

Un aperçu diachronique

Norberto Bottani

Améliorer les performances des systèmes d'enseignement et réduire le taux des échecs, c'est l'objectif des enquêtes menées par des organismes internationaux depuis désormais plusieurs décennies.

L'ÉVALUATION DES ÉLÈVES : UNE PRATIQUE QUOTIDIENNE ET UNIVERSELLE

L'une des pratiques les plus universelles de tous les systèmes d'enseignement est l'évaluation des élèves : évaluation d'une dictée, évaluation de la solution de problèmes mathématiques, évaluation d'un calcul, évaluation d'un dessin, évaluation de la récitation d'une poésie par cœur. Depuis toujours les enseignants évaluent les élèves, attribuent des notes à leurs travaux, décident s'il faut les faire passer à la classe suivante ou leur faire redoubler la classe, selon des procédures plutôt arbitraires voire très subjectives. On peut estimer que l'évaluation des élèves dans les classes est une composante caractéristique du profil professionnel des enseignants. Cependant, depuis longtemps, on se rend compte que l'évaluation des élèves n'est pas une pratique appliquée objectivement et correctement. On a maintes fois établi qu'un même devoir, évalué par des enseignants différents, est apprécié avec des notes différentes.

L'ÉVALUATION DES SYSTÈMES D'ENSEIGNEMENT

Nous ne voulons pas parler ici de l'évaluation des élèves qui a été passablement étudiée jusqu'au point de donner lieu à une science spécifique, la docimologie. Très tôt, au début du XX siècle, des voix se sont élevées pour

contester les évaluations des enseignants, démontrer leur nature arbitraire et trouver des méthodes qui n'exposent pas les élèves à des appréciations injustes.

On n'abordera donc pas ici ces questions, mais un volet différent, celui de l'évaluation des systèmes d'enseignement pour en apprécier la qualité, l'efficacité et les coûts. Ces opérations évaluent les acquis d'un grand nombre d'élèves qui fréquentent un même système d'enseignement et qui sont testés au même moment selon une procédure contrôlée, soustraite à l'ingérence des enseignants.

LES DÉBUTS DE L'ÉVALUATION DES SYSTÈMES D'ENSEIGNEMENT

L'évaluation des systèmes d'enseignement a surgi au début des années trente lorsque plusieurs pédagogues se sont rendu compte qu'il fallait inventer des structures d'enseignement nouvelles pour passer d'un enseignement secondaire réservé à une petite élite à un enseignement secondaire de masse. Ce passage a débuté avant la deuxième guerre mondiale, tout d'abord aux États-Unis, et sa mise en œuvre a été interrompue dans la plupart des systèmes européens d'enseignement quand la guerre a éclaté, mais elle a été reprise tout de suite après, au cours des années cinquante. La création de l'école moyenne unique au début des années soixante en Italie,



école qu'on appelle maintenant *enseignement secondaire de premier degré*, s'insère parfaitement dans ce mouvement.

LE PASSAGE DE L'ÉDUCATION PRIMAIRE POUR TOUS À L'ÉDUCATION SECONDAIRE DE MASSE

Les responsables politiques et les acteurs de l'enseignement, confrontés à la nécessité de généraliser l'enseignement secondaire du premier degré, étaient conscients qu'ils ne pouvaient pas appliquer à l'enseignement de masse le modèle qui fonctionnait pour une petite élite d'élèves. Sous l'impulsion des milieux scientifiques américains, et plus particulièrement des centres de recherche en éducation de New York, on a commencé à s'interroger sur le plan international, des deux côtés de l'Atlantique, s'il existait un modèle plus performant qu'un autre pour réaliser la généralisation de l'enseignement secondaire du premier degré. En d'autres termes, il fallait inventer un modèle permettant de prolonger la scolarisation jusqu'à 13-14 ans sans massacrer la population scolaire pour laquelle on estimait qu'il fallait un niveau d'instruction supérieur à celui visé par l'enseignement élémentaire. De là est née la nécessité de mieux connaître ce qui se pratiquait dans les systèmes d'enseignement les plus développés de l'époque et de comparer leurs résultats. La manière d'évaluer les élèves constituait un volet de

cette recherche, mais ce n'était pas le seul : les programmes d'enseignement, la formation des enseignants, l'organisation des classes et des établissements, la manière d'enseigner, les ressources allouées aux écoles pour faire face à ce défi étaient autant de thèmes qu'il fallait connaître, analyser et évaluer, en mesurant leurs effets sur les acquis des élèves.

À l'époque, on était convaincus que l'évaluation des systèmes d'enseignement devait amener à identifier la structure la meilleure et la plus performante pour atteindre l'objectif de la généralisation de l'enseignement secondaire. Aujourd'hui, on n'en est plus aussi sûrs.

UNE EXPÉRIENCE D'AVANT-GARDE

La première expérience connue d'évaluation des performances d'un système d'enseignement a été réalisée en 1932 en Écosse¹. Le département écossais de l'éducation a mis en œuvre une évaluation des connaissances des élèves à la fin de l'enseignement primaire pour connaître l'efficacité de l'école élémentaire écossaise. Pour ce faire, le département a organisé une évaluation sur vaste échelle comprenant tous les élèves nés en 1921 et qui en 1932 avaient donc onze ans. Environ 100 000 élèves de l'école publique et privée ont participé à l'enquête. Le test a été soumis le même jour, le mercredi 1^{er} juin 1932. Pour vaincre la méfiance et les résistances des enseignants et des autorités locales, le département a préparé le terrain en engageant beaucoup d'enseignants dans la préparation du test et dans l'organisation de l'enquête. Ce travail est exemplaire, car aujourd'hui, dans la plupart des cas, on ne prête plus attention à ces aspects et les hiérarchies imposent aux enseignants les tests qu'ils doivent soumettre à leurs élèves.

L'« INTERNATIONAL EXAMINATION INQUIRY » (1931-1938)

Simultanément à cette initiative, sans toutefois qu'il y ait un lien de cause à effet entre les deux, la fondation américaine *Carnegie* décida, en 1931, de financer un projet international pour savoir comment étaient organisés les examens et la sélection des élèves dans l'enseignement secondaire de différents systèmes d'enseignement. À l'orée de la généralisation de l'enseignement secondaire du premier degré il fallait s'enquérir sur les modalités adoptées pour évaluer et sélectionner les élèves. Le projet se déroula entre 1931 et 1938. L'initiative ayant été prise par une fondation privée, la participation était sur invitation. Les systèmes d'enseignement suivants ont été invités à participer à la recherche : États-Unis, Angleterre, Écosse, France, Allemagne, Suisse, Finlande, Suède et Norvège.² Le but du projet était de comparer les méthodes d'évaluation appliquées dans différents sys-



tèmes scolaires. La question à la base de la recherche était la suivante : quelle était la méthode la meilleure ou la plus correcte d'évaluation dans un système d'enseignement de masse ? Fallait-il envisager l'élaboration de tests pour évaluer les élèves ou est-ce qu'on pouvait continuer à se servir de la pratique bien connue des examens traditionnels pratiqués dans les écoles accueillant un nombre restreint d'élèves ? Pouvait-on faire évoluer de la même manière l'ensemble d'une génération d'élèves et le faible pourcentage d'élèves sélectionnés et privilégiés qui était admis au lycée ?

À l'époque on commençait à connaître les méthodes statistiques de mesure et d'évaluation mises au point par plusieurs chercheurs. Ces travaux apportaient la preuve de la nature arbitraire des évaluations traditionnelles et recherchaient des solutions pour corriger les erreurs. Par ailleurs, on commençait à s'interroger sur les relations existant entre les performances des élèves dans les tests et la qualité du système d'enseignement, dont les résultats ne dépendaient pas uniquement de la bonne ou mauvaise volonté des élèves et des enseignants, mais aussi de la qualité de l'organisation, de la qualité des enseignants, de l'efficacité de l'administration scolaire, de la richesse du soutien aux écoles, etc.

Malheureusement, l'éclatement de la deuxième guerre mondiale en 1939 bloqua le projet qui néanmoins parvint à produire une série d'études remarquables. Il faut toutefois signaler que, pour des raisons politiques, l'Allemagne se retira du projet après quelques années et que la Suisse, qui était représentée par deux chercheurs de grande qualité et de renommée internationale comme Jean Piaget et Pierre Bovet de Genève, se retira aussitôt après, car les chercheurs helvétiques estimaient que la mise en œuvre d'un dispositif d'évaluation des systèmes d'enseignement présentait des risques considérables de manipulation, bien connus à l'époque, qu'il ne fallait pas courir à nouveau. Cependant, le projet *International Examination Inquiry* eut le mérite d'ouvrir les portes des sciences de l'éducation aux statisticiens. Pour la première fois, autour de la table de discussion ne se trouvaient pas seulement des psychopédagogues ou des docimologues, mais aussi des statisticiens.

L'ÉPOQUE CONTEMPORAINE

Dans l'après-guerre, les travaux pour faire évoluer la qualité des différents systèmes d'enseignement reprirent, cette fois-ci sous le patronage de l'*UNESCO*. Au cours des années cinquante, les directeurs des principaux instituts mondiaux de recherche en éducation, lors de leurs rencontres annuelles au sein de l'*UNESCO*, se penchèrent sur la question et décidèrent de mener une enquête pour comparer les systèmes d'enseignement. Personne ne savait à l'époque si l'opération pouvait réussir. Les travaux de préparation se déroulèrent entre 1959 et 1961.

Le tout premier test d'évaluation comparée des systèmes d'enseignement fut effectué il y a cinquante ans, en 1959 dans douze pays. Chaque système d'enseignement ayant accepté de participer à la comparaison internationale sur une base empirique devait sélectionner au hasard 1000 étudiants de treize ans auxquels faire passer un test centré sur la compréhension de la lecture, sur les mathématiques, sur les connaissances scientifiques et sur la géographie. Les résultats furent probants : malgré les nombreux défauts méthodologiques constatés après coup qui rendaient les résultats obtenus incomparables, on réussit à démontrer qu'il était possible de traiter une grande masse de données et de parvenir à des informations intéressantes sur le fonctionnement des systèmes d'enseignement.

LES OBJECTIFS DE LA PREMIÈRE ENQUÊTE DE L'IEA EN 1964

Les objectifs de l'enquête expérimentale de l'*IEA*³ (*International Association for the Evaluation of Educational Achievement - Association Internationale pour l'Évaluation de l'Enseignement*) n'étaient pas les mêmes que ceux poursuivis avant la guerre dans le projet de la *Carnegie Foundation*. Après la deuxième guerre mondiale il était acquis, parmi les militants de l'éducation pour tous, que les politiques de l'éducation devaient viser à la généralisation de l'enseignement secondaire du premier degré. Néanmoins, les oppositions à ce projet étaient nombreuses. Il fallait dès lors convaincre les décideurs politiques d'entreprendre cette réforme dans l'intérêt de la reconstruction des pays, de l'économie nationale, du développement de sociétés détruites par la guerre et de l'égalité sociale. Un long tronc unique commun pour tous les élèves était l'objectif à poursuivre, mais personne n'avait une idée précise sur comment il fallait le configurer. Dès lors on estimait que des études comparées, méthodiques, rondement menées et contrôlées auraient pu fournir des indications sur les solutions à adopter en ce qui concerne l'organisation du tronc commun, le curriculum, la formation des enseignants en prenant en compte les acquis des élèves obtenus dans des tests identiques réalisés dans des systèmes d'enseignement différents.

Pour faire cette démonstration, les directeurs de recherche en éducation et l'*Unesco* décidèrent de mettre sur pied en 1961 une organisation spécialisée, justement l'*IEA*. La première opération d'envergure de cette association scientifique fut la réalisation en 1964 d'une grande enquête internationale sur les connaissances en mathématiques⁴.

Depuis lors l'*IEA* a réalisé toute une série d'enquêtes internationales comparées permettant de connecter les acquis des élèves à des âges différents avec des modèles d'organisation de l'enseignement et des curricula diffé-

rents, dans le but de vérifier s'il existait des variables plus efficaces que d'autres pour renforcer et améliorer le niveau d'instruction d'élèves provenant de tous les milieux sociaux. Par exemple, on se demandait s'il y avait une corrélation de cause à effet entre le coût unitaire et les acquis en mathématiques, ou entre le nombre d'heures hebdomadaires ou annuelles d'enseignement dans une discipline et les résultats dans les tests. Par ailleurs, la multiplication de ces enquêtes a permis de perfectionner les méthodes de mesure, de construction des tests, de passation, de traitement de données. Il faut se rendre compte que la comparaison des résultats obtenus avec un même instrument appliqué dans des systèmes d'enseignement différents d'un point de vue des traditions administratives et culturelles n'allait pas de soi. C'est en répétant ces enquêtes qu'on a appris à maîtriser les problèmes, à mieux comprendre les résultats et à identifier les aspects problématiques à mesurer et à observer pour améliorer les performances des systèmes d'enseignement et pour réduire le taux des échecs.⁵

L'IEA n'a pas réussi à résoudre tous les problèmes méthodologiques posés par ce défi et, surtout, n'a pas réussi à fournir aux décideurs politiques des indications claires sur les réformes scolaires à réaliser, en donnant la garantie qu'elles auraient permis d'atteindre les résultats espérés. Néanmoins, on peut dire que l'IEA a été un laboratoire expérimental à l'échelle mondiale dans le domaine de l'éducation. C'est grâce à ses travaux que l'OCDE (*Organisation pour la Coopération et le Développement Économique*) a pu lancer en 2000 le cycle d'enquêtes internationales connu sous l'acronyme PISA (*Programme International pour le Suivi des Acquis des élèves*).

Notes

¹ Le *Scottish Mental Survey*.

² L'Italie fasciste n'avait pas été invitée.

³ En français l'AIE.

⁴ Cette enquête est connue sous l'acronyme FIMS (*First Mathematics and Science Study*). L'Italie n'a pas participé à cette enquête.

⁵ Étant entendu que le redoublement n'est pas une solution, comme le prouvent une quantité de recherches.

Norberto Bottani - Consultant indépendant sur les politiques de l'éducation, Paris (norberto.bottani@oxydiane.net; www.oxydiane.net).



Un ouvrage pour l'école secondaire du premier degré du Val d'Aoste

Le texte que nous avons conçu vise à aider les enseignants qui sont engagés dans l'élaboration des épreuves de langue française à l'examen d'État de fin du premier cycle à harmoniser le format et l'évaluation des épreuves avec le *Cadre Européen Commun de Référence pour les Langues : Apprendre, Enseigner, Évaluer* du Conseil de l'Europe. Ce travail s'inscrit dans la lignée du projet *Continuità*, démarré en 2004, qui se proposait d'amorcer une comparaison, horizontale et verticale, entre différents établissements scolaires, sur l'évaluation des compétences en langues (italien, français et anglais) et en maths, à la fin du premier cycle.

Son objectif est d'encourager la réflexion et de faire évoluer la pratique en fonction du contexte, du temps et des moyens dont les enseignants disposent. Les remarques qui l'accompagnent vont dans le sens d'aider cette réflexion et de permettre aux concepteurs des épreuves d'ancrer leur travail sur des principes clairs et des démarches cohérentes et partagées qui devraient déboucher sur une évaluation transparente. Il ne se propose aucun but normatif, ni prescriptif. C'est une contribution à une réflexion individuelle et collective : ce qui compte, c'est que les enseignants impliqués dans le processus d'évaluation des compétences des élèves soient conscients des variables que l'on peut manipuler lors de ce processus et prennent des décisions de manière claire et rationnelle.

Ce travail d'approfondissement sur le format et les critères d'élaboration et d'évaluation des épreuves à la fin du 1^{er} cycle d'instruction s'avère essentiel en considération du contexte institutionnel dessiné par l'autonomie scolaire et il s'inscrit entièrement dans le mandat confié par le Gouvernement régional à la *Structure Régionale pour l'Évaluation du système scolaire de la Vallée d'Aoste (SREV)*.

L'analyse qui est faite récupère et s'approprie les connaissances fondamentales en matière d'évaluation qui ont fait l'objet du *CECR*, les différentes éditions du *Projet OCDE-PISA* et, plus particulièrement, les recommandations issues du *Profil de la politique linguistique éducative de la Vallée d'Aoste*.

W. Tonetta, G. Vernetto, *Les épreuves de français à l'examen d'État à la fin de l'école secondaire du 1^{er} degré, SREV Unité 2 – Résultats de l'apprentissage*, Le Château, 2009.

Wilma Tonetta - Chargée de mission pour la dimension bi-plurilingue de l'enseignement - *USAS (Ufficio Supporto all'Autonomia Scolastica)* de la Vallée d'Aoste.

Gabriella Vernetto - Chargée de mission pour la dimension européenne de l'enseignement - *USAS* de la Vallée d'Aoste.