

"Méthodologie d'évaluation et mesure."

Conclusions personnalisées de Jean CARDINET
(IRDP Neuchâtel)

"Méthodologie d'évaluation et mesure."

1. Regard en arrière

J'ai beaucoup apprécié le titre original qui m'avait été proposé pour cette intervention de clôture : «Conclusion personnalisée». Cela m'évite de prétendre à une synthèse d'apports beaucoup trop divers pour qu'on puisse les rassembler dans un cadre unique. Cela me donne aussi un peu moins mauvaise conscience de présenter en fait mes propres idées, en semblant citer celles des autres ...

Ce qui m'a frappé, dans l'ensemble de ces journées, c'est «l'élargissement» de l'évaluation (pour reprendre une expression dont je faisais un slogan, il y a une dizaine d'années). Plus exactement, j'ai été surpris de l'évolution du sens donné à ce terme, avec la transformation d'une problématique de mesure en une problématique de recherche de sens.

Il y a 10 ans, le mot «évaluation» était encore peu connu. Chacun lui donnait son sens étymologique : attribution d'une valeur. Pour Noizet et Caverni (1978), c'était essentiellement un processus de comparaison entre une production et une échelle de référence subjective. Les spécialistes discutaient de l'isomorphisme existant ou non entre les objets évalués et les propriétés des nombres, pour permettre l'affectation de ces objets à des classes d'équivalence (soit un niveau d'une échelle quantitative, soit une catégorie d'une échelle nominale).

C'est un peu dans cette perspective que je conclusais en 1978 le Colloque de Genève (Allal, Cardinet, Perrenoud, 1979) en proposant la théorie de la généralisabilité comme cadre de référence pour situer les recherches de développement qui restaient à faire. Le problème paraissait toujours de porter un jugement, qu'il soit fondé sur une mesure ou sur une catégorisation.

Aujourd'hui, au contraire, l'évaluation est pour moi plutôt une recherche de feedback, c'est-à-dire une démarche de prise d'information multiréférenciée sur une situation, ou le résultat d'une action. Je sais que d'autres chercheurs voudraient même éviter la référence systémique du terme de feedback. De toute façon, on peut tenir pour acquis cet élargissement de sens. Il suffit d'examiner les thèmes discutés cette année à Bruxelles pour le démontrer.

Nous verrons successivement les objets évalués, les problèmes soulevés à leur sujet, c'est-à-dire les disciplines de référence, et les méthodes essayées, ou envisagées, pour faciliter cette évaluation.

2. Les objets évalués

Evaluation scolaire

Très peu de communications portaient sur la mesure du résultat de l'apprentissage. C'était pourtant le thème central de la docimologie, il y a quelques années, et qui oserait affirmer aujourd'hui que la construction des épreuves de connaissances ne pose plus de problèmes ?

J'ai pu rattacher à ce domaine l'exposé de Marie-Jeanne Kuczynska et Colette Lecocq sur la mesure des compétences verbales, celui de Clément Dassa sur l'évaluation des acquis en français, et celui d'Antoine Bodin, où il s'agissait de vérifier le savoir des élèves par domaines des mathématiques. La méthodologie de mesure était aussi le thème de Mulondwa Misenga, qui comparait divers indices de fidélité et de généralisabilité.

Pour la plupart des autres chercheurs, d'autres visées étaient premières, même si une mesure de compétence était souvent impliquée dans leur projet.

Evaluation de la formation

Le contrôle de la démarche d'enseignement s'est révélé au Colloque une problématique beaucoup plus fréquente que la précédente. Je m'en réjouis puisqu'à Namur, en 1983, je demandais d'évaluer les conditions d'apprentissage plutôt que les élèves.

Jean-Emile Charlier, Aletta Grisay et Marie Duru-Bellat ont évalué, selon des approches méthodologiques différentes, mais toutes les trois fructueuses, l'efficacité ou le succès des établissements scolaires. Jacqueline Beckers a suivi le développement d'un projet pilote de formation en alternance. Jacques Colomb a rendu compte du fonctionnement du système didactique français, lors du passage du primaire au secondaire. Marie-Jeanne Kuczynska a étudié les causes du dysfonctionnement d'une école d'ingénieurs. Anne-Marie Garçon et Robert Prat ont cherché à évaluer une formation à l'évaluation.

Dans chaque cas, ce sont les déterminants généraux et communs, par opposition aux facteurs personnels, qu'il s'agissait de mettre en évidence.

Evaluation des moyens d'évaluation

Prendre une certaine distance par rapport à la démarche qu'on utilise, pour essayer de l'améliorer, est essentiel, particulièrement dans un colloque consacré à la méthodologie. Si la mise au point d'un instrument de mesure a rarement été présentée (elle aurait ramené à la problématique docimologique), d'autres moyens d'évaluation ont fait l'objet de communications.

Plusieurs exposés concernaient la recherche de nouvelles sources d'information. Linda Allal a traité de l'auto-évaluation des étudiants, et Jacqueline Marsenach de l'observation de la démarche d'évaluation elle-même, chez les professeurs d'éducation physique et sportive. Concetta Biasin a proposé l'utilisation d'un jeu pour analyser les stratégies d'évaluation des enseignants.

La réflexion sur les moyens utilisés pouvait aussi prendre la forme d'une recherche d'utilisations multiples, ou nouvelles, pour un même instrument : c'est ce que proposaient Nicole Dechanxhe et Alain Lammé pour l'exploitation des résultats d'une épreuve commune. Dagmar Hexel et Daniel Bain ont cherché à analyser ces fonctions multiples, mais souvent conflictuelles, des épreuves communes genevoises.

Evaluation de modèles théoriques

Un dernier domaine a été traité, plus rarement il est vrai, mais son apparition est particulièrement significative de l'évaluation annoncée. Le fait que plusieurs chercheurs aient cherché à vérifier des théories montre bien l'élargissement apporté au concept d'évaluation.

Bernadette Noël a proposé la validation d'un modèle métacognitif, François Tochon celle d'un modèle didactique, relatif à la planification des leçons. Bernard Plaitin cherchait à vérifier dans sa pratique d'enseignant la viabilité des propositions des chercheurs. Dany Laveault a voulu tester un modèle de communication des objectifs à une classe, qui devait renforcer l'anticipation, l'auto-évaluation et la préparation des étudiants.

En conclusion de cette analyse des thèmes des communications au Colloque, on voit que l'évaluation-mesure, à elle-seule, ne peut absolument pas rendre compte des évaluations auxquelles songeaient chacun des intervenants. La plupart s'adressent à des concepts théoriques. Leurs recherches constituent des validations de constructs. Leur but est d'insérer les objets étudiés dans un réseau de relations théoriques, qui en démontrent la pertinence et les chargent de sens.

3. Disciplines de référence

La même évolution apparaît, par rapport aux premières rencontres sur l'évaluation, si l'on examine le point de vue à partir duquel sont abordés les problèmes d'évaluation.

Alors que la thématique psychométrique (comment construire des épreuves adaptées aux diverses finalités de l'évaluation ?) avait motivé la réunion des chercheurs suisses et belges autour d'Yvan Tourneur en 1977, on ne retrouve plus, onze ans plus tard, qu'une minorité de travaux centrés sur les propriétés métriques des instruments.

Des mesures continuent à être recueillies, mais leurs résultats viennent directement contribuer à l'étude de problèmes psychopédagogiques, ou même de psychologie cognitive ou de psychologie sociale, voire de sociologie ou de psychologie des organisations. Il suffirait de reprendre les principaux thèmes annoncés pour le Colloque pour s'en convaincre.

4. Méthodes proposées.

Certaines propositions méthodologiques évoquées au Colloque sont originales, en tout cas dans ce contexte de l'évaluation. Elles débordent largement les techniques habituelles utilisées pour l'estimation de traits personnels.

Ainsi Antoine Bodin propose d'utiliser l'analyse implicative (de Régis Gras) pour l'étude conjointe de séries de questions, en vue d'y découvrir des hiérarchies, ou configurations de réponses. Aletta Grisay veut aussi étudier des configurations d'indicateurs, en précisant qu'il s'agit de rendre compte de relations fonctionnelles, c'est-à-dire de trouver quelles conditions d'enseignement sont favorables à l'apprentissage, quand elles se trouvent réunies. Jacques Colomb cherche aussi une méthode qui mettrait en relation des données très diverses, pour faire apparaître des relations interprétables. Citons encore l'utilisation suggérée de fonctions discriminantes pour distinguer les savoir-faire critique; qui caractérisent les experts par rapport aux novices (Clément Dassa et Gilles Tremblay), ou encore l'analyse de séries temporelles, pour démontrer l'effet d'une intervention (Jacqueline Beckers).

Il est clair que toutes ces méthodes débordent largement le cadre de pensée de la théorie de la généralisabilité que je proposais il y a dix ans. Plutôt que de chercher à estimer plus valablement un paramètre particulier, ces procédures visent à mettre en évidence une relation fonctionnelle, à parvenir à une interprétation ou à une explication.

(Qu'il me soit permis néanmoins de remarquer que la théorie de la généralisabilité, elle aussi, a pris des dimensions imprévues à l'époque. Elle peut servir à contrôler qu'un effet attendu existe effectivement, et même que son influence dépasse certains seuils d'utilité pratique. Par ailleurs, elle peut aider à planifier et choisir des plans d'observation, ou d'expérience ... Mais il est certain qu'elle n'a pas réponse à tout, et notamment pas aux questions précédentes).

5. Conclusion

L'analyse que je viens de faire des questions soulevées au sein des ateliers a été naturellement biaisée par le thème même du Colloque, où les participants devaient traiter de la méthodologie de la recherche en évaluation. Tout avait été fait pour décourager la présentation de travaux qui appliqueraient simplement des méthodes déjà parfaitement rodées, comme c'est le cas dans la construction des tests étalonnés.

Cependant, si ce Colloque a suscité tant d'intérêt et de participation, ce n'est pas seulement à cause de sa parfaite organisation, fruit du remarquable travail de préparation de Pascal Simon à l'Université Libre de Bruxelles et de son Comité scientifique inter-universitaire. C'est que la préoccupation existait partout dans les pays francophones de développer des conceptions nouvelles de l'évaluation, et de fertiliser de façon croisée des champs nouveaux, à partir de concepts et de méthodes venus d'ailleurs.

Il est donc possible de conclure que l'évaluation, telle la nostalgie, n'est plus ce qu'elle était ... mais qu'elle s'élargit à la dimension des problèmes que rencontre aujourd'hui l'évaluation. Il aurait été intéressant d'effectuer des comparaisons non seulement dans le temps, mais dans l'espace, et de situer les conceptions actuelles de l'évaluation dans les pays anglo-saxons, face aux conceptions européennes.

On aurait remarqué alors une étrange disparité entre évaluation des élèves et évaluation des innovations. Le premier domaine est encore dominé, dans les travaux de langue anglaise, par la conception d'une démarche instrumentée, préoccupée avant tout d'objectivité, où les

enseignants n'ont d'autre choix que de fabriquer leurs tests eux-mêmes, s'ils ne trouvent pas à leur goût des méthodes qu'on leur propose.

Pour l'évaluation des innovations, au contraire, la rigidité et l'irréalisme des plans expérimentaux a été dénoncée déjà depuis des décennies par les anglo-saxons et les approches qualitatives les plus diverses ont été introduites pour prendre en compte toutes les dimensions de la réalité sociale et apprécier l'influence de la subjectivité des acteurs, afin de mieux éclairer le contexte des décisions à prendre.

Ces deux champs d'application (individuel et institutionnel) de l'évaluation, aussi distincts soient-ils dans leurs instruments et dans leurs méthodes, ne sont pourtant pas totalement étrangers l'un à l'autre. Une sorte de démarche commune pourrait être définie, à partir d'une analyse détaillée de ce qu'implique le concept d'évaluation qui les réunit.

On verrait alors que l'élargissement que chaque région linguistique a introduit dans un champ particulier de l'évaluation serait souhaitable aussi dans l'autre. De part et d'autre, il faut dépasser cette idée que seule la mesure des résultats absolument objectifs peut fonder une évaluation utilisable. De part et d'autre, en effet, le critère de la vérité ne peut être qu'un certain degré d'accord intersubjectif.

D'après le contenu des ateliers de la rencontre de Bruxelles, il apparaît que l'ADMEE-Europe se situe clairement du côté de la conception élargie de l'évaluation. On peut donc trouver d'ores et déjà dans cette orientation la confirmation du bien-fondé de la décision de Dijon : les membres fondateurs de l'association ont eu raison d'introduire dans le sigle de l'ADMEE, qui nous venait de nos amis canadiens, les termes de «Méthodologies de l'évaluation», en lieu et place de «Mesure et évaluation».

C'est bien de méthodologies très ouvertes que nous avons discuté à Bruxelles et c'est sans aucun doute de leur renouvellement que nous devons débattre encore pendant longtemps.

REFERENCES

- Allal, L., Cardinet, J. & Perrenoud, Ph. L'évaluation formative dans un enseignement différencié. Berne : Peter Lang, 1979, 223 p.
- Noizet, G. & Gaverni, J.P. Psychologie de l'évaluation scolaire. Paris : P.U.F., 1978, 232 p.