans l'information plus « objective » que véhicule ce type de données, la prise de décisions risque d'être indûment influencée par les préjugés personnels des individus, les droits acquis des directeurs d'école ou des syndicats d'enseignants et les éléments anecdotiques mis en avant par les représentants d'intérêts commerciaux, les journalistes et d'autres. Cela ne veut pas dire que les décisions soient jamais prises sur la seule base de l'information issue de l'évaluation, quand bien même elle serait disponible. La prise en compte de certaines valeurs et les pressions exercées par un électorat puissant auront toujours un rôle à jouer. Cependant, la disponibilité de l'information issue de l'évaluation doit servir à rendre la prise de décisions plus réfléchie et à renforcer sa base objective.

Les décideurs, les planificateurs et les gestionnaires doivent également veiller à ce que l'évaluation soit en accord avec les autres mécanismes d'orientation de l'enseignement dans le système éducatif. Par exemple, si les savoir-faire et les connaissances qui font l'objet d'une évaluation nationale diffèrent de ceux qui sont soumis à un examen public auquel sont attachés des enjeux considérables, il est peu probable que les enseignants prêtent beaucoup d'attention aux résultats de l'évaluation dans leur activité. Pour remédier à cette situation, il est nécessaire de s'assurer de la compatibilité des connaissances et des savoir-faire qui font l'objet d'une évaluation et d'un examen, et dont l'acquisition peut donc être considérée comme

importante pour les élèves. Les autres composantes du processus éducatif qui véhiculent une orientation de politique générale (par exemple, les manuels, les matériels didactiques et la formation pédagogique avant l'emploi et en cours d'emploi) doivent aussi être alignées sur l'évaluation et venir la renforcer.

Si les conclusions d'une évaluation indiquent, comme cela est probable, la nécessité d'approches nouvelles de l'enseignement et de l'apprentissage, les enseignants de l'ensemble du système éducatif devront recevoir l'information appropriée. Lorsqu'on dispose de données portant sur l'ensemble des écoles, la publication de classements des écoles en fonction de leurs résultats peut être interprétée comme signalant à certaines écoles la nécessité de s'améliorer. En soi, cette indication peut inciter à l'action, mais même dans ce cas, on admet généralement qu'il convient de recueillir davantage d'informations sur les déficiences et sur la manière de les traiter.

Une évaluation nationale peut exercer un impact sur les enseignants de manières diverses. Dans certaines évaluations, les enseignants participent à l'élaboration des instruments et/ou à la notation des réponses des élèves. Dans d'autres cas, ils effectuent le travail d'évaluation sur le terrain. L'expérience acquise dans l'accomplissement de ces tâches doit contribuer à l'amélioration de leurs compétences en matière d'évaluation tout en développant leur compréhension des réformes sous-jacentes à l'évaluation. Cependant, des approches plus formelles et mieux focalisées sont également nécessaires pour atteindre l'ensemble des enseignants.

Il existe plusieurs exemples d'initiatives visant à communiquer aux enseignants les résultats d'une évaluation nationale. À la suite d'une évaluation menée en Colombie, les résultats ont été diffusés par les médias et un séminaire national a été organisé pour déterminer les stratégies appropriées en vue de l'amélioration de la qualité de l'enseignement. Il a été suivi d'ateliers nationaux et locaux consacrés à l'examen des résultats et de leurs implications. L'information sur les résultats nationaux et régionaux a été envoyée à chaque enseignant sous la forme de bulletins et de brochures. Des « profils » des écoles qui avaient obtenu des scores élevés ont été préparés. Ils contenaient des informations sur des facteurs tels que le nombre d'heures

hebdomadaires consacrées à un secteur du programme, l'accent mis par les enseignants sur un élément spécifique du programme, le degré d'instruction des enseignants, les installations scolaires et, enfin, le nombre de manuels disponibles par élève.

En Argentine, l'administration responsable de l'évaluation nationale a produit une série de publications à l'intention des enseignants, des principaux et des superviseurs scolaires. Les déficiences spécifiques des écoles ont été cernées et des recommandations en vue d'y remédier ont été intégrées dans les cours de formation pédagogique avant l'emploi et en cours d'emploi. Une série de vidéos destinées aux enseignants a également été produite. En Uruguay, des rapports confidentiels sur les performances détaillées réalisées par les élèves lors d'une évaluation ont été transmis aux écoles afin de fournir une base de diagnostic des difficultés rencontrées par les enfants.

En Ouganda, la diffusion a été axée sur le fait d'aider les responsables de l'éducation au niveau de la circonscription à élaborer des stratégies d'application, dans les salles de classe, de l'information dérivée des évaluations nationales effectuées par l'Uganda National Examinations Board. Le processus comprenait une série d'ateliers et de séminaires pour les conseillers pédagogiques, les inspecteurs scolaires et les enseignants, au cours desquels ont été présentés les résultats d'une évaluation. Leurs implications en matière d'enseignement ont été examinées, ainsi que la manière d'utiliser l'information pour adapter le processus éducatif en vue d'améliorer l'apprentissage. On attendait des enseignants qu'ils recentrent leur manière de présenter le matériel d'enseignement et qu'ils prêtent davantage d'attention aux secteurs de la connaissance et du savoir-faire dont l'évaluation nationale avait discerné la faiblesse relative. Il a été décidé de faire participer tous les enseignants en poste dans les écoles, et non pas seulement ceux qui exerçaient dans l'année d'études ciblée par l'évaluation : d'abord parce que les racines des problèmes qui pourraient être cernés se trouvaient probablement dans d'autres années d'études que celle dont les élèves étaient soumis à l'évaluation, et ensuite parce que l'on considérerait comme souhaitable de traiter les problèmes de manière systématique dans l'ensemble de l'école. La nécessité de continuer d'appuyer les enseignants pendant qu'ils s'efforçaient de mettre des changements en œuvre a été reconnue.

Si louables que soient de telles initiatives, la difficulté de faire comprendre la signification des normes spécifiées au niveau central et des informations dérivées des évaluations à ceux qui doivent agir sur ces bases ne doit pas être sous-estimée. Dans une analyse de l'évaluation nationale au Brésil, il a été constaté que les rapports n'étaient pas écrits dans un langage compréhensible par tous et que très peu nombreux étaient ceux qui, au niveau des États, possédaient la capacité critique d'interpréter les résultats. Il serait étonnant qu'il n'en aille pas de même dans d'autres pays.

Les difficultés de communication avec les enseignants risquent fort d'être encore plus grandes. Le fossé entre ce que demandent les réformes et ce que comprennent et font effectivement les enseignants peut être très large. On ne peut pas présumer que les enseignants sauront ce qu'impliquent réellement les propositions de réformes, ni ce qu'ils doivent faire sur la base des résultats d'une évaluation. Même lorsque les enseignants bénéficient d'une formation et d'une orientation en cours d'emploi, les significations attribuées respectivement à la réforme par les décideurs et par les enseignants peuvent être disparates, car l'orientation sera probablement interprétée par ces derniers dans le contexte de leur compréhension et de leur pratique existantes dans le domaine du programme scolaire et de l'enseignement, lesquelles peuvent être très différentes de celles des réformateurs.

L'utilisation des données de l'évaluation nationale

Au niveau du système

On dispose de peu d'informations sur l'utilisation des conclusions des évaluations nationales, même lorsqu'elles sont communiquées à ceux qui participent au processus éducatif. Cela ne veut pas dire que ces conclusions ne sont pas utilisées, mais que leur utilisation est rarement documentée publiquement. En même temps, étant donné que l'activité d'évaluation nationale est relativement nouvelle dans de nombreux pays, il ne serait pas surprenant de découvrir que sa valeur potentielle n'est pas appréciée pleinement ou encore que les procédures d'action sur la base de ses conclusions ne sont pas entièrement développées. En fait, il a fallu longtemps avant que les

conclusions de l'Évaluation nationale de l'état de l'éducation (NAEP), aux États-Unis, ne suscitent beaucoup d'attention, et plus longtemps encore avant qu'elles n'exercent un impact quelconque sur la politique ou la pratique de l'enseignement.

Certains éléments attestent de l'utilisation des conclusions des évaluations nationales au niveau du système dans un certain nombre de pays. L'une de ces utilisations a impliqué la révision du programme scolaire. Ainsi, en Thaïlande, les performances médiocres des élèves dans les domaines des mathématiques et des sciences ont conduit à mettre davantage l'accent, dans le programme, sur l'appréhension des processus et l'application des connaissances. Au Brésil, les conclusions d'évaluations nationales ont aussi été associées à des changements dans les programmes et à des innovations pédagogiques et ont influencé la formation des ressources humaines et les politiques de financement des écoles qui desservent des zones défavorisées. Un autre exemple de l'utilisation des informations dérivées des évaluations nationales provient du Chili, où ces informations ont été utilisées pour décider du mode d'utilisation du personnel de supervision.

Au niveau des écoles

L'utilisation des conclusions des évaluations nationales pour évaluer les écoles est décrite dans le *chapitre VII*. Nous ferons ici un certain nombre de remarques générales, concernant notamment les déductions sur les responsabilités qui peuvent découler de cette utilisation.

On peut supposer que la responsabilité de traiter toute déficience discernée à l'occasion d'une évaluation nationale incombe à l'institution ou à la personne chargée d'en rendre compte (voir *chapitre II*). Le processus de collecte de données peut fournir quelques indices à cet égard. Quand des données sont collectées dans un échantillon d'écoles, les écoles participantes sont généralement assurées que les résultats individuels de chacune d'elles ne seront pas dégageés. De toute manière, seul un nombre d'écoles relativement limité (bien que représentatif) prendront part à l'évaluation. Il ne sera donc pas possible d'associer des sanctions aux performances individuelles des écoles, des enseignants ou des élèves dans l'ensemble du pays. Dans ce cas, on

peut supposer, en l'absence de preuves du contraire, que les écoles ne seront pas tenues pour les premières responsables des performances des élèves.

Au contraire, les évaluations dans lesquelles on obtient des informations sur toutes les écoles (par exemple au Royaume-Uni, en France, au Chili ou en Uruguay) ont la capacité de discerner les écoles et les enseignants aux performances médiocres, et de là à leur attribuer la responsabilité première des performances des élèves, le pas est vite franchi. La déduction selon laquelle l'école est tenue pour responsable semble avérée quand les résultats de l'évaluation sont publiés sous forme de classements.

Cependant, le fait de collecter des informations dans toutes les écoles n'implique pas nécessairement que la responsabilité des performances scolaires sera imputée à l'école ou à l'enseignant. En Uruguay, bien que l'on collecte des données dans toutes les écoles, le gouvernement central, plutôt que d'imputer des responsabilités à chacune des écoles, considère qu'il est de son ressort d'assurer la qualité de l'enseignement et d'agir au nom des communautés appauvries. En fait, son évaluation nationale a été centrée sur l'identification et le traitement des besoins sociaux qui conditionnent l'apprentissage des élèves. Les résultats concernant chaque école et chaque enseignant ne sont pas publiés et les données issues de l'évaluation sont utilisées essentiellement à des fins de diagnostic (Benveniste, 2000a).

Il existe une situation semblable dans le Territoire de la capitale, en Australie, qui s'est prononcé contre l'utilisation des données issues de son programme d'évaluation de l'alphabétisation et de l'aptitude au calcul pour évaluer les performances des enseignants, pour formuler des commentaires positifs ou négatifs sur les performances des écoles ou encore pour comparer les institutions étatiques et les autres. Cependant, ces données sont utilisées pour comparer les résultats d'ensemble à des repères nationaux, pour fournir aux écoles des informations confidentielles sur leurs performances, pour donner aux enseignants des résultats détaillés sur les performances des élèves et pour fournir aux parents des données qui leur permettront de

comparer le niveau de performances de leurs enfants avec les normes établies dans les programmes scolaires. Les écoles doivent également préparer des rapports sur leurs projets d'amélioration du niveau des performances.

Conclusion

Des procédures diverses ont été utilisées dans les évaluations nationales pour décrire les acquis des élèves, et certaines d'entre elles décrivent les performances de manière très détaillée (par exemple, le pourcentage d'élèves qui ont répondu correctement à chaque item d'une tâche d'évaluation). Si celles-ci présentent l'avantage de pouvoir être utilisées pour analyser les points forts et les faiblesses des résultats scolaires, elles ont parfois l'inconvénient de ne pas décrire d'importants mécanismes de compréhension ou d'aptitudes sous-jacents, tout en fournissant une masse de détails accablante. Les personnes dont la préoccupation essentielle n'est pas l'analyse des programmes scolaires préféreront recourir à des données plus récapitulatives.

Une procédure qui suscite beaucoup d'intérêt est la représentation des performances des élèves en fonction du niveau de « compétence » dont ils font preuve dans une tâche d'évaluation. En réponse à ce besoin, les performances ont été décrites dans certaines évaluations, avec des degrés de justification variables, en fonction du pourcentage d'élèves qui ont atteint des niveaux variables de « compétence ». Dans certains cas, la spécification de ces niveaux relevait de l'arbitraire. Dans d'autres, des efforts sérieux ont été accomplis pour déterminer des seuils entre les niveaux et pour décrire les connaissances et les savoir-faire qu'ils représentent. Cependant, l'adéquation technique de ces efforts eux-mêmes a été l'objet de débats et de controverses.

Il est inutile d'effectuer une évaluation nationale si ses conclusions ne sont pas communiquées aux utilisateurs potentiels et n'initient aucune action. Dans certains cas, il semble que peu d'efforts aient été investis dans ces tâches ; dans d'autres cas, des stratégies ont été élaborées pour communiquer les conclusions à des niveaux variables du système éducatif. Elles admettent que les principaux utilisateurs des données issues des évaluations nationales seront les décideurs et les

administrateurs du secteur de l'éducation, qui auront la responsabilité de prendre des décisions sur des questions comme l'adéquation des programmes scolaires ou l'attribution de ressources ou d'intégrer une évaluation dans d'autres mécanismes d'orientation de l'enseignement. Il sera également nécessaire d'impliquer les enseignants si l'action envisagée exige d'eux qu'ils introduisent des changements dans les connaissances et les savoir-faire qu'ils enseignent, ou dans leur manière de les enseigner. Les stratégies visant à modifier le comportement des enseignants varieront selon que l'évaluation nationale fournira des informations sur toutes les écoles d'un système ou seulement sur un échantillon d'écoles choisies pour représenter les performances générales du système. Dans les deux cas, les problèmes associés au changement de comportement des enseignants devront être pris en compte.

VI. Les évaluations internationales des résultats scolaires

Dans ce chapitre, nous décrirons quelques-unes des caractéristiques des évaluations internationales, ainsi que des exemples des modes d'utilisation des informations fournies par les évaluations. Les évaluations internationales ont dans leur procédure beaucoup de caractéristiques communes à celles des évaluations nationales, bien qu'elles en diffèrent par certains aspects, le plus évident étant le fait qu'elles doivent être conçues de manière à pouvoir être administrées dans plus d'un pays (Beaton, Postlethwaite, Ross, Spearritt et Wolf, 1999 ; Goldstein, 1996 ; Greaney et Kellaghan, 1996).

Qu'est-ce qu'une évaluation internationale ?

Les évaluations comparatives internationales trouvent leur origine dans une prise de conscience de l'absence de normes de résultats scolaires valables à l'échelle internationale et permettant aux différents pays de comparer les performances de leurs élèves respectifs. Outre l'apport de données utilisables pour ces comparaisons, il a été envisagé que les études tirent profit de la diversité existant entre les systèmes éducatifs et exploitent les conditions offertes par « un seul grand laboratoire pédagogique » aux structures et aux programmes scolaires variables, non seulement pour décrire la situation existante, mais aussi pour suggérer ce qui pourrait être réalisable dans le domaine éducatif (Husén et Postlethwaite, 1996). Ces études ont donc été encouragées comme ouvrant une perspective de recherche tout en ayant des implications de politique générale plus immédiates.

Comme pour les évaluations nationales, des instruments sont élaborés dans les évaluations internationales afin d'évaluer les connaissances et les savoir-faire des élèves. Cependant, au lieu de représenter le programme scolaire d'un seul système éducatif, ces instruments doivent être considérés comme utilisables dans l'ensemble des systèmes participants. La classe d'âge ou l'année d'études dans lesquelles les instruments doivent être utilisés feront l'objet d'un accord,

de même que les procédures de sélection des écoles ou des élèves. Toutes les études internationales ont été basées sur des échantillons de populations scolaires.

La plupart des évaluations internationales ont été effectuées sous les auspices de l'International Association for the Evaluation of Educational Achievement (IEA). Depuis 1961, où une étude pilote a été effectuée, plus de 60 pays ont participé à des études sur les résultats scolaires en lecture, en mathématiques, en sciences, en écriture, en littérature, en langues étrangères, en instruction civique et en informatique. La participation a davantage concerné les pays industrialisés que les pays moins développés. Plusieurs raisons peuvent expliquer ce déséquilibre, parmi lesquelles des limitations en matière de finances, d'infrastructures et de ressources humaines, ainsi que la conscience du fait que les différences de situation entre systèmes éducatifs rendent la documentation des différences corrélatives en matière de résultats superflue, et éventuellement embarrassante.

En 2000, après un certain nombre d'années de préparation, l'Organisation de coopération et de développement économiques (OCDE) a lancé un Programme international pour le suivi des acquis des élèves (PISA), qui est conçu pour collecter, à intervalles réguliers, des données sur l'état de préparation à la vie d'adulte des élèves âgés de 15 ans dans les pays membres de l'OCDE et dans un certain nombre d'autres pays. Des tâches d'évaluation de la maîtrise de la lecture, des mathématiques et des sciences seront administrées tous les trois ans (32 pays y ont participé en 2000). Une caractéristique notable du PISA est que les tâches d'évaluation sont basées sur un modèle « dynamique » d'apprentissage permanent, dans lequel les connaissances et compétences nouvelles nécessaires à l'adaptation à un monde en mouvement sont acquises de manière continue tout au long de la vie. Ainsi, tout en évaluant les connaissances des élèves âgés de 15 ans, le PISA examine aussi leur aptitude à réfléchir sur leurs connaissances et leur expérience et à les appliquer aux problèmes du monde réel (OCDE, 2001).

Pourquoi participer à une évaluation internationale ?

Avant que l'on s'engage à participer à une évaluation internationale, ses avantages par rapport à l'évaluation nationale méritent d'être considérés. Une caractéristique unique des évaluations internationales est qu'elles indiquent la position des acquis des élèves d'un pays donné par rapport à ceux des élèves d'autres pays, notamment dans ceux qui peuvent être considérés comme des concurrents économiques. C'est sans doute l'aspect de ces études qui suscite le plus l'attention des médias, et il peut y avoir des ramifications en attirant des investisseurs extérieurs vers tel ou tel pays. Pourtant, quelque intérêt que puisse éveiller le classement attribué à un pays, les informations que peuvent fournir des approches plus analytiques de l'examen des performances scolaires sont sans doute beaucoup plus précieuses. Toute la diversité des méthodes servant à décrire les résultats scolaires dans une évaluation nationale (cf. *chapitre V*) peut être utilisée, en fonction de la nature de l'instrument d'évaluation.

Les études internationales permettent aussi l'examen de questions aussi diverses que l'incidence de l'âge de la scolarisation (qui varie d'un pays à l'autre) ou la répartition des résultats dans les écoles d'un pays par rapport à la répartition observée dans d'autres pays, et les raisons qui peuvent expliquer ces différences (Beaton *et al.*, 1999). Il faut cependant reconnaître que même si un pays constate qu'il ne se conforme pas aux pratiques existant ailleurs en matière d'âge de la scolarisation ou que la répartition des résultats observée dans ses écoles est atypique, il arrive que l'on ne puisse faire grand-chose pour modifier cette situation.

Un autre avantage perçu de l'évaluation internationale (avantage qui peut avoir son mauvais côté) est qu'elle tend à attirer largement l'attention des hommes politiques et des médias. Cet aspect peut être utile si la publicité obtenue se traduit dans le public par une meilleure appréciation de la valeur de l'éducation et entraîne un accroissement du soutien aux dépenses d'éducation.

La participation à une évaluation internationale a aussi pour effet de présenter à certains pays l'expérience d'une certaine rigueur en

matière d'échantillonnage, d'examen détaillé des résultats, de supervision, d'analyse et de rédaction de rapports. Tout cela peut être très bénéfique pour un pays dans lequel il n'existe pas de tradition affirmée en matière de recherche pédagogique empirique et de technologies associées à celle-ci (notamment la méthodologie des enquêtes et l'élaboration d'instruments d'évaluation).

Les problèmes des évaluations internationales

Un certain nombre de problèmes qui ont été identifiés à l'occasion d'évaluations internationales méritent également d'être examinés avant toute décision de participation (Kellaghan et Grisay, 1995). L'un d'eux est lié à la difficulté de concevoir une procédure d'évaluation qui apprécie de manière adéquate les résultats de programmes scolaires divers, étant donné que les systèmes éducatifs diffèrent dans leurs buts et dans l'importance et le temps qu'ils accordent aux différents domaines d'un programme. Cette difficulté peut cependant présenter un aspect positif. Si l'analyse révèle des différences dans les résultats entre pays, lesquelles reflètent à leur tour des différences dans les programmes scolaires (par exemple dans le contenu ou dans l'enchaînement selon lequel les thèmes sont présentés), les décideurs peuvent être encouragés à réfléchir sur l'opportunité des programmes proposés dans leurs écoles. La différence résulte-t-elle d'un choix, parce que le programme est approprié à la situation locale, ou peut-être est-ce qu'il a perdu le contact avec ce qui se passait ailleurs ?

Un deuxième problème risque de se poser dans les évaluations internationales (ainsi que dans certaines évaluations nationales) lorsqu'il est nécessaire de traduire les instruments d'évaluation en une ou plusieurs langues. Si l'on doit établir des comparaisons entre des performances évaluées dans des langues différentes, il convient de réaliser que les différences qui pourront se faire jour seront peut-être imputables à des différences d'origine linguistique dans la difficulté des tâches d'évaluation. Dans la pratique, il est très difficile de s'assurer que la formulation des questions et l'opportunité culturelle du contenu sont équivalentes dans toutes les versions linguistiques d'une tâche d'évaluation.

Un troisième problème des évaluations internationales est lié à l'équivalence entre différents pays des populations et des échantillons d'élèves qui sont l'objet de comparaisons. Par exemple, des différences de performances peuvent apparaître parce que les pays diffèrent dans le degré de maintien des élèves dans leur système éducatif, ou dans le cas où certaines catégories d'élèves sont mises à l'écart de la vie scolaire normale et risquent donc d'être exclues d'une évaluation (par exemple les élèves inscrits dans des programmes spéciaux ou les élèves d'écoles dans lesquelles la langue d'enseignement diffère de celle de l'évaluation).

Un quatrième problème se pose quand le rapport relatif aux résultats d'une évaluation internationale est essentiellement centré sur le classement des pays en fonction des scores moyens de leurs élèves. Outre le fait que cette présentation peut conduire à négliger des aspects plus intéressants des performances, comme la répartition des résultats à l'intérieur d'un pays, ou les différences entre genres ou entre groupes ethniques, elle a le défaut de ne pas reconnaître que les différences traduites par les classements peuvent ne pas être indicatives de différences qui seraient considérées comme « sensibles » si des tests statistiques avaient été appliqués. Ce problème a été traité (bien que les médias l'aient souvent ignoré) de deux manières dans les rapports d'études internationales. Dans la première approche, les rapports indiquent si la moyenne obtenue par un pays diffère sensiblement (au sens statistique) du score moyen de l'ensemble des pays participants. Dans la seconde, on détermine l'importance statistique des différences entre l'ensemble des pays, considérés deux par deux.

Encadré 6.1 L'expérience de la TIMSS et de la TIMSS-R en Afrique du Sud

La participation aux études TIMSS et TIMSS-R a permis à l'Afrique du Sud de comparer ses programmes scolaires et ses normes de résultats en mathématiques et en sciences avec les normes internationales. Elle a entretenu les capacités techniques en matière d'évaluation de l'enseignement, généré trois rapports à l'usage des décideurs et des présentations destinées à chacun des neuf départements de l'éducation provinciaux, et fourni une base d'informations sur les normes des résultats scolaires et un stimulus pour des recherches complémentaires.

Une grande partie du temps de la TIMSS a été consacrée à la résolution de problèmes dus à l'inadéquation des données sur les populations scolaires par province, au manque d'expérience de la direction et du personnel administratif et à une tradition de travail selon des « horaires normaux ». Pour l'étude TIMSS-R, le contenu des manuels du projet fournis par le centre d'études international a fait l'objet d'une plus grande attention, un système de classement a été installé, des procédures de contrôle de qualité ont été mises en œuvre et un appui extérieur a été obtenu pour l'échantillonnage (qui a permis de découvrir 4 000 écoles de plus que celles qui étaient recensées). Au niveau logistique, on a rencontré des difficultés à localiser les écoles désignées, et il a été nécessaire de recourir à un agent local extérieur pour transporter les tests (du fait de l'insuffisance du système postal) et les administrer. Les problèmes rencontrés au cours de l'administration comprenaient l'absentéisme ou l'arrivée tardive des élèves, quelques fraudes dans des classes surchargées, la nervosité des élèves qui croyaient que le test contribuerait à leurs scores dans les examens publics et la réticence à demander de l'aide en cas d'incertitude sur les instructions relatives aux tests. Beaucoup d'élèves ont eu des difficultés avec les questions ouvertes, ce qui n'était pas surprenant étant donné qu'environ 80 % d'entre eux subissaient les tests dans leur deuxième, troisième et même quatrième langue.

Source : Howie, 2001

Un certain nombre d'autres facteurs méritent d'être examinés avant tout engagement à participer à une évaluation internationale. Certains sont d'ordre politique. Un pays peut souhaiter ne pas participer en cas de probabilité de performances médiocres, qui risquent fort d'être une cause d'embarras. Certains pays ont réellement collecté des données en vue d'une évaluation, puis, lorsqu'ils ont eu connaissance des résultats, ont refusé d'autoriser l'intégration de celles-

ci dans des comparaisons entre pays. D'autres facteurs sont liés aux limitations en matière de finances, d'infrastructures et de ressources humaines. La nécessité de respecter des dates limites peut se révéler très contraignante dans des pays qui manquent de personnel administratif et doivent affronter la médiocrité des infrastructures de communication (voir *encadré 6.1*).

Il peut sembler plus approprié d'obtenir des données comparatives concernant les pays voisins que des données émanant de pays du monde entier aux niveaux de développement socio-économique très différents. Pour cette raison, dix pays d'Amérique latine et des Caraïbes ont effectué en 1997 une évaluation conjointe des compétences de base en langues et en mathématiques. Les évaluations effectuées en Afrique australe et orientale par un réseau de ministères dans les années 1990 permettent aussi les comparaisons internationales au niveau régional.

L'utilisation des données de l'évaluation internationale

Il existe un certain nombre de comptes rendus de l'utilisation des conclusions d'études internationales pour informer les politiques générales et la prise de décisions. Certains décrivent des réactions générales devant les résultats. On rapporte ainsi qu'un ministre italien de l'Éducation a déclaré que les résultats de l'enquête menée par l'IEA sur six matières ont agi comme « un choc électrique » sur le système éducatif. Aux États-Unis et au Canada, les faibles scores obtenus par les élèves dans des évaluations internationales en mathématiques et en sciences ont influencé la perception qu'avait le public de la qualité de l'enseignement dispensé dans ces matières, ce qui a incité le gouvernement à leur accorder une priorité.

Des changements spécifiques dans les programmes scolaires ont aussi été attribués aux résultats d'études internationales (Kellaghan, 1996b). Au Japon, on est arrivé à la conclusion que si les performances réalisées par les élèves en mathématiques dans leurs études antérieures étaient bonnes en général, elles étaient relativement meilleures en calcul qu'en mathématiques appliquées et en analyse, qui exigent vraisemblablement des processus mentaux d'un niveau

supérieur. En conséquence, le programme scolaire a été modifié pour réduire l'importance attachée aux compétences en calcul, changement qui a été associé à une amélioration des performances en mathématiques appliquées et en analyse dans une enquête ultérieure. Par la suite, les préoccupations relatives aux attitudes négatives des élèves japonais envers les mathématiques ont conduit les autorités à proposer des programmes scolaires moins rigoureux.

Sur la base des différences observées dans les programmes projetés entre l'Irlande et d'autres pays dans la Deuxième étude internationale sur les mathématiques (SIMS), des modifications ont affecté l'importance accordée à certains sujets dans les programmes de mathématiques des écoles irlandaises. Certains engagements en faveur de modifications des programmes scolaires ont aussi été attribués à la Troisième étude internationale sur les mathématiques et les sciences (TIMSS) à Chypre et en Afrique du Sud. En Nouvelle-Zélande, pour la préparation de nouveaux programmes détaillés et de livres du maître, l'accent a été mis spécialement sur le renforcement des secteurs des mathématiques dans lesquels les résultats indiquaient une relative médiocrité des performances scolaires. En Hongrie, les résultats des études de l'IEA ont été interprétés comme soulignant l'importance du foyer familial dans les progrès scolaires de l'enfant.

L'effort le plus vaste pour obtenir des éléments de preuve concernant l'impact des études internationales réside dans la TIMSS (Robitaille, Beaton et Plomp, 2000). Vingt-neuf systèmes éducatifs ont fourni des informations sur l'impact des conclusions de l'étude au niveau national, informations dont ont été tirées un certain nombre de conclusions générales. Premièrement, dans la plupart des cas, les résultats ont été perçus comme significatifs de la nécessité d'une amélioration. Même dans quelques-uns des pays les plus performants, des préoccupations ont été exprimées sur certains aspects des conclusions (par exemple, les attitudes des élèves ou les performances réalisées sur des items exigeant des capacités de réflexion d'un niveau supérieur). Deuxièmement, les résultats semblent avoir eu un impact plus grand sur le public et dans les cercles politiques, par le biais des médias, que sur les éducateurs. Troisièmement, la TIMSS a fait

comprendre aux décideurs que les pratiques pédagogiques et les programmes scolaires de mathématiques et de sciences étaient différents dans d'autres pays : une telle prise de conscience pouvait conduire à une révision des programmes.

Conclusion

Le principal avantage des évaluations internationales par rapport aux évaluations nationales est qu'elle fournissent des données comparatives entre pays sur les résultats scolaires, et aussi, très vraisemblablement, sur la situation de l'enseignement dans les pays participants. De plus, le partage de la planification et de la mise en œuvre d'une étude internationale réduira les coûts et fournira les services d'experts en évaluations provenant de l'extérieur du pays.

Les études internationales présentent cependant des inconvénients lorsqu'on les compare aux évaluations nationales. Elles ont moins de chances d'être en accord avec les préoccupations et les problèmes locaux et, donc, de répondre aux besoins des décideurs, car il est nécessaire qu'un compromis s'établisse entre les participants sur les buts, l'ampleur et le calendrier d'une évaluation. Il est notamment peu probable, étant donné qu'un compromis sera nécessaire dans leur élaboration, que les instruments d'évaluation reflètent parfaitement les efforts déployés par chaque pays en matière de programmes scolaires.

On ne dispose que d'informations limitées sur l'utilisation qui a été faite des conclusions des études internationales. Le plus souvent, elles ont servi à apporter des modifications aux programmes envisagés, bien que l'on puisse déduire qu'elles ont aussi exercé un certain impact sur les programmes mis en œuvre dans les écoles si, comme on l'affirme, des changements sont intervenus dans le niveau des acquis des élèves japonais à la suite de changements dans le programme scolaire envisagé. Certaines conclusions (relatives, par exemple, au rôle du foyer familial dans la détermination des progrès scolaires des enfants) mettent en lumière le fait que certains facteurs affectant l'apprentissage des élèves peuvent être de ceux que ne modifient pas aisément les actions des décideurs.

Un certain nombre d'organisations mondiales (IEA, OCDE) et de consortiums locaux organisent des évaluations internationales des résultats scolaires. Un pays qui envisage de participer à une évaluation aura besoin d'établir des contacts avec ce type d'organisme. La participation à une évaluation régionale pourra être plus attirante pour un pays en développement que la participation à une évaluation à l'échelle mondiale, ne serait-ce que parce que les évaluations mondiales ont eu tendance à être dominées par les pays industrialisés et économiquement développés. Quand les pays participant à une évaluation connaissent des situations largement similaires sur les plans socio-économique et éducatif, les données comparatives générées devraient être plus significatives et appropriées à chaque pays concerné.

VII. Utiliser les données des résultats scolaires pour évaluer les écoles

Dans ce chapitre, nous examinerons l'utilisation des informations sur les résultats scolaires dérivées des évaluations nationales et des évaluations par État auxquelles participent toutes les écoles afin de former des jugements sur la qualité de chaque école (ou de chaque enseignant). On prendra également en considération les données obtenues dans les examens publics, puisque un certain nombre de pays (comme la France ou le Royaume-Uni) les utilisent au même titre que les données dérivées des évaluations nationales pour établir des classements des performances des écoles. On décrira les conséquences de la mise à disposition des informations issues des évaluations sous la forme de tableaux de classement, ainsi que les initiatives visant à répondre à la critique selon laquelle il n'est pas équitable de classer les écoles, que ce soit sur la base d'évaluations par État ou nationales ou de performances réalisées lors d'examens publics, car cette méthode ne prend pas en compte les différences existant entre les écoles dans les caractéristiques de leurs élèves ou leurs conditions de fonctionnement.

Évaluer la qualité des écoles

La plupart des parents émettent sans doute, à l'occasion, des jugements sur la qualité individuelle des écoles. Il est probable que ces jugements sont fondés sur leur observation des installations et des ressources d'une école, sur leur perception de la qualité de l'enseignement, éventuellement en fonction de la possibilité de participer à des activités hors programme, et lorsque les examens extérieurs font partie du système, les parents ont sans doute une idée précise du niveau des performances des élèves. De telles considérations peuvent les influencer dans le choix de l'école où ils enverront leurs enfants, le cas échéant. On a également vu des parents choisir leur lieu de résidence en fonction de la proximité d'une école particulière.

Dans de nombreux pays, on a recours à un inspectorat des établissements scolaires pour évaluer la qualité et l'efficacité de chaque école et de chaque enseignant. Pour ce faire, les inspecteurs observent habituellement les performances des enseignants en classe et évaluent les connaissances et les compétences des élèves sur certains points du programme scolaire. Des rapports individuels sur les écoles et les enseignants peuvent être préparés pour les archives du ministère de l'Éducation. Ces systèmes d'inspection ne font pas partie de tous les systèmes éducatifs, et même s'ils existent, leur rôle diminue à mesure qu'augmentent le pouvoir des syndicats d'enseignants et la reconnaissance de l'autonomie professionnelle des enseignants. Dans certains pays, le manque de ressources rend l'inspection scolaire inefficace.

Des procédures plus formelles sont maintenant utilisées dans un certain nombre de pays pour arriver à juger l'adéquation des performances d'une école, voire d'un enseignant, sur la base des résultats obtenus par les élèves. Elles peuvent se situer dans le contexte d'une évaluation au niveau des États (par exemple au Brésil ou aux États-Unis), d'une évaluation nationale (par exemple en Angleterre, en France, au Népal, à Maurice) ou de l'analyse des résultats d'un examen public (extérieur) (par exemple en Angleterre, en France ou au Kenya).

Actions consécutives à l'évaluation des écoles

Le type d'actions entreprises au niveau de chaque école sur la base des informations issues de l'évaluation varie très largement. On peut discerner, sur la base de la comparaison des résultats d'une école avec les données concernant d'autres écoles, les points forts et les faiblesses de cette école : il en va ainsi, par exemple, dans les évaluations nationales effectuées au Chili, en France, en Thaïlande et en Uruguay. On peut émettre des suggestions pour remédier aux difficultés d'apprentissage discernées dans les résultats et encourager les écoles à utiliser les informations à des fins d'auto-évaluation. Dans la province de Mendoza, en Argentine, l'évaluation est conçue pour influencer les enseignants dans le choix du contenu et des compétences qui auront la priorité dans les classes, et des copies des instruments

d'évaluation sont distribuées afin que les enseignants puissent modeler leurs activités scolaires sur l'évaluation (voir *encadré 7.1*). On peut enfin assurer des ressources supplémentaires aux écoles qui réalisent des performances médiocres et les aider à cerner les problèmes, à fixer des objectifs, à susciter des attentes et à élaborer des stratégies appropriées.

Encadré 7.1 Utilisation des résultats scolaires dans la province de Mendoza, en Argentine

À la suite d'une évaluation, chaque école a reçu au début de l'année scolaire un rapport individualisé contenant les résultats obtenus par l'école ainsi que les scores moyens par province, par département et par secteur (zone urbaine, limite zone urbaine et zone rurale). Les scores obtenus aux tests par l'école ont été décomposés par classe et par élève, ainsi que par secteur d'aptitude (géométrie, probabilités, etc.) et par item. Les écoles ont fait l'objet d'un classement d'ensemble par rapport à toutes les autres écoles de leur secteur. Une série de documents analytiques, de vidéos et d'ateliers sur les difficultés spécifiques rencontrées par les élèves a été fournie, qui proposait des changements pédagogiques destinés à remédier aux carences perçues. Des modèles des résultats scolaires en fonction du genre, de la localisation géographique et des variables de la gestion publique/privée ont fait l'objet de descriptions.

Source : Benveniste, 2000b.

Certaines évaluations nationales et par État tentent d'influencer le comportement des enseignants en publiant les résultats obtenus par les écoles ou en proposant des incitations à une amélioration des performances. Les évaluations qui entrent dans cette dernière catégorie sont celles dans lesquelles les enseignants reçoivent une rémunération supplémentaire si les élèves atteignent un objectif (par exemple, si 85 % atteignent un niveau de compétence satisfaisant). Les évaluations qui proposent des incitations (y compris des avantages financiers pour

les enseignants et les écoles) ont caractérisé les systèmes éducatifs d'Australie, du Chili et des États-Unis (voir *encadré 7.2*).

En Angleterre, les autorités locales, sur la base des résultats des examens, apportent un appui aux écoles dans la mise en œuvre d'un « plan d'action » conçu par l'école pour faire face à ses problèmes. Si les résultats sont particulièrement médiocres, on fixe aux écoles un délai pour leur amélioration ; en cas d'échec, elles peuvent être fermées ou reprises.

Encadré 7.2 Chili : table des indicateurs des récompenses attribuées aux écoles selon leur mérite

Le Chili utilise un système de récompenses en espèces basé sur les performances réalisées au niveau de l'école. Les écoles sont en concurrence avec d'autres écoles de la région qui présentent des caractéristiques géographiques et socio-économiques similaires. Celles qui réalisent les meilleures performances reçoivent une récompense monétaire équivalant à près d'un mois de salaire par enseignant. Les récompenses sont basées sur les scores obtenus pour les indicateurs figurant sur la table ci-dessous :

Variables	Indicateurs	Pondération 1998-1999 %
Efficacité	Scores SIMCE en mathématiques et espagnol	37
Valeur ajoutée	Moyenne d'augmentation du score SIMCE	28
Capacité d'initiatives	<ul style="list-style-type: none">• Création de conseils d'enseignants• Participation aux réunions mensuelles d'un centre pédagogique• Activités pédagogiques• Conseil d'élèves• Plan de développement de l'école• Ateliers d'enseignants	6
Amélioration des conditions de travail	<ul style="list-style-type: none">• Effectifs du personnel au complet• Remplacement des enseignants absents	2
Egalité des chances	<ul style="list-style-type: none">• Taux de rétention des élèves• Taux d'obtention du diplôme de fin d'études	22

	<ul style="list-style-type: none">• Groupement selon les aptitudes• Projets d'intégration• Absence de pratiques discriminatoires	
Intégration des parents et des enseignants	Acceptation du travail éducatif par les parents, les tuteurs et les élèves, et création de centres parentaux	5
<p>Si cette table générale des indicateurs de l'efficacité des écoles englobe une grande diversité de variables, on remarque que les résultats scolaires représentent près des deux tiers de la notation d'ensemble.</p> <p>Source : Delannoy, 2000.</p>		

Au Chili, 900 écoles (soit environ 10 % des écoles du pays) sont reconnues comme ayant besoin d'assistance sur la base de leurs performances dans les évaluations de langue et de mathématiques. On s'efforce dès lors d'améliorer leurs performances par des moyens divers : en fournissant des manuels, des bibliothèques de classe et des matériels didactiques, et en organisant le perfectionnement professionnel des enseignants par des ateliers basés dans les écoles. Des activités extrascolaires sont prévues pour les élèves. Il est établi que l'écart entre les résultats de ces écoles et ceux des autres a diminué avec le temps (Benveniste, 2000b).

Les classements des établissements scolaires

Dans certains pays, les résultats des performances réalisées par les élèves à l'occasion d'une évaluation sont publiés sous la forme d'un classement des écoles en fonction du pourcentage d'élèves qui ont atteint des objectifs de niveau. Cette pratique, qui suscite beaucoup de commentaires médiatiques et professionnels, semble être perçue comme une manière relativement peu onéreuse de résoudre les problèmes de qualité. Dans certains cas, il semble que l'on en attende une élévation du niveau des résultats du simple fait de la concurrence ainsi générée entre les écoles, les enseignants et les élèves (voir Guskey et Kifer, 1990). Elle pourrait, par exemple, servir à créer une situation de « marché » dans laquelle les parents se renseigneraient à droite et à gauche avant de choisir une école pour leurs enfants, armés des « preuves » de ses performances.

Lorsque l'on publie des données issues des évaluations indiquant comment fonctionnent les écoles – en bien ou en mal –, cela peut sans doute avoir des incidences sur leur comportement. Plusieurs facteurs indiquent cependant que l'utilisation de tableaux de classement est plus complexe qu'il pourrait y paraître à première vue, qu'elle est loin d'être sans problèmes et peut même induire en erreur. Premièrement, les performances des écoles (et par conséquent leur position relative par rapport à celles des autres écoles) peuvent varier selon la provenance des résultats utilisés. Ainsi, les classements des écoles pourront différer selon qu'ils seront basés sur les résultats d'examens publics, sur les performances réalisées dans des secteurs de base du programme, sur une évaluation effectuée par des inspecteurs scolaires ou sur une appréciation globale de l'efficacité des écoles. Deuxièmement, les classements basés sur une *même* évaluation peuvent eux aussi varier selon le critère de « réussite » qui est utilisé. Ainsi, un classement fondé sur le pourcentage d'élèves qui obtiennent un rang élevé (qui sont par exemple classés comme « avancés » sur la base de leurs performances) peut différer d'un classement basé sur le pourcentage d'élèves qui obtiennent un niveau de performances plus modéré mais acceptable (par exemple, ceux qui sont classés comme « compétents »). Troisièmement, le manque de précision qui affecte même des instruments d'évaluation bien conçus signifie que les petites différences entre écoles dans les scores appliqués aux résultats relèvent du hasard. Une différence de quelques points entre ces scores peut engendrer une différence très importante dans un classement des écoles d'un très vaste système scolaire. Quatrièmement, quelle que soit la mesure des résultats utilisée, les données obtenues peuvent varier d'une année à l'autre. Cette variation peut être due à des différences entre les cohortes d'élèves aussi bien qu'entre les enseignants.

Si l'on ne doit être tenu pour responsable que des questions que l'on maîtrise (voir *chapitre II*), un autre problème naît du fait que les résultats des élèves dépendent de diverses influences étrangères à celles sur lesquelles l'école peut agir (voir *encadré 7.3*). Si l'on a l'intention d'émettre une appréciation sur la contribution d'une école aux résultats obtenus par les élèves, il faut alors aborder un problème : comment prendre en compte de manière adéquate des facteurs qui

échappent au contrôle de l'école ? Il est donc raisonnable de demander si les élèves ont réalisé de mauvaises performances parce qu'ils n'étaient pas motivés ou qu'ils recevaient un appui insuffisant dans leur famille, ou parce que l'école disposait de ressources insuffisantes, ou encore parce que la qualité de l'enseignement dispensé était mauvaise. De plus, si le niveau des résultats obtenus par les élèves diffère d'une école à l'autre lors de leur admission, toute appréciation ultérieure qui ne prendra pas cet élément en compte risque de ne pas refléter de manière adéquate le degré de réussite qu'aura connu cette école en les faisant passer de leur niveau d'entrée initial à leur niveau de résultats actuel (voir Goldstein et Thomas, 1996 ; Torrance, 1997).

Encadré 7.3 Les facteurs qui affectent les résultats scolaires

- *Les caractéristiques des élèves* lors de leur entrée dans l'école où ils feront l'objet de l'évaluation. Parmi celles-ci figurent l'héritage génétique, les résultats scolaires, l'état sanitaire et physique, les aspirations et l'image de soi.
- *Les conditions de vie des élèves.* Celles-ci incluent la situation physique et les ressources de leur foyer et de leur communauté, ainsi que l'appui, l'orientation et les modèles de rôles dont dispose l'élève chez lui et dans sa communauté, certains de ces éléments dépendant du revenu de la famille et du degré d'instruction, des attitudes, des attentes et des aspirations des parents.
- *Les conditions de scolarité.* Il en existe trois catégories :
 - Certaines conditions sont sous le contrôle de l'école : ce que les enseignants et d'autres adultes font à l'intérieur de l'école pour aider les élèves à apprendre au cours de l'enseignement formel, et ce qui leur est offert en matière de sports, de consultation avec les parents et de tutorat.
 - D'autres conditions ne sont généralement pas sous le contrôle de l'école. Elles incluent l'espace et les autres ressources disponibles.
 - Certaines conditions de scolarité peuvent être ou ne pas être sous le contrôle de l'école. Ainsi, les caractéristiques du corps des études (qui peuvent avoir un impact sur les résultats et les aspirations des élèves considérés individuellement) peuvent être manipulées par l'école au moyen de procédures de sélection. Dans certains systèmes éducatifs, cette option n'est pas disponible.

Mesurer la contribution de l'école aux résultats scolaires

Pour répondre à la critique selon laquelle les comparaisons peuvent être inéquitables du fait que les écoles diffèrent selon les caractéristiques de leurs élèves et les conditions dans lesquelles elles fonctionnent, on peut adopter une des approches décrites ci-dessous. Dans la première, les comparaisons sont limitées à des écoles similaires par les caractéristiques socio-économiques de leurs élèves. Par exemple, en Angleterre, les écoles sont réparties en cinq catégories sur la base du pourcentage de leurs élèves qui ont droit à des repas gratuits. Elles sont ensuite incitées à comparer les performances de leurs élèves lors d'une évaluation avec celles des élèves d'autres écoles appartenant à la catégorie dans laquelle elles ont été placées. Dans cette approche, le droit à des repas gratuits est utilisé comme supplétif du deuxième facteur (conditions de vie des élèves) présenté dans l'*encadré 7.3*.

En France, les rapports concernant les résultats de l'examen de fin d'études secondaires (*Baccalauréat*) constituent un pas en avant dans la tentative pour contextualiser les performances des élèves et prendre en compte des facteurs autres que l'enseignement au sein de l'école qui peuvent les avoir affectés. Ce pas en avant consiste à fournir des informations sur ce que l'on appelle la « valeur ajoutée ». Pour cela, on calcule le taux de réussite à l'examen attendu après la prise en compte de l'origine socio-économique des élèves (classés en « très favorisés », « favorisés », « moyennement favorisés » et « défavorisés »). Les écoles reçoivent deux ensembles de données sur l'examen : le taux de réussite réel des élèves de l'école, et le taux de réussite que l'on attendrait d'une école présentant ses caractéristiques socio-économiques particulières. Une école dans laquelle le taux de réussite réel dépasse le taux attendu sera considérée comme ajoutant efficacement de la « valeur » aux résultats scolaires, alors qu'une autre dans laquelle le taux attendu dépasse le taux réel sera considérée comme ayant un fonctionnement médiocre.

Dans ce cas, il n'y a en fait aucune mesure réelle de la valeur qui a été ajoutée, puisque l'on ne fournit aucune donnée concernant

la base préexistante. Pour résoudre ce problème, certains systèmes éducatifs (par exemple dans certains États américains et dans certaines administrations locales de l'enseignement au Royaume-Uni) obtiennent des données sur les résultats des élèves à une époque antérieure afin d'obtenir une estimation des « progrès » réalisés par les élèves à partir de cette base. Cela peut être combiné avec d'autres informations (concernant par exemple l'origine socio-économique et ethnique des élèves) pour calculer les résultats attendus. Ici encore, c'est la différence entre les résultats réels et attendus des élèves qui est utilisée pour estimer la « valeur » ajoutée par les écoles.

Des classements différents se dégagent selon que l'on compare les performances des écoles en utilisant des scores de performances non ajustés, des scores ajustés en fonction de facteurs contextuels (par exemple le pourcentage d'élèves ayant droit à des repas gratuits) ou des scores ajustés en fonction de résultats antérieurs. Ainsi, une école qui mérite un rang très élevé sur la base de ses scores bruts peut se révéler « moyenne » après la prise en compte des résultats antérieurs de ses élèves. Par contre, une école qui mérite un rang inférieur sur la base de ses scores bruts peut être jugée « moyenne » après la prise en compte du pourcentage de ses élèves qui ont droit à des repas gratuits, ou encore des résultats obtenus antérieurement par ses élèves (Goldstein, 2001). Ainsi se trouve soulignée l'importance de la prise en compte des conditions de fonctionnement des écoles lors des tentatives d'estimation de leur efficacité.

Les estimations ajustées des résultats scolaires peuvent avoir une utilité pour distinguer les écoles nettement divergentes dans lesquelles il peut paraître approprié de procéder à des recherches plus approfondies afin de leur apporter une certaine assistance. Cependant, les limites que présentent ces estimations elles-mêmes quand il s'agit d'apprécier l'efficacité des écoles ne doivent pas être oubliées. Elles incluent le fait que l'éventail des résultats soumis à l'évaluation est généralement restreint, que les données sont souvent incomplètes à cause de l'absentéisme des élèves ou de leur taux de renouvellement, que les nombreux facteurs personnels, scolaires et contextuels qui peuvent influencer sur les résultats ne sont pas spécifiés de manière adéquate (tous les facteurs énoncés dans l'*encadré 7.3* ne sont pas

pris en compte), que les écoles peuvent obtenir des résultats très différents selon que les résultats initiaux sont faibles ou élevés et que les estimations ne sont pas fiables lorsqu'une école compte un petit nombre d'élèves.

L'impact des données de l'évaluation nationale sur les écoles

On a pu discerner des effets variables des évaluations nationales sur les écoles. Ces effets sont surtout prononcés quand les résultats sont placés dans un contexte d'enjeux importants, à savoir quand les informations concernant les performances des écoles sont publiées, ou quand des sanctions (envers les écoles, les enseignants ou les élèves) sont attachées aux performances.

Un certain impact peut être obtenu par la simple participation à une évaluation. Par exemple, l'évaluation britannique du programme scolaire national a cherché à influencer sur la manière dont les enseignants exerçaient et évaluaient leurs élèves, et il semblerait qu'elle ait, dans une certaine mesure, réussi à les faire davantage réfléchir sur le contenu du programme à enseigner et sur le processus éducatif. En Thaïlande, des appréciations des conséquences affectives, des compétences pratiques et de la perception sociale ont été incluses dans une évaluation nationale dans le cadre d'une initiative visant à recentrer les préoccupations des enseignants, axées sur le développement cognitif dans des secteurs spécifiques du programme (préoccupations qui étaient renforcées par les examens de fin d'études secondaires ouvrant l'accès à l'université). Cette initiative a conduit les enseignants à mettre davantage l'accent sur ces conséquences dans leur enseignement et leurs évaluations (Pravalpruk, 1996).

Aux États-Unis, un certain nombre d'effets positifs ont été relevés à la suite d'évaluations dont les résultats ont été publiés. Ainsi, les États dans lesquels la diffusion des notes obtenues aux tests avait été demandée ont enregistré, au cours des années, une légère augmentation du nombre d'élèves orientés vers des études universitaires par rapport à ceux qui n'avaient pas fait cette demande. Certains éléments indiquent aussi que la responsabilisation des écoles a aidé à réduire

les préjugés des enseignants, à élever leurs attentes concernant les performances des élèves et à augmenter les taux de collation du diplôme de fin d'études secondaires (Scheurich, Skrla et Johnson, 2000).

Des conséquences négatives ont aussi été discernées quand des enjeux importants (sous la forme d'un type quelconque de sanction) sont attachés aux performances d'une école dans une évaluation nationale ou par État. Conformément aux indications recueillies dans le cas des examens extérieurs (publics), les élèves sont préparés à partir des instruments d'évaluation et les enseignants concentrent leurs efforts sur ce qui est évalué, ce qui a pour effet une restriction du programme scolaire et une concentration sur l'obtention de hauts niveaux de performances aux tests, au détriment du développement cognitif général et de formes d'apprentissage approfondies (Kellaghan et Greaney, 1992 ; Kellaghan, Madaus et Raczek, 1996 ; Madaus, 1988). Il arrive que l'on ne prête pas l'attention voulue à d'autres aspects des acquis scolaires, tels que le développement de la conscience sociale et de l'affectivité. Certains éléments indiquent aussi que les évaluations nationales et par État sont associées, même quand les enjeux sont relativement faibles, à des comportements de fraude chez les élèves, les enseignants et les administrateurs scolaires.

Une autre conséquence de l'association d'enjeux élevés aux performances est que, dans les efforts visant à améliorer le niveau moyen de performances d'une école, les ressources et les stratégies didactiques peuvent être focalisées sur les élèves situés « à la limite » (afin d'augmenter le nombre de ceux qui sont classés comme « compétents »), au risque de négliger ceux qui obtiennent des résultats inférieurs et supérieurs. Les écoles peuvent aussi adopter des procédures visant à réduire le nombre des élèves de faible niveau qui participent aux tâches d'évaluation. Un certain nombre de stratégies peuvent être utilisées à cette fin : se montrer plus sélectif dans le choix des élèves acceptés par l'école, maintenir les élèves dans des classes qui ne font pas l'objet d'évaluations, ou encore augmenter le nombre des élèves classés comme atteints d'un handicap si la réglementation permet de les exclure de l'évaluation. Chacune de ces stratégies réduira le nombre des élèves de faible niveau qui

participent à une évaluation, ce qui aura évidemment des incidences sur les performances moyennes de l'école.

Il n'est pas difficile de trouver des exemples de l'utilisation de ce type de stratégie. Aux États-Unis, l'introduction de l'évaluation de dixième année au Texas a été associée à une augmentation du taux de rétention en neuvième année, du taux d'abandon scolaire et du temps consacré aux matières testées, aux dépens des autres (Haney, 2000). De plus, l'évaluation a été associée à une augmentation du nombre d'élèves présentés comme ayant besoin de prestations éducatives spéciales, qui ont été exemptés des évaluations. Une approche assez différente de la manipulation d'une évaluation peut être observée au Chili, avec un programme spécial destiné aux écoles aux performances médiocres. Dans ce cas, les enseignants, conscients du fait que l'interprétation des résultats en serait affectée, ont sous-évalué la situation socio-économique de leurs élèves (Himmel, 1996).

Les parents peuvent également réagir aux résultats d'évaluations à enjeux élevés en choisissant les écoles qui, selon eux, offriront à leurs enfants le plus haut niveau de résultats, aggravant ainsi les difficultés de certaines écoles et de leurs élèves. Enfin, les tests à enjeux élevés ont été associés à des problèmes de recrutement et de rétention des enseignants. Ces derniers quittent les écoles dans lesquelles les performances des élèves sont médiocres et sont attirés vers celles dans lesquelles de bonnes performances leur vaudront une prime.

Conclusion

Le but primordial d'une évaluation nationale est d'obtenir des informations qui seront utiles aux décideurs et aux gestionnaires pour la prise de décisions concernant la répartition des ressources au sein du système éducatif. Cette information doit indiquer le niveau des acquis des élèves du système pour des centres d'intérêt particuliers et, en fonction de sa conception, les performances relatives des élèves dans des composantes variables des résultats, ainsi que les résultats obtenus par des sous-groupes de la population. Cependant, les données

de certaines évaluations nationales et par État sont aussi utilisées pour évaluer individuellement les écoles et pour influencer de manière immédiate sur le comportement des enseignants dans les classes.

Si l'on prend la décision d'appliquer l'évaluation aux écoles, il convient de décider aussi si des sanctions seront ou non liées aux performances pour chacune des écoles et chacun des enseignants, voire même des élèves. Si l'on décide de ne pas faire intervenir de sanctions, on peut supposer que l'administration de l'enseignement endosse la responsabilité majeure de remédier à toute carence qui aurait été décelée. Au contraire, s'il est proposé d'associer des sanctions aux performances, un certain nombre de considérations méritent l'attention.

Premièrement, il convient de reconnaître que l'amélioration des écoles est un problème complexe dynamique, qui demande un large éventail d'informations et d'actions. L'information dérivée d'une seule évaluation ne peut en elle-même répondre à ce besoin. Elle ne nous dit pas en quoi et pourquoi une école a été « inefficace », ni comment traiter les problèmes qui peuvent exister. Elle se limite à fournir un point de départ pour des recherches plus approfondies, qui pourront prendre la forme de visites d'écoles, d'entretiens et d'observations en vue de l'analyse de la situation et de la conception de stratégies permettant de remédier aux carences. Deuxièmement, la responsabilité des écoles doit être considérée dans le contexte d'autres institutions, organismes et individus qui sont également responsables, d'une manière ou d'une autre, du fonctionnement des écoles et des ressources qui leur sont attribuées. Troisièmement, il est nécessaire, dans l'évaluation des écoles, d'équilibrer le besoin de responsabilité politique et la reconnaissance de l'autonomie professionnelle des enseignants. Une insistance excessive sur le premier de ces termes peut aisément engendrer une approche standardisée et mécanique de l'enseignement.

Quatrièmement, les performances scolaires doivent être évaluées dans le contexte du fonctionnement global de l'école : sa mission, son organisation, ses activités et toutes ses aspirations en matière de programmes scolaires, et non les seuls résultats des élèves. Cinquièmement, l'utilisation des données issues de l'évaluation pour

évaluer les écoles peut susciter de sérieux problèmes d'équité si l'on ne tient pas assez compte du fait que les écoles diffèrent largement en fonction des caractéristiques des élèves qui les fréquentent. Elle peut aussi être un mauvais service rendu aux écoles qui travaillent dans un environnement difficile, et qui peuvent réussir dans d'autres domaines que ceux qui font l'objet de l'évaluation. Lorsqu'on établit des comparaisons entre les écoles, on doit s'efforcer de prendre en compte les conditions dans lesquelles elles fonctionnent, tout en reconnaissant qu'il est impossible de spécifier et de prendre en considération la totalité de ces conditions. Sixièmement, les écoles qui inscrivent dans une marge étroite les scores obtenus pour une tâche d'évaluation ne doivent pas être considérées comme différant « sensiblement » les unes des autres dans le domaine qui a fait l'objet de l'évaluation. Les erreurs d'appréciation engendrent inévitablement de petites différences. Enfin, il convient de reconnaître que le fait d'associer des enjeux importants aux performances sous la forme de sanctions destinées aux écoles, aux enseignants ou aux élèves peut générer des pratiques pédagogiques mal fondées, différentes formes de négligence professionnelle et une culture d'« enseignement en fonction des tests ».

VIII. Les problèmes de conception d'une évaluation

Dans le *premier chapitre*, nous avons affirmé, en termes peut-être simplistes à l'excès, que les décideurs, les planificateurs et les gestionnaires du secteur de l'éducation devraient répondre à deux questions fondamentales lorsqu'ils se demandent s'il convient ou non d'investir dans l'utilisation de l'évaluation pour améliorer la qualité de l'enseignement. Faut-il investir des ressources dans les activités d'évaluation ? Et si la réponse est positive, quel est le type d'évaluation et quels sont les usages des informations issues de l'évaluation qui auront probablement un impact sur la qualité de l'enseignement dans les écoles ?

Pour répondre à la première question, il est sans doute nécessaire d'améliorer les pratiques de l'évaluation dans les examens extérieurs (publics) et dans la vie scolaire quotidienne. Des carences dans ces deux domaines ont été cernées, et l'on peut raisonnablement attendre que les initiatives pour y remédier aient des effets positifs sur la qualité de l'apprentissage des élèves. Remédier aux carences des examens extérieurs (publics), lorsqu'ils existent, n'exigerait qu'une petite augmentation des dépenses, puisque des sommes considérables leur sont déjà consacrées. Si on l'accepte, la question que doivent aborder les décideurs peut être restreinte à un choix entre l'augmentation des investissements dans les capacités d'évaluation des enseignants en salles de cours ou l'investissement dans de nouvelles formes d'évaluation représentées par les évaluations nationales et internationales.

Si le seul objectif est d'améliorer la qualité de l'apprentissage scolaire, et si les coûts n'entrent pas en ligne de compte, alors la décision doit pencher en faveur de l'amélioration des capacités d'évaluation des enseignants en salles de cours. C'est en fait la seule catégorie d'évaluation qui ait comme objectif premier et immédiat la facilitation de l'apprentissage scolaire. Si d'autres catégories d'évaluation peuvent prétendre influencer sur l'apprentissage, elles ne peuvent espérer le faire que de manière indirecte. Il est malheureusement vraisemblable que l'amélioration des capacités d'évaluation des enseignants sera aussi l'option la plus coûteuse.

Cependant, les améliorations apportées à l'apprentissage scolaire ne sont pas le seul but de l'évaluation (voir *chapitre II*). Si les décideurs, les planificateurs et les gestionnaires du secteur de l'éducation ont besoin d'informations sur l'adéquation des performances d'un système éducatif ou de ses composantes, ce besoin ne peut être satisfait que par une évaluation nationale. Si certaines informations sur les acquis des élèves dans le système éducatif peuvent être – et sont souvent – obtenues à d'autres sources, elles ne peuvent être considérées comme équivalentes à celles que l'on obtient dans une évaluation nationale. Ainsi, un système de supervision ou d'inspection peut fournir des informations importantes sur les acquis des élèves, mais celles-ci ne seront ni systématiques ni comparables entre écoles. Si une analyse des résultats des examens publics peut également fournir des informations sur les acquis scolaires, de tels résultats ne posséderont pas la base de comparabilité requise pour effectuer un suivi dans le temps, puisque, d'une part, les populations soumises aux examens subissent des changements imprévisibles et, d'autre part, que les normes ne peuvent être considérées comme constantes dans le temps (Kellaghan, 1996a).

L'extension rapide de l'activité d'évaluation nationale à travers le monde, notamment dans les années 1990, souvent dans un contexte où l'influence sociétale mettait l'accent sur les indicateurs, la responsabilité et la transparence, semblerait souligner le besoin d'informations objectives et systématiques sur la qualité de l'apprentissage des élèves à plusieurs niveaux de planification et de prise de décisions du système éducatif, informations qui ne sont pas disponibles autrement. Une autre affaire est évidemment de savoir dans quelle mesure ces informations pourraient aussi être utilisées pour améliorer la qualité de l'apprentissage scolaire. Le fait de savoir si elles se révéleront utiles à cet égard semble dépendre d'un certain nombre de facteurs : la qualité de l'évaluation, les mesures prises pour communiquer les résultats à ceux qui peuvent les étudier (administrateurs nationaux, régionaux et locaux ; responsables de l'élaboration des programmes scolaires ; enseignants et conseillers pédagogiques) et, enfin, le degré d'harmonisation de l'évaluation avec les autres mécanismes d'orientation de l'enseignement existant au sein du système éducatif.

Si l'on décide d'entreprendre une évaluation nationale, alors le besoin s'impose de prendre une série d'autres décisions. Les grandes lignes de celles-ci seront présentées dans la section suivante. Les décisions à prendre dans le cadre d'une évaluation internationale sont semblables, mais elles exigeront bien entendu l'accord des pays participants. Une brève description des avantages et des inconvénients de la participation à une évaluation internationale sera suivie de la prise en considération de la question du coût. Enfin seront examinés les problèmes qui se posent dans l'utilisation des informations issues d'une évaluation.

Les considérations qui interviennent dans la conception d'une évaluation

À la suite de la décision d'entreprendre une évaluation nationale, il convient de décider du secteur du programme scolaire qui sera évalué. Si l'on se fonde sur la pratique générale des évaluations nationales ou internationales, il semble que le besoin primordial soit de disposer d'informations sur les résultats obtenus dans les secteurs de base du programme scolaire. Si les compétences de base en alphabétisme et en calcul sont évidemment d'une importance cruciale, elles ne sont pas les seuls aspects du développement des élèves sur lesquels les écoles tentent d'influer. D'autres compétences, cognitives et non cognitives, sont également importantes et peuvent, lorsqu'elles sont incluses dans une évaluation, inciter les enseignants à leur prêter davantage d'attention dans leurs activités. Bien qu'elle soit difficile à évaluer de manière fiable, l'information sur les attitudes, les activités, les centres d'intérêt et les valeurs des élèves peut aussi être utile pour interpréter les conclusions relatives aux acquis scolaires.

Il sera également nécessaire de prendre une décision sur la forme d'évaluation qui sera employée. Il s'agira de choisir entre (a) une procédure complexe s'étendant sur plusieurs jours et incluant l'observation, les tests et l'établissement de listes de contrôle effectués sur les élèves par leurs propres enseignants (au Royaume-Uni) ou par des enseignants provenant d'une autre école (en Nouvelle-Zélande) et (b) un test administré dans des conditions standardisées, ce qui est la procédure adoptée dans la plupart des pays. La première procédure

est censée favoriser une évaluation plus « authentique » des performances des élèves dans des situations diverses et est conçue pour avoir un impact direct sur les pratiques des enseignants en matière d'évaluation et pour influencer leur choix des thèmes particuliers du programme ; la seconde est considérée comme plus « objective » et fournissant une meilleure base de comparaisons entre les écoles ou dans le temps.

La qualité des instruments utilisés pour évaluer les acquis mérite plus d'attention qu'elle n'en reçoit. Si les instruments utilisés dans les études internationales tendent à être de haute qualité technique, on ne peut en dire autant des instruments utilisés dans toutes les évaluations nationales. La nécessité d'améliorations est indiquée même dans le cas d'instruments conçus pour évaluer les compétences scolaires de base. Il est également nécessaire, dans le cas d'instruments utilisés pour l'évaluation d'élèves plus âgés, de s'assurer qu'ils évaluent bien les compétences et les connaissances de niveau supérieur dont on affirme le besoin pour les élèves, ou que les compétences qui sont évaluées ont une certaine pertinence par rapport à la vie dans une économie concurrentielle mondialisée.

Une autre décision est en rapport avec la « population » qui sera ciblée dans l'évaluation. À moins que l'on ne décide d'y inclure tous les élèves de toutes les classes d'une école, il convient de décider si la base choisie sera l'âge des élèves ou le niveau scolaire auquel ils sont inscrits. Le choix de l'année d'études plutôt que de l'âge a pour avantage qu'il est plus facile d'administrer une évaluation à des élèves qui sont dans la même année d'études qu'à des élèves dispersés dans un certain nombre de classes : dans les systèmes éducatifs où la plupart des élèves d'un âge donné sont inscrits dans la même année d'études, cela ne posera guère de problèmes. Cependant, dans les systèmes où l'âge de la scolarisation est variable et/ou la rétention dans certaines années d'études est courante, il peut être intéressant d'obtenir des données sur les résultats scolaires en rapport avec l'âge plutôt que des données en rapport avec le niveau d'études, ou s'ajoutant à celles-ci.

Si l'on se conforme à la pratique suivie dans la plupart des pays, l'âge ou les années d'études spécifiques retenus pour une évaluation

nationale s'inscriront dans la période de l'enseignement obligatoire (très probablement au niveau de l'enseignement primaire) ; de plus, les élèves auront fréquenté l'école assez longtemps pour que l'enseignement ait eu un impact. On pourra choisir plus d'une seule année d'études afin d'obtenir des indications sur les changements associés aux années de scolarité (par exemple, de 3^e en 5^e année) : cependant, si cela est prévu, quelques-uns des items des instruments d'évaluation devront être communs aux deux années.

Une autre décision porte sur le choix de mener l'opération d'évaluation dans toutes les écoles, ou bien de baser des déductions concernant les performances du système éducatif sur une évaluation réalisée dans un échantillon d'écoles. Il ne sert à rien d'adopter le premier parti, qui sera beaucoup plus onéreux, si l'on n'a pas mis en place des plans d'action au niveau de chaque école selon les informations obtenues au cours de l'évaluation.

Une autre décision doit encore être prise en ce qui concerne les données qui seront collectées en complément des données relatives aux résultats (par exemple dans les questionnaires remplis par les enseignants, par les élèves et éventuellement par les parents), ce qui peut aider à interpréter les conclusions relatives aux acquis des élèves. Il est habituel, par exemple, de collecter de l'information sur les intrants, comme les bâtiments, la qualification des enseignants et la situation socio-économique des familles et des communautés des élèves. Parfois, on obtient également des informations sur les activités scolaires et didactiques des écoles. S'il peut être tentant de collecter autant d'informations de ce type que possible, il convient de réfléchir sérieusement, au préalable, à la manière dont les données seront utilisées et à la disponibilité des ressources nécessaires pour effectuer les analyses. Il conviendra également d'être conscient du fait que certains problèmes intéressants pourraient être traités de manière plus précise et moins onéreuse dans d'autres types de recherche.

Lorsque l'on prépare une évaluation, il importe de garder présents à l'esprit les besoins d'information de publics divers, car ils peuvent avoir des implications pour la conception des instruments et la dimension des échantillons. Il peut être proposé, par exemple,

d'analyser les données par domaine du programme scolaire, en fournissant des éléments de preuves pouvant être utiles aux gestionnaires de l'éducation, aux responsables des programmes scolaires, aux rédacteurs de manuels et aux enseignants, sur les points forts et les faiblesses des élèves (par exemple, dans leur aptitude à résoudre les problèmes en mathématiques ou à intégrer les informations ou à effectuer des déductions en lecture). S'il en est ainsi, la conception d'un instrument doit être telle qu'elle produise le type d'informations demandé. Si des informations sur une sous-population (par exemple sur des différences régionales) sont demandées – informations susceptibles d'intéresser les hommes politiques, les planificateurs et ceux qui participent à la répartition des ressources –, cela aura alors des implications pour la conception et la dimension de l'échantillon requis.

Avant d'effectuer une évaluation, il convient aussi de prendre sérieusement en considération les ressources politiques et techniques qui seront nécessaires pour l'appuyer et la mener à son terme. Au niveau politique, l'implication d'un large éventail d'intervenants (par exemple dans un comité de pilotage) peut aider à s'assurer que les préoccupations de tous les participants au système éducatif sont représentées, tout en fournissant un appui à cette opération. Au niveau technique, des compétences en matière d'élaboration d'instruments d'évaluation, d'échantillonnage, d'analyse et de gestion d'une enquête à grande échelle seront requises.

Dans beaucoup de pays en développement, la situation peut rendre difficile la mise en œuvre d'une évaluation. Au niveau administratif, il peut y avoir un financement inadéquat, un manque de capacités techniques, des données incomplètes sur les écoles et des problèmes logistiques associés à la médiocrité des infrastructures des transports, des télécommunications et des services postaux. Il peut aussi être nécessaire de faire face à la diversité linguistique. Au Nigéria, l'évaluation de l'Éducation pour tous a été limitée à la 4^e année, car toute évaluation à un niveau inférieur aurait nécessité la traduction des tâches d'évaluation en 270 langues. La tradition du redoublement, particulièrement répandue dans les pays africains francophones et en Amérique latine, crée des difficultés d'échantillonnage et de logistique.

Au niveau de l'école, des problèmes peuvent être posés par les classes surchargées, le caractère inapproprié des installations et des conditions d'administration des tâches d'évaluation, et l'absentéisme des enseignants et des élèves.

La participation à une évaluation internationale

Une évaluation internationale risque moins qu'une évaluation nationale d'être centrée sur les besoins d'un système éducatif particulier. De plus, les exigences inévitables des études internationales en matière de respect des délais peuvent poser des problèmes à certains pays. Ces éléments doivent cependant être mis en balance avec certains des avantages de la participation. Le coût de l'élaboration et de l'analyse peut être inférieur, car il sera partagé avec d'autres pays, tandis que la capacité d'un système éducatif national pourra être renforcée par l'interaction avec des personnes possédant des compétences en matière d'évaluation dans d'autres systèmes éducatifs. Une évaluation internationale est également l'occasion de réviser les programmes scolaires à la lumière de ceux d'autres pays.

La dimension politique de la participation à une évaluation internationale ne doit pas être négligée. Certains pays peuvent ne pas souhaiter participer du fait que leurs élèves réaliseraient sans doute des performances médiocres, lesquelles pourraient à leur tour provoquer un certain embarras politique. S'il y a des avantages à participer à une évaluation dans laquelle sont impliqués les pays voisins, une tradition d'antagonisme avec ceux-ci peut conduire, compte tenu des conséquences politiques prévisibles, à une décision négative.

Les coûts

La question des coûts est de celles qui doivent inévitablement être abordées lorsque l'on envisage une évaluation nationale ou la participation à une évaluation internationale. Il n'est pas possible d'évaluer avec précision le rapport coûts-avantages des différents types d'évaluation en matière d'amélioration de la qualité de l'enseignement, rapport qui pourrait guider les planificateurs et les administrateurs dans leurs décisions sur l'allocation de ressources. Si les coûts peuvent

être calculés, il est impossible de déterminer exactement les avantages. Cela ne veut pas dire que la prise en compte des coûts et des avantages attendus ne doit pas figurer dans le processus de prise de décisions. Il est à tout le moins nécessaire de former une appréciation sur l'utilité prévisible d'une évaluation en matière d'aide à la prise de décisions sur l'emploi des ressources, par exemple en déterminant les faiblesses d'un programme scolaire ou les manquements du système éducatif à l'équité.

Le coût d'une évaluation peut sembler élevé lorsqu'on l'envisage dans sa globalité. Il peut cependant paraître beaucoup plus intéressant lorsqu'on calcule le coût par élève du système éducatif. De plus, un certain nombre de décisions peuvent aider à contenir les coûts. Ainsi, la localisation de l'opération d'évaluation dans une organisation de recherche, lorsqu'elle existe, présente l'avantage de faire usage des compétences appropriées existantes (par exemple, pour la conception des instruments, l'échantillonnage et l'analyse). Certains pays qui ne disposent pas de cette facilité ont localisé leur évaluation nationale dans un service des examens. Dans ce cas, des investissements peuvent être nécessaires pour apporter à certains les compétences appropriées (notamment en ce qui concerne l'échantillonnage et l'analyse). Des connaissances spécialisées peuvent aussi être nécessaires dans les bureaux de recensement scolaire (pour aider à l'échantillonnage) et parmi les conseillers pédagogiques et le corps des inspecteurs (pour l'analyse des programmes scolaires).

L'utilisation de l'information issue d'une évaluation

Étant donné l'importance considérable des ressources investies dans les évaluations nationales et internationales, il semblerait rentable de prévoir quelques dépenses supplémentaires pour s'assurer du retour des informations ainsi obtenues dans le système éducatif. Si quelques éléments indiquent que les données issues des évaluations ont été utilisées pour des décisions de politique générale (par exemple, pour modifier les programmes scolaires ou traiter des problèmes d'équité), il est vraisemblable que le potentiel des données n'est pas pleinement exploité.

Les indications provenant de nombreux pays dans lesquels des évaluations ont été administrées prouvent abondamment que les informations dérivées de ces opérations ne sont pas communiquées de manière efficace, et ne sont donc pas intégrées dans les structures décisionnelles des ministères de l'Éducation ou à d'autres niveaux du système éducatif. Dans cette situation, il ne serait pas réaliste d'attendre que les évaluations aient beaucoup d'impact sur la qualité de l'apprentissage scolaire. La communication des informations peut demander la mise en œuvre d'un certain nombre de stratégies. S'il peut être possible de fournir toutes les informations dérivées d'une évaluation dans un seul rapport, il sera peut-être plus efficace et commode de préparer un certain nombre de rapports distincts. Des données récapitulatives pourront être utiles dans un rapport destiné au public. Les décideurs demanderont davantage d'informations (par exemple, des détails sur les performances de sous-groupes de la population), alors que les analystes des programmes scolaires, les rédacteurs de manuels et les enseignants auront besoin de savoir comment les élèves se sont comportés dans des secteurs particuliers du programme scolaire. Une description des aspects techniques d'une évaluation (par exemple, l'instrument d'évaluation, la méthode d'échantillonnage, les règles d'exclusion) sera particulièrement importante si l'évaluation doit être répétée à une date ultérieure.

Obtenir des résultats sous une forme significative pour les enseignants pose des problèmes particuliers. On ne dispose en effet que d'une expérience limitée des questions que peut susciter cette initiative ou de l'impact qu'elle peut avoir. Dans certains pays, on a utilisé des récapitulatifs imprimés, des graphiques présentant les différences en fonction des matières scolaires, ainsi que l'organisation d'ateliers. Il a également été nécessaire d'assurer un appui continu à la suite de la communication initiale. Lorsqu'on communique des informations, il convient de prendre en compte la situation des écoles, par exemple l'inadéquation des programmes scolaires et des manuels, l'influence des examens extérieurs, la dimension des classes et la compétence des enseignants, chacun de ces éléments pouvant faire obstacle au changement dans le style d'enseignement.

Dans certaines évaluations nationales ou par État (mais non internationales), un problème particulier se pose quand on utilise les résultats pour accomplir des fonctions qui étaient antérieurement accomplies par d'autres méthodes d'évaluation. Par exemple, les résultats peuvent être utilisés pour évaluer les écoles (fonction traditionnelle de l'inspection scolaire), ou pour prendre des décisions sur la promotion des élèves ou la collation d'un diplôme (ce qui se faisait, dans le passé, sur la base d'un examen extérieur ou public dans certains pays, et par l'intermédiaire des enseignants dans d'autres). Quand les résultats des évaluations nationales sont utilisés à ces fins, il se pose des problèmes de responsabilité et de conséquences non prévues. Conférer aux écoles la responsabilité première des résultats scolaires est problématique pour trois raisons au moins. Premièrement, cette démarche ne reconnaît pas que les enseignants ne sont pas les seuls responsables des performances des élèves. Deuxièmement, elle ne tient pas nécessairement compte de la situation dans laquelle fonctionnent les écoles ou des caractéristiques de leurs élèves. Les tentatives allant dans ce sens n'ont pas toujours été entièrement satisfaisantes. Troisièmement, plutôt que d'améliorer la qualité de l'apprentissage, l'utilisation des résultats des évaluations pour responsabiliser les écoles peut conduire à l'adoption de stratégies et de pratiques pédagogiques diverses (telles que l'« enseignement en fonction des tests ») visant à améliorer le niveau moyen des performances de l'école sur les instruments d'évaluation, amélioration qui peut cependant ne pas concorder avec l'aptitude à des constructions mentales plus vastes que la réforme était censée promouvoir.

Conclusion

Les efforts considérables déployés à travers le monde dans les années 1980 et 1990 pour élaborer des procédures descriptives des résultats des systèmes éducatifs et pour développer la capacité de mise en œuvre de ces procédures peuvent être considérés comme attestant la perception d'un besoin d'obtenir des informations objectives et systématiques sur l'apprentissage des élèves. Au cours de cette période, des progrès importants ont été réalisés dans l'élaboration des techniques requises pour une évaluation, notamment en ce qui concerne la construction d'instruments d'évaluation et

l'échantillonnage, ainsi que le transfert des connaissances requises pour effectuer une évaluation. Il existe une technologie relativement développée qui permet de décrire la qualité de l'apprentissage dans un système éducatif, de discerner les sous-groupes dont les performances sont médiocres et, peut-être dans une moindre mesure, de déterminer si les résultats des élèves varient avec le temps. Une série d'options est maintenant à la disposition des décideurs, des planificateurs et des gestionnaires de l'éducation qui souhaitent ajouter des informations sur l'apprentissage scolaire à celles dont ils disposent déjà sur les ressources des systèmes éducatifs.

Cela ne signifie pas que les techniques de mise en œuvre des évaluations soient parfaites, ni que la qualité de toutes les évaluations nationales qui ont été effectuées puisse être considérée comme satisfaisante. Il y a encore beaucoup d'améliorations à apporter à la conception des instruments d'évaluation, et des problèmes sont apparus en matière de présentation, de communication et d'utilisation des résultats. Les responsables des politiques d'évaluation dans les ministères de l'Éducation doivent être conscients des limitations imposées par cette situation. Ils doivent aussi se rappeler que l'évaluation revêt des formes multiples et sert à plusieurs fins ; s'engager dans une évaluation nationale ne doit donc pas conduire à négliger la nécessité d'améliorer d'autres formes d'évaluation, dont le système de supervision et d'appui aux écoles, les pratiques des enseignants dans leurs propres salles de cours et les examens extérieurs (publics). Les décideurs doivent aussi s'efforcer de faire en sorte que les évaluations nationales soient harmonisées avec d'autres mécanismes majeurs d'orientation de l'enseignement au sein du système éducatif, non seulement avec d'autres systèmes d'évaluation, mais avec les programmes scolaires, la formation pédagogique, le développement des capacités de l'école et les mesures en faveur de l'équité. En l'absence de cette harmonisation, il est peu probable que l'évaluation nationale soit un instrument efficace d'amélioration de la qualité de l'enseignement.

Bibliographie

- AFT (American Federation of Teachers) ; NCME (National Council on Measurement in Education) ; NEA (National Education Association). 1990. « Standards for teacher competence in educational assessment of students ». Dans : *Educational Measurement: Issues and Practices*, 9(4), p. 30-32.
- Baker, E.L. ; Linn, R.L. 1995. « United States ». Dans : *Performance standards in education. In search of quality*. Paris : OCDE.
- Beaton, A.E. ; Postlethwaite, T.N. ; Ross, K.N. ; Spearritt, D. ; Wolf, R.M. 1999. *The benefits and limitations of international educational achievement studies*. Paris : UNESCO/Institut international de planification de l'éducation ; International Academy of Education.
- Benveniste, L. 2000a. « Student assessment as a political construction: The case of Uruguay ». Dans : *Education Policy Analysis Archives*, 8(32).
- Benveniste, L. 2000b. *The politics of student testing: a comparative analysis of national assessment systems in Southern Cone countries*. Thèse de doctorat, Stanford University.
- Burnstein, L. ; Oakes, J. ; Guiton, G. 1992. « Education indicators ». Dans : Alkin, M.C. (dir. publ.), *Encyclopedia of educational research* (6^e éd.). New York : Macmillan.
- Carroll, D. 1996. « The grade 3 and 5 assessment in Egypt ». Dans : Murphy, P. *et al.* (dir. publ.), *National assessments: testing the system*. Washington DC : Banque mondiale.
- Chinapah, V. 2000. *L'éducation pour tous : quelle qualité ? Manuel pour le suivi permanent des acquis scolaires*. Paris : UNESCO.

Bibliographie

- Cosgrove, J. ; Kellaghan, T. ; Forde, P. ; Morgan, M. 2000. *The 1998 National Assessment of English Reading*. Dublin : Educational Research Centre.
- Crespo, M. ; Soares, J.F. ; de Mello e Souza, A. 2000. « The Brazilian national evaluation system of basic education: context, process and impact ». Dans : *Studies in Educational Evaluation*, 26, p. 105-125.
- Daugherty, R. 1997. « National curriculum assessment: the experience of England and Wales ». Dans : *Educational Administration Quarterly*, 33, p. 198-218.
- Déclaration mondiale sur l'Éducation pour tous*. 1990. Adoptée par la Conférence mondiale sur l'éducation pour tous, Répondre aux besoins éducatifs fondamentaux. New York : PNUD ; UNESCO ; UNICEF ; Banque mondiale.
- Delannoy, F. 2000. *Educational reforms in Chile 1980-98: a lesson in pragmatism*. Washington DC : Banque mondiale.
- Flockton, L. ; Crooks, T. 1998. *Social studies: assessment results, 1997. National Education Monitoring Project Report 8*. Dunedin, Nouvelle-Zélande : Educational Assessment Research Unit, Université d'Otago ; Ministère de l'Éducation.
- Goldstein, H. 1996. « International comparisons of student achievement ». Dans : Little, A. et Wolf, A. (dir. publ.), *Assessment in transition. Learning, monitoring and selection in international perspective*. Oxford : Pergamon.
- Goldstein, H. 2001. « Using pupil performance data for judging schools and teachers: scope and limitations ». Dans : *British Educational Research Journal*, 27, p. 433-442.
- Goldstein, H. ; Thomas, S. 1996. « Using examination results as indicators of school and college performance ». Dans : *Journal of the Royal Statistical Society*, 159, 1^{re} partie, p. 149-163.

- Greaney, V. ; Kellaghan, T. 1996. *Monitoring the learning outcomes of education systems*. Washington DC : Banque mondiale.
- Greaney, V. ; Khandker, S.R. ; Alam, M. 1998. *Bangladesh. Assessing basic learning skills*. Dhaka : University Press ; Banque mondiale.
- Guskey, T.R. ; Kifer, E.W. 1990. « Ranking school districts on the basis of statewide test results: is it meaningful or misleading? ». Dans : *Educational Measurement: Issues and Practice*, 9(1), p. 11-16.
- Haney, W. 2000. « The myth of the Texas miracle in education ». Dans : *Education Policy Analysis Archives*, 8(41).
- Haney, W. ; Raczek, A. 1994. *Surmounting outcomes accountability in education*. Document présenté pour le Bureau d'évaluation des technologies du Congrès des États-Unis. Chestnut Hill MA : Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- Himmel, E. 1996. « National assessment in Chile ». Dans : Murphy, P. *et al.* (dir. publ.), *National assessments. Testing the system*. Washington DC : Banque mondiale.
- Howie, S.J. 2001. *Mathematics and science performance in grade 8 in South Africa 1998/1999*. Pretoria : Human Sciences Research Council.
- Husén, T. ; Postlethwaite, T.N. 1996. « A brief history of the International Association for the Evaluation of Educational Achievement (IEA) ». Dans : *Assessment in Education*, 3, p. 129-141.
- Johnson, E.G. 1992. « The design of the National Assessment of Educational Progress ». Dans : *Journal of Educational Measurement*, 29, p. 95-110.

- Kane, M.T. 2001. « So much remains the same: conception and status of validation in setting standards ». Dans : Cizek, G.J., *Setting performance standards. Concepts, methods and perspectives*. Mahwah NJ : Lawrence Erlbaum.
- Kellaghan, T. 1996a. « Can public examinations be used to provide information for national assessment? ». Dans : Murphy, P. *et al.* (dir. publ.), *National assessments. Testing the system*. Washington DC : Banque mondiale.
- Kellaghan, T. 1996b. « IEA studies and educational policy ». Dans : *Assessment in Education*, 3, p. 143-160.
- Kellaghan, T. 1997. « Seguimiento de los resultados educativos nacionales ». Dans : Álvarez H.B. et Ruiz-Casares, M. (dir. publ.), *Evaluación y reforma educativa. Opciones de política*. Washington DC : US Agency for International Development.
- Kellaghan, T. ; Greaney, V. 1992. *Using examinations to improve education. A study in fourteen African countries*. Washington DC : Banque mondiale.
- Kellaghan, T. ; Greaney, V. 2001. « The globalisation of assessment in the 20th century ». Dans : *Assessment in Education*, 8, p. 87-102.
- Kellaghan, T. ; Grisay, A. 1995. « Comparaisons internationales des résultats des élèves : problèmes et perspectives ». Dans : *Mesurer les résultats scolaires*. Paris : OCDE.
- Kellaghan, T. ; Madaus, G.F. ; Raczek, A. 1996. *The use of external examinations to improve student motivation*. Washington DC : American Educational Research Association.
- Kelly, M.J. ; Kanyika, J. 2000. *Learning achievement at the middle basic level. Summary report on Zambia's national assessment project 1999*. Lusaka : Examinations Council of Zambia.

- Kirst, M.W. 2000. « Accountability : implications for state and local policymakers ». Dans : Stufflebeam, D.L. ; Madaus, G.F. ; Kellaghan, T. (dir. publ.). *Evaluation models: viewpoints on educational and human services evaluation* (2^e éd.). Boston : Kluwer Academic.
- Linn, R.L. ; 1983. «Testing and instructions: links and distinctions ». Dans : *Journal of Educational Measurement*, 20, p. 179-189.
- Linn, R.L. 2000. « Assessments and accountability ». Dans : *Educational Researcher*, 29(2), p. 4-16.
- Machingaidze, T. ; Pfukani, P. ; Shumba, S. 1998. *The quality of education: some policy suggestions based on a survey of schools in Zimbabwe*. Paris : UNESCO/Institut international de planification de l'éducation.
- Madaus, G.F. 1988. « The influence of testing on the curriculum ». Dans : Tanner, L.N. (dir. publ.), *Critical issues in curriculum*. Quatre-vingt-septième annuaire de la National Society for the Study of Education. 1^{re} partie. Chicago : University of Chicago Press.
- Madaus, G.F. ; Kellaghan, T. 1992. « Curriculum evaluation and assessment ». Dans : Jackson, P. (dir. publ), *Handbook of research on curriculum*. New York : Macmillan.
- Michel, A. 1995. « France ». Dans : *Performance standards in education. In search of quality*. Paris : OCDE.
- Ministère palestinien de l'Éducation. 1998. *Level of achievement in Arabic for sixth-grade students at the end of the lower basic stage in Palestine*. Ramallah : Ministère palestinien de l'Éducation.
- Murphy, P. ; Greaney, V. ; Lockheed, M.E. ; Rojas, C. (dir. publ.). 1996. *National assessments. Testing the system*. Washington DC : Banque mondiale.

- OCDE (Organisation de coopération et de développement économiques). 2001. *Connaissances et compétences : des atouts pour la vie. Premiers résultats de PISA 2000*. Paris : OCDE.
- Popham, W.J. 1987. « The merits of measurement-driven instruction ». Dans : *Phi Delta Kappan*, 68, p. 679-682.
- Pravalpruk, K. 1996. « National assessment in Thailand ». Dans : Murphy, P. *et al.* (dir. publ.), *National assessments. Testing the system*. Washington DC : Banque mondiale.
- Robitaille, D.F. ; Beaton, A.E. ; Plomp, T. (dir. publ.). 2000. *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver : Pacific Educational Press.
- Rojas, C.C. ; Esquivel, J.M. 1998. *Los sistemas de medición del logro académico en Latinoamérica*. Washington DC : Banque mondiale.
- Ross, K.N. (dir. publ.) *et al.* 2000. « Translating educational assessment findings into educational policy and reform measures: lessons from the SACMEQ initiative in Africa ». Document présenté au Forum mondial sur l'éducation, Dakar, Sénégal, 26-28 avril.
- Ross, K.N. ; Mählck, L. (dir. publ.). 1990. *Planning the quality of education. The collection and use of data for informed decision-making*. Paris : UNESCO/Institut international de planification de l'éducation.
- Scheurich, J.J. ; Skrla, L. ; Johnson, J.E. 2000. « Thinking carefully about equity and accountability ». Dans : *Phi Delta Kappan*, 82, p. 293-299.
- Shore, A. ; Pedulla, J. ; Clarke, M. 2001. *The building blocks of state testing programs*. Chestnut Hill MA : National Board on Educational Testing and Public Policy, Boston College.

- Torrance, H. 1997. « Assessment, accountability and standards: using assessment to control the reform of schooling ». Dans : Halsey, A.H. *et al.* (dir. publ.), *Education, culture, economy and society*. Oxford : Oxford University Press.
- UNESCO. 2000. *Cadre d'action de Dakar. L'Éducation pour tous : tenir nos engagements collectifs*. Texte adopté par le Forum mondial sur l'éducation, Dakar, Sénégal, 26-28 avril. Paris : UNESCO.

Publications et documents de l'IPE

Plus de 1 200 ouvrages sur la planification de l'éducation ont été publiés par l'Institut international de planification de l'éducation. Un catalogue détaillé est disponible ; il présente les sujets suivants :

Planification de l'éducation

Généralité- contexte du développement

Administration et gestion de l'éducation

Décentralisation - participation - enseignement à distance - carte scolaire - enseignants

Économie de l'éducation

Coûts et financement - emploi - coopération internationale

Qualité de l'éducation

Évaluation - innovations - inspection

Différents niveaux d'éducation formelle

De l'enseignement primaire au supérieur

Stratégies alternatives pour l'éducation

Éducation permanente - éducation non formelle - groupes défavorisés - éducation des filles

Pour obtenir le catalogue, s'adresser à :

IPE, Diffusion des publications (information@iiep.unesco.org).

Les titres et les résumés des nouvelles publications peuvent être consultés sur le site web de l'IPE, à l'adresse suivante :

<http://www.unesco.org/iiep/>

L'Institut international de planification de l'éducation

L'Institut international de planification de l'éducation (IIEP) est un centre international, créé par l'UNESCO en 1963, pour la formation et la recherche dans le domaine de la planification de l'éducation. Le financement de l'Institut est assuré par l'UNESCO et les contributions volontaires des États membres. Au cours des dernières années, l'Institut a reçu des contributions volontaires des États membres suivants : Allemagne, Danemark, Inde, Irlande, Islande, Norvège, Suède et Suisse.

L'Institut a pour but de contribuer au développement de l'éducation à travers le monde par l'accroissement aussi bien des connaissances que du nombre d'experts compétents en matière de planification de l'éducation. Pour atteindre ce but, l'Institut apporte sa collaboration aux organisations dans les États membres qui s'intéressent à cet aspect de la formation et de la recherche. Le Conseil d'administration de l'IIEP, qui donne son accord au programme et au budget de l'Institut, se compose d'un maximum de huit membres élus et de quatre membres désignés par l'Organisation des Nations Unies et par certains de ses institutions et instituts spécialisés.

Président :

Dato' Asiah bt. Abu Samah (Malaisie)
Directrice, Lang Education, Kuala Lumpur, Malaisie.

Membres désignés :

Pekka Aro
Directeur, Département de l'amélioration des compétences, Bureau international du travail (BIT), Genève, Suisse.

Eduardo A. Doryan
Représentant spécial de la Banque mondiale aux Nations Unies, New York, États-Unis d'Amérique.

Carlos Fortín
Secrétaire-général adjoint, Conférence des Nations Unies sur le commerce et le développement (CNUCED), Genève, Suisse.

Edgar Ortégón
Directeur, Division de la programmation des projets et investissements, Institut de la planification économique et sociale (ILPES) pour l'Amérique latine et les Caraïbes, Santiago, Chili.

Membres élus :

José Joaquín Brunner (Chili)
Directeur, Programme d'Éducation, Fundación Chile, Santiago, Chili.

Klaus Höffner (Allemagne)
Professeur, Université Libre de Berlin, Berlin, Allemagne.

Zeineb Faïza Kefi (Tunisie)
Ambassadrice extraordinaire et plénipotentiaire de Tunisie en France, Déléguée permanente de Tunisie auprès de l'UNESCO.

Philippe Mehaut (France)
Directeur adjoint, Centre d'études et de recherches sur les qualifications, Marseille, France.

Teboho Moja (Afrique du Sud)
Professeur de l'enseignement supérieur, Université de New York, New York, États-Unis d'Amérique.

Teiichi Sato (Japon)
Conseiller spécial auprès du Ministre de l'Éducation, des Sports, des Sciences et de la Technologie, Tokyo, Japon.

Tuomas Takala (Finlande)
Professeur, Université de Tampere, Tampere, Finlande.

Pour obtenir des renseignements sur l'Institut, s'adresser au :
Secrétariat du Directeur, Institut international de planification de l'éducation,
7-9, rue Eugène Delacroix, 75116 Paris, France.