

Mesure sans démesure: la contribution de Jean Cardinet aux méthodologies de la mesure et de l'évaluation en éducation

Dany Laveault

Université d'Ottawa

dlaveault@uottawa.ca

Résumé

Cet article dresse un bilan de l'œuvre de Jean Cardinet autour de deux grands principes: le besoin de résultats fidèles et valides pour informer et fonder la recherche en éducation et l'utilisation appropriée de cette information afin de soutenir les prises de décisions à tous les niveaux du système éducatif. L'article conclut en démontrant comment l'œuvre de Jean Cardinet fournit des pistes pour traiter la problématique et les défis d'aujourd'hui.

Mots clés

Théories de la mesure, fidélité, généralisabilité, validité de conséquence, standards, redevabilité.

Pour citer cet article : Laveault, D. (2015). Mesure sans démesure: la contribution de Jean Cardinet aux méthodologies de la mesure et de l'évaluation en éducation. *Evaluer. Journal international de Recherche en Education et Formation*, 1(2), pp. 129-138.

Cet article a initialement été publié dans la revue « Mesure et évaluation en éducation » en 2008 à la suite d'un colloque organisé en janvier de cette même année à l'occasion des 80 ans de Jean Cardinet. Il est reproduit ici avec l'aimable autorisation de son auteur et du directeur de la revue, Eric Dionne. Les références originales de l'article sont : Laveault, D. (2008). Mesure sans démesure : la contribution de Jean Cardinet aux méthodologies de la mesure et de l'évaluation en éducation. *Mesure et évaluation en éducation*, 31(2), pp. 5-17.

1. Introduction

La pertinence de l'œuvre de Jean Cardinet se confirme à chaque fois qu'un auteur fait référence à ses écrits, s'en inspire et s'en sert pour poursuivre l'avancement de la mesure et de l'évaluation en éducation. Que ce soit en anglais ou en français, l'indicateur du nombre de citations (*citation index*) de Jean Cardinet atteste du rayonnement de ses idées. Il n'est pas facile, dans de telles circonstances, d'ajouter à ces nombreux témoignages un nouveau compte rendu de ses publications, surtout qu'un livre hommage publié en 1990 en avait déjà établi un bilan (Institut romand de recherches et de documentation pédagogiques (IRDPP), 1990). Depuis cette date, le corpus de ses publications s'est encore enrichi.

L'objectif de cet article est de rechercher, dans l'œuvre de Jean Cardinet, ce qui permet de répondre à la question posée par le thème du colloque organisé en janvier 2008, à l'occasion de ses 80 ans : « Méthodologie de la mesure et de l'évaluation des élèves : où en est-on aujourd'hui? ». C'est ce qui explique en partie le choix des références. Il a également été guidé par les conseils de ses proches collaborateurs qui ont bien voulu attirer mon attention sur les aspects de son travail qui leur semblaient les plus pertinents.

Cette approche de l'œuvre de Jean Cardinet, toute qualitative et en subjectivité, rappelle l'échantillonnage « boule de neige » (*snowball sampling*, Fraenkel & Wallen, 2006) où une piste d'information conduit à une autre et ainsi de suite jusqu'à saturation de l'information. Elle m'aura permis d'organiser cet article autour de deux préoccupations qui ressortent de l'ensemble de son œuvre: le besoin de résultats fidèles et valides pour *informer* et fonder la recherche en éducation et l'utilisation appropriée de cette information afin de *soutenir* les prises de décisions à tous les niveaux du système éducatif. Même s'il n'a jamais employé lui-même le terme de « validité de conséquence » (Messick, 1988 ; Reckase, 1998 ; Shepard, 1997), plusieurs des écrits de Cardinet en reflètent l'esprit (Cardinet, 1977/1986a, p. 109 ; 1979/1990c, pp. 115, 133 ; 1987/1990d, p. 226). En tentant de répondre à la question posée par le thème du colloque, je chercherai à clarifier les questionnements et les propositions de Jean Cardinet en rapport avec les deux grands axes précités. Je chercherai également à illustrer comment son travail fournit des pistes pour traiter les problématiques et les défis d'aujourd'hui.

2. Fonction d'information de la mesure en éducation

Le développement de la mesure en éducation est une préoccupation constante de l'œuvre de Cardinet et la généralisabilité de la mesure en éducation y occupe une place de choix. Traditionnellement, les études de généralisabilité (études G) étaient utilisées « afin de prendre en considération tous les plans d'observation possibles utilisés pour une étude D » (étude servant à des prises de décision). Comme l'ont réaffirmé Crocker et Algina (1986), « il n'est pas possible de classer une étude G ou une étude D en se basant seulement sur le *plan d'observation* ; l'objectif poursuivi par le chercheur est déterminant » (p. 158). Une contribution majeure de Cardinet a été d'accroître le nombre de situations pour lesquelles les études G permettent de guider nos décisions. Cardinet a démontré que les études G permettent de différencier non seulement les sujets, comme Cronbach concevait les études G à l'origine (Cronbach, Gleser, Nanda & Rajaratnam, 1972), mais aussi n'importe quelles facettes de la mesure, tels les items ou les objectifs d'apprentissage (Cardinet & Allal, 1983 ; Cardinet & Tourneur, 1985 ; Cardinet, Tourneur & Allal, 1976, 1981/1982). Cardinet (1994) fait ainsi valoir toute la portée de la théorie de la généralisabilité pour mesurer les progrès des élèves :

The concept of facet slicing [...] expands even further the types of applications and optimization that are made possible using generalizability theory. New measurement designs are offered, at the same time intra-individual and based on large scale studies, promising educational benefits completely different from those of traditional psychometrics, while belonging to the same conceptual framework (p. 210).

En retour, un tel élargissement des applications des études G a permis à Cardinet d'étudier systématiquement et rigoureusement les différents plans d'observation et de mesure à même de répondre aux besoins de mesures généralisables et valides en éducation selon l'utilisation qui sera faite des résultats : évaluation sommative, évaluation diagnostique et évaluation formative (Cardinet, 1977/1986a, p. 97). Certaines de ces applications se sont avérées parfois des voies sans issues. En effet, les exigences de la mesure en éducation pour différencier adéquatement les personnes, les apprentissages et les progrès individuels peuvent être difficiles à satisfaire. Pour différencier « avec assurance » les résultats des élèves en fonction de plusieurs facettes de leurs apprentissages et atteindre des degrés acceptables d'erreur de mesure, il faut multiplier les prises d'information en augmentant le nombre d'observateurs ou d'observations à un point tel que les conditions de la mesure deviennent souvent impraticables dans le milieu de l'éducation. Lorsque de telles situations se présentent, Cardinet (1986b) se tourne alors vers des approches plus qualitatives, notamment celles qui ont trait à l'utilisation de l'évaluation formative pour venir en aide aux élèves et les soutenir dans leur apprentissage.

Le travail de Jean Cardinet témoigne de l'importance tant des méthodes qualitatives que quantitatives. « Le réel est qualitatif », souligne Cardinet (1979/1990c, p. 10). Par contre, les méthodes quantitatives ont l'avantage de permettre de mesurer l'incertitude ou l'intervalle de confiance des décisions, ce que les approches qualitatives ne permettent pas de manière aussi précise. Le travail de Jean Cardinet emprunte aux traditions épistémologiques des deux méthodologies. Dans un certain sens, il se réfère à une « méthodologie de recherche mixte » avant même que ce concept ne soit formalisé (Johnson, Onwuegbuzie & Turner, 2007). Ardoino (1990) fait bien ressortir cette dualité dans le développement de la pensée de Cardinet en matière d'évaluation :

C'est justement ce parallélisme entre l'idéal de la mesure des phénomènes, pour contribuer à leur intelligibilité, et l'intention plus délibérément praxéologique d'optimisation de l'action et d'aide à la décision qui va conduire à une première intelligence de l'évaluation cherchant, ensuite, son dépassement, par une problématique plus largement ouverte, dans la perspective d'un élargissement (p. 73).

L'élargissement du domaine de la mesure en éducation proposé par Cardinet (1979/1990c) est plus d'actualité que jamais. Cet élargissement met à l'épreuve les modèles actuels des tests. En voici quelques exemples :

- a. L'évaluation d'objets complexes, tels que les compétences, soit à travers des méthodes de *testing* traditionnelles, soit par l'intermédiaire de grilles d'appréciation de la performance, suscite de nouveaux défis. Voici un aperçu de quelques-uns de ces défis :
 - Dans un test de compréhension en lecture, les items se référant au même texte sont décrits de façon plus appropriée par la notion de *testlet* (Wainer & Kiely, 1987). Les *testlets* sont constitués de grappes d'items et se retrouvent chaque fois que ceux-ci sont reliés par un problème, un cas ou un texte commun. De telles grappes d'items posent un problème particulier de traitement statistique ; car ils ne constituent pas des observations indépendantes.

- Pour observer et coter une performance complexe, le choix entre une échelle de mesure analytique ou globale n'est pas évident. Le choix du type d'échelle et les modalités de codage ont une influence sur la fidélité et la validité des appréciations des juges (Smee, 2007).
 - L'évaluation de questions intégrant plusieurs matières scolaires ou plusieurs objectifs crée également des difficultés dans l'étude de la généralisabilité des degrés de difficulté des objectifs. Dans de telles situations, le nichage des items dans plus d'un objectif à la fois pose problème.
- b. Il existe un intérêt grandissant pour les méthodes d'établissement de standards et de seuils de réussite ainsi que pour leur validité et leur fidélité respectives. Buckendahl, Smith, Impara, et Plake (2002) ont récemment comparé une méthode d'Angoff modifiée avec la méthode des signets (*bookmark method*). Quelle est la généralisabilité des standards établis par ces différentes méthodes? Combien faut-il d'observations et de juges pour atteindre des seuils de réussite en lesquels nous pouvons avoir confiance?
- c. Les chercheurs sont conscients du besoin d'intégrer des modèles cognitifs au sein des modèles de la mesure en éducation. Plusieurs modèles de traits latents ont été développés à cet effet. La plupart d'entre eux reposent sur des avis d'experts afin de construire une matrice Q qui relie chaque item à un attribut cognitif. Jusqu'à présent, la documentation sur la modélisation cognitive ne fournit pas vraiment de pistes en ce qui concerne le nombre d'experts nécessaires à l'élaboration d'une matrice Q généralisable (Loye, 2007). Si les matrices Q ne sont pas généralisables, les attributs cognitifs risquent de ne pas être correctement reliés aux items, conduisant à des diagnostics et à des pronostics différents selon l'échantillon d'experts consultés.

Tous ces développements récents nécessitent l'input de jugements d'experts à un moment ou à un autre. Malgré leur degré élevé de sophistication et probablement en raison même de celui-ci, ces modèles font intervenir les facteurs « humain » et « subjectif ». Nous sommes bien loin des exigences des prétendus tests « objectifs » fondés sur des questions à choix multiple. Dans chacun des cas précédents, il faut avoir une idée, ne serait-ce qu'approximative, du nombre d'experts requis sous certaines conditions afin de pouvoir effectuer des jugements généralisables. Même si ces questions ont peu d'écho ou d'intérêt pour l'évaluation en salle de classe, elles prennent de plus en plus d'importance particulièrement dans le contexte d'évaluations à grandes échelles ou aux enjeux élevés.

Les défis importants auxquels doit faire face l'évaluation scolaire aujourd'hui révèlent bien les limites des modèles expérimentaux, comme des modèles de science sociales, qui ont façonné le rôle et l'utilisation de la mesure en éducation. Cardinet (1990b) démontre comment ces deux modèles doivent être complétés par un troisième modèle, qu'il nomme « l'évaluation négociée » :

L'évaluation a nécessairement, dans ce cas, des référentiels multiples, parce qu'elle prend comme cadre la réunion des référentiels individuels des acteurs. La réalité est définie alors comme la somme des faits objectifs de la première approche [modèles expérimentaux] et des représentations subjectives de la deuxième [modèles des sciences sociales], La meilleure façon d'explorer cette réalité, c'est de confronter tous les points de vue possibles et pour cela de susciter des interactions entre partenaires (p. 150).

Selon Cardinet (1990b), l'utilité de l'évaluation ne se trouve pas uniquement dans sa capacité à contrôler le système éducatif, comme en sciences expérimentales, ni seulement dans sa capacité à le comprendre et l'expliquer, comme en sciences sociales, mais plutôt dans sa capacité à produire du changement et à introduire des innovations afin de l'améliorer. La valeur de l'évaluation ne se limite pas à la qualité de l'information. Elle s'exprime aussi à travers le soutien à la prise de décisions et à l'anticipation des conséquences que celles-ci entraînent. C'est à ce stade que le travail de Cardinet quitte l'univers bien assuré de la théorie de la généralisabilité pour entrer dans celui plus instable mais non moins rigoureux de soutien à l'intervention en éducation.

3. Fonction de prise de décision

Lorsque l'on envisage de nouvelles applications de la mesure en éducation, il existe un stade à partir duquel il devient nécessaire de chercher au-delà des modèles purement métriques. Un tel déplacement vers l'intégration de modèles extérieurs à la mesure est nécessaire à l'amélioration de nos actions, que ces actions concernent la mesure elle-même, sa communication ou son application. Cardinet a contribué à enrichir le champ de la mesure en éducation en empruntant des notions en provenance d'autres disciplines telles que la cybernétique, l'approche systémique de la recherche, la théorie piagétienne du développement et de l'apprentissage, la psychologie sociale, l'épistémologie et l'éthique, pour n'en citer que quelques-unes.

« Objectifs éducatifs et évaluation individualisée » fournit un bel exemple d'intégration de plusieurs disciplines aux modèles de mesure. En utilisant un modèle cybernétique, Cardinet (1977/1986a) dresse un parallèle entre le lancement d'une fusée vers Mars et les différentes *boucles de rétroaction et les régulations* qui se produisent à différentes étapes d'un système d'évaluation de l'éducation. L'évaluation de l'éducation implique une suite d'opérations complexes qui nécessitent d'être réalisées à temps et au moment opportun afin que la mission puisse réussir. Cardinet (1977/1986a) identifie trois objectifs principaux à la mission de l'éducation: améliorer les conditions d'apprentissage, améliorer les processus d'apprentissage et, enfin, améliorer les résultats d'apprentissage. La confusion entre ces objectifs est un problème récurrent qui conduit à l'utilisation inappropriée d'instruments ou de tests, ou encore à des interprétations injustifiées des résultats, parce que les « contrôles » ou « régulations » mis en place ne tiennent pas suffisamment compte des objectifs poursuivis. Utilisant un concept majeur de la cybernétique, le concept de « régulation », il l'applique à la question de l'évaluation individualisée, mettant l'accent sur la primauté de sa fonction formative. Depuis lors, le concept de régulation n'a cessé d'être développé par Allal (2007) et plusieurs autres chercheurs francophones afin de traiter d'autres fonctions ou « missions » de la mesure en éducation, notamment en ce qui concerne l'autoévaluation et les différentes formes que peut emprunter l'évaluation pour supporter l'élève dans ses apprentissages (Allal & Mother Lopez, 2005 ; Laveault, 2004, 2007).

Trente ans après la « mission pour Mars » de Jean Cardinet, que peut-on dire de la mission de la mesure en éducation? La mission actuelle de la mesure en éducation fait intervenir de nombreux acteurs et le concept d'*évaluation négociée* (Cardinet, 1990b) rend davantage compte de la responsabilité collective de tous les intervenants. Cette nouvelle mission se rapproche davantage du modèle des négociations organisées par l'ONU sur le changement climatique à Bali en 2007 que du modèle de la mission pour Mars. À plusieurs occasions, toute la question des changements climatiques et des actions à entreprendre afin de limiter leurs impacts négatifs rappelle l'ensemble des actions prises collectivement par les enseignants et les administrateurs afin de modifier le « climat scolaire » de façon à favoriser l'apprentissage

des élèves. On trouve ci-dessous certaines des conditions se trouvant dans ces deux types de négociations intervenant dans le « changement de climat » :

- Les intervenants ne s'entendent pas toujours sur la nature complexe du problème, et certaines personnes vont jusqu'à remettre en question son existence.
- Même au sein des personnes qui sont d'accord sur l'existence du problème, il peut y avoir des divergences sur les objectifs et sur le type de moyens à mettre en œuvre afin de les atteindre. Certains participants soulignent, légitimement, les déficiences des connaissances scientifiques sur lesquelles sont basées les décisions. De tels arguments sont souvent utilisés pour justifier l'inaction.
- Même au sein des personnes qui sont d'accord sur les actions à entreprendre, certains individus n'agiront pas sans qu'une pression ne les y contraigne. La question du leadership occupe une place centrale.

Ces freins à l'innovation et au changement rappellent que beaucoup de subjectivité et d'intérêts personnels sont impliqués dans de telles négociations, que celles-ci portent sur l'école ou les changements climatiques. Au lieu de les ignorer, Cardinet (1987/1990d, pp. 224-227) propose de les utiliser au mieux. Il met à profit la subjectivité inhérente à la prise de décision d'une manière qui rappelle l'art martial japonais Aikido. Le principe de l'Aikido est d'utiliser la force de l'adversaire afin de le faire vaciller et de le contrôler de manière à ce qu'il ne représente plus une menace. Dans l'évaluation négociée de Cardinet, un modèle interactionniste de gestion scolaire, le manque d'objectivité des intervenants avec leurs différents partis pris représente l'adversaire à maîtriser. Ce modèle fait le pari que, de l'ensemble des « subjectivités » partagées de tous les intervenants en éducation ressortiront des représentations et valeurs collectives dotées d'un statut quasi objectif.

Le modèle d'évaluation négociée de Cardinet anticipe quelques-uns des plus récents concepts concernant le changement en éducation. Par exemple, le modèle de « communauté d'apprentissage » de Fullan (2001) est très semblable au modèle d'évaluation négociée de Cardinet. On trouve ci-après un résumé des caractéristiques principales du modèle de Fullan, telles que relevées par Hammonds (2002) :

The key to successful change is the improvement in relationships between all involved and not simply the imposition of top down reform. The new emphasis in educational change is based on creating the conditions to develop the 'capacity' of both organizations' and individuals to learn [en ligne].

4. Conclusion: Où en sommes-nous aujourd'hui?

Cardinet (1979/1990c) appelait à un élargissement du domaine de la mesure en éducation. Il a dégagé par la suite de nouvelles directions et a également signalé les terrains minés. Sans aucun doute, le champ de la mesure en éducation a fait depuis d'énormes progrès. La théorie de la généralisabilité, les modèles de réponses aux items ainsi que les autres articles de ce numéro spécial de la revue *Mesure et évaluation en éducation* attestent de ces progrès. Le domaine a pris de l'assurance et s'attaque à des défis et à des problématiques que l'on n'aurait pas osé imaginer il y a de cela quelques années.

La mesure en éducation est peut-être aujourd'hui victime de son propre succès. Elle a créé de grandes attentes, pas toujours réalistes, quant à sa capacité de guider les prises de décision pour faire progresser les systèmes éducatifs à tous les niveaux. Ses retombées sont plus importantes que jamais, notamment en ce qui concerne l'exploitation des résultats des évaluations externes et tout ce qui touche la redevabilité des systèmes éducatifs. Tout ceci

fait en sorte que l'attention dont elle est l'objet est peut-être disproportionnée par rapport à son efficacité réelle. Cardinet (1977/1986a) observait que l'utilisation valide de la mesure en éducation pour améliorer nos systèmes éducatifs est tout aussi essentielle que la collecte de mesures fiables. L'une ne saurait aller sans l'autre et c'est pourquoi les attentes élevées et les retombées importantes attendues de la mesure et évaluation en éducation exigent des standards élevés, tant en ce qui concerne les résultats que l'exploitation qui en sera faite.

Les spécialistes de la mesure en éducation se doivent d'informer les utilisateurs de tests des limites de ces instruments et des conclusions qu'il est possible d'en tirer. Même en faisant l'hypothèse qu'il est possible d'obtenir des résultats de tests valides et fidèles, la validité et la justesse des décisions et des choix dépendent du respect de standards élevés de qualité non seulement par les développeurs de tests, mais aussi par ceux qui en exploiteront les résultats. Les cas de l'Ontario et des États-Unis ci-dessous ne sont que quelques exemples de conséquences non désirables de la mesure et du *testing* en éducation :

- Cas #1. L'introduction de tests à grande échelle dans les classes de troisième et de sixième en Ontario a eu pour effet que les enseignants expérimentés de ces niveaux ont migré vers d'autres degrés. Au bout de seulement quelques années, les élèves des classes de troisième et de sixième bénéficiaient d'enseignants beaucoup moins expérimentés que ceux des autres degrés. Ce facteur pourrait expliquer pourquoi, entre autres choses, les résultats des élèves se sont mis à stagner, même à décroître dans certains cas, au début de l'introduction de ces tests.
- Cas #2. Même si le processus administratif et les ressources financières nécessaires pour instaurer un programme national de tests n'ont pas mis longtemps à être mis en place après le vote de la loi en Ontario, il a fallu plusieurs années avant que les enseignants, les directions d'établissements scolaires et les administrateurs de conseils scolaires commencent à prêter attention aux résultats des tests, à se les approprier et à les utiliser avec efficacité. La communication de résultats de manière à soutenir des décisions administratives et à stimuler les échanges professionnels entre enseignants soulève des difficultés qui sont d'une importance tout aussi grande que la validité et la fidélité des résultats des tests.
- Cas #3. En dépit de mises en garde largement diffusées, des comparaisons non valides et des conclusions non fondées sont effectuées, tant par des professionnels que par des non professionnels. C'est ainsi que les journaux de l'Ontario comparent les résultats des tests administrés en français aux résultats de tests administrés en anglais, alors même qu'ils ne sont pas équivalents pour ce qui est des échelles de mesure. De même, des demandes répétées sont formulées afin d'utiliser les résultats de ces tests à des fins diagnostiques alors que la fidélité des résultats pour différentes échelles de ces tests ne justifie pas un tel usage à cause du petit nombre d'items par domaine.
- Cas #4. Aux États-Unis, les classements scolaires qui se fondent sur l'utilisation de tests ont un impact, non seulement sur les décisions concernant l'éducation, mais aussi sur le marché de l'immobilier. Les agents immobiliers utilisent fréquemment ces classements pour justifier les prix de vente des propriétés, une maison située près d'une « bonne » école justifiant un prix plus élevé (Simner, 2000).

Les défis posés par la mesure et l'évaluation sont non seulement de nature psychométrique, mais également de nature sociale. L'œuvre de Cardinet atteste de l'importance de faire progresser tant le front de l'amélioration des dispositifs de mesure que celui de l'utilisation sociale qui sera faite des résultats. C'est pourquoi il n'est pas possible de conclure ce survol sans mentionner les contributions, non seulement de l'homme de conviction, mais aussi de

l'homme d'action. L'engagement social de Jean Cardinet a servi à élever les standards de notre discipline et à plus d'un titre. Je soulignerai les deux contributions majeures suivantes :

- Jean Cardinet a contribué à la fondation de l'*International Test Commission*, une organisation dont le but est de concevoir des Standards internationaux pour les tests. Dans un article publié dans le premier numéro du *International Journal of Testing*, Oakland, Poortinga, Schlegel et Hambleton (2001) reconnaissent le rôle fondateur de Jean Cardinet dans la création de cette organisation internationale. Il en montrait déjà le besoin en 1968 lors d'une présentation à l'assemblée générale de la Société suisse de psychologie sur l'application de standards éthiques dans l'emploi des tests. De tels standards sont de plus en plus d'actualité et sont devenus essentiels pour assurer une progression continue de la discipline.
- Jean Cardinet a également joué un rôle décisif dans l'extension des réunions des chercheurs belgo-suisse, dans le développement de collaborations avec les chercheurs canadiens dans le cadre de l'ADMEE (Association pour le développement des méthodologies d'évaluation en éducation) et dans l'établissement de l'ADMEE-Europe, qui a fêté ses 21 ans en 2008. Il a également joué un rôle important au sein de la revue *Mesure et évaluation en éducation*, qui est devenue depuis l'une des rares publications francophones de rang international dans le domaine de l'éducation.

Cardinet mérite d'être reconnu pour sa contribution à l'élargissement de la mesure en éducation tout en maintenant une perspective cohérente du domaine entre ses méthodes et ses objectifs. En parlant des pédagogues, Meirieu (dans une intervention auprès du Syndicat des enseignants romands, 2007) affirmait: « Rien n'est plus subversif dans une société que quelqu'un qui oblige à la cohérence. » Il y a en effet quelque chose de subversif lorsque Cardinet (1989) pose la question suivante: « Ne vaudrait-il pas mieux admettre, une fois pour toutes, que les possibilités de mesure seront toujours radicalement insuffisantes par rapport à nos besoins et qu'il est plus sage de modifier nos attentes et notre conception de l'évaluation? (p.46) ». Au moment où les méthodologies de la mesure et de l'évaluation sont de plus en plus sollicitées pour réguler les systèmes éducatifs, cette question et l'ensemble de l'œuvre de Cardinet résonnent comme un vibrant appel à *la mesure sans la démesure*.

4. Références

- Allal, L. (2007). Régulations des apprentissages: orientations conceptuelles pour la recherche et la pratique en éducation. In L. Allal & L. Mother Lopez (éds). *Régulation des apprentissages en situation scolaire et en formation* (pp. 7-23). Bruxelles : De Boeck.
- Allal, L., & Mortier Lopez, L. (2005). L'évaluation formative de l'apprentissage: revue de publications en langue française. In OECD-CERI (éd.), *L'évaluation formative - pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Paris : Auteur.
- Ardoino, J. (1990). L'apport de Jean Cardinet à l'évaluation des systèmes éducatifs. In Institut romand de recherche et de documentation pédagogique (éd.), *Hommage à Jean Cardinet* (pp. 71-76). Neuchâtel: IRDP-Delval.
- Buckendahl, C.W., Smith, R.W., Impara, J.C., & Plake, B.S. (2002). A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational measurement*. 39(3), 253-263.
- Cardinet, J. (1977). *La coordination de l'information dans le système éducatif*. Neuchâtel : IRDP, 12 pages.
- Cardinet, J. (1977). Objectifs éducatifs et évaluation individualisée. Neuchâtel: IRDP, Rapport IRDP/R 77.05 (2^e éd.). Reproduit dans Cardinet, J. (1986a), *Évaluation scolaire et mesure* (pp. 64-112). Bruxelles : De Boeck.
- Cardinet, J. (1979). L'élargissement de l'évaluation. *Éducation et recherche*, 1(1), 15-34. Reproduit dans Cardinet, J. (1990c), *L'élargissement de l'évaluation*. In Institut romand de recherche et de documentation

- pédagogique (éd.), *Hommage à Jean Cardinet* (pp. 109-177). Neuchâtel: IRDP-Delval.
- Cardinet, J. (1986b). *Évaluation scolaire et mesure*. Bruxelles: De Boeck.
- Cardinet, J. (1987). L'évaluation en classe: mesure ou dialogue? *European Journal of Psychology of Education*, 2(2), 133-144. Reproduit dans Cardinet, J. (1990d). L'évaluation en classe: mesure ou dialogue? In Institut romand de recherche et de documentation pédagogique (éd.), *Hommage à Jean Cardinet* (pp. 215-228). Neuchâtel : IRDP - Delval.
- Cardinet, J. (1988). The generalisability of an individual's progress. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 8(4), 247- 255.
- Cardinet, J. (1989). Évaluer sans juger. *Revue française de pédagogie*, 88, 41-52.
- Cardinet, J. (1990a). *Remettre le quantitatif à sa place en évaluation scolaire*. Neuchâtel : Institut romand de recherche et de documentation pédagogique.
- Cardinet, J. (1990b). Evaluation externe, interne, ou négociée? In Institut romand de recherche et de documentation pédagogique (éd.), *Hommage à Jean Cardinet* (pp. 139-157). Neuchâtel: IRDP - Delval.
- Cardinet, J. (1994). Control of the value of an intra-subject measurement design. In D. Laveault, B. Zumbo, M. Gessaroli & M. Boss (éds). *Modern Theories of Measurement: Problems and Issues*. Ottawa: Faculty of Education, University of Ottawa.
- Cardinet, J., & Allal, L. (1983). Estimation of generalizability parameters. *New Directions for Testing and Measurement*, (18), 17-48.
- Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure*. Berne: Peter Lang.
- Cardinet, J., Tourneur, Y., & Allal, L. (1976). The symmetry of generalizability theory: Applications to educational measurement. *Journal of Educational Measurement*, 13(2), 119-135.
- Cardinet, J., Tourneur, Y., & Allal, L. (1981/1982). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement*, 18(4), 183-204. ERRATA(1982) *Journal of Educational Measurement*, 19(4), 331- 332.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: generalizability for scores and profiles*. New York: John Wiley.
- Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and evaluate research in education*. (6th ed.). Montréal: McGraw-Hill.
- Fullan, M. (2001). *The new meaning of educational change, 3rd edition*. New York: Teachers College Press.
- Hammonds, B. (janvier 2002). The latest ideas on school reform by Michael Fullan. Leading and learning for the 21st Century, 1(3). In *Quality Learning*. Document en ligne consulté le 2 juin 2008 à l'adresse suivante: [http://www.leading-learning.co.nz/newsletters/vol01-no03-2002.html].
- Institut romand de recherches et de documentation pédagogiques (IRDP). (1990). *Hommage à Jean Cardinet*. Neuchâtel : IRDP - Delval.
- Johnson, R.B., Onwuegbuzie, A.J., & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*. 1(2), 112-133.
- Laveault, D. (2004). Interactions entre formation et évaluation: de la régulation entre acteurs et leurs rôles à la régulation de l'apprentissage. *Mesure et évaluation en éducation*. 27(1), 51-67.
- Laveault, D. (2007). De la régulation au réglage: étude des dispositifs d'évaluation favorisant l'autorégulation des apprentissages. In L. Allal & L. Mottier Lopez (éds). *Régulation des apprentissages en situation scolaire et en formation* (pp. 207-234). Bruxelles : De Boeck.
- Loye, N. (2007). *Conditions d'élaboration de la matrice Q des modèles cognitifs et impact sur sa validité et sa fidélité*. Thèse de doctorat inédite : Université d'Ottawa.
- Messick, S. (1988). The once and the future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braun (éds). *Test validity* (pp. 33- 46). Hillsdale, NJ: Lawrence Erlbaum.

- Oakland, T., Poortinga, Y.H., Schlegel, J., & Hambleton, R.K. (2001). International Test Commission: Its history, current status, and future directions. *International Journal of Testing*, 1(1), 3-32.
- Reckase, M.D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13-16.
- Shepard, L.A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-24.
- Simner, M.L. (2000). *Enoncé de position conjoint de la Société canadienne de psychologie et de l'Association canadienne des psychologues scolaires relativement à la couverture, par la presse canadienne, des résultats des examens de compétences administrés à l'échelle de la province*. Groupe de travail sur l'utilisation erronée de tests. Document en ligne consulté le 2 juin 2008 à l'adresse suivante : [http://www.cpa.ca/documents/joint_position_fr.html].
- Smee, S. (2007). *Comparing scoring instruments for the performance assessment of complex competencies*. Thèse de doctorat inédite: Université d'Ottawa.
- Syndicat des enseignants romands (2007). *Morceaux choisis de l'intervention de Philippe Meirieu*. Site de L'Éducateur. Document en ligne consulté le 2 juin 2008 à l'adresse suivante: [http://www.le-ser.ch/ser/even/07_congres_SER_meirieu.html].
- Wainer, H., & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24(3), 185-201.