

VI. Psychométrie et éduimétrie

EDUMETRIE ET PSYCHOMETRIE: DEUX DISCIPLINES DIFFERENTES ?¹

Marc Demeuse

En préparant le cours de « Méthodes psychométriques et éduométriques », il a semblé utile d'envisager la distinction entre ces deux termes: s'agit-il de deux disciplines ? comment s'articulent-elles ? qu'ont-elles en propre et en particulier ? C'est surtout du côté de l'éduométrie que l'investigation a été poussée dans cette partie, la composante psychométrique ayant été largement couverte jusqu'ici dans les chapitres précédents.

Une première mention « historique » en langue française du terme « éduométrie » peut être trouvée dans l'ouvrage de Viviane de Landsheere intitulé « *Faire réussir, faire échouer. La compétence minimale et son évaluation* » publié en 1988 aux Presses universitaires de France (p. 59) : « *Edumétrie. Mot créé par Carver (1974), sur le modèle de psychométrie, pour désigner l'étude quantitative des variables relatives aux apprentissages suscités par l'éducation : influence d'une action pédagogique, performance effective par rapport à une performance attendue, épreuves centrées sur les objectifs...* ». La publication de Carver, mentionnée dans la bibliographie de l'ouvrage est un peu antérieure à 1974 : Carver, R.P. (1970). *Special Problems in Measuring Change with Psychometric Devices, Evaluative Research : Strategies and Methods*, Washington, American Institute for Research. Le texte de Carver, publié en 1974 dans *American Psychologist* (29(7), p. 512-518) est nettement plus aisé à consulter et est intitulé « *Two Dimensions of Tests. Psychometric and Edumetric* ».

Carver (1974), comme dans son texte de 1970, mais sur un ton peut-être un peu moins virulent, y argumente en faveur d'une distinction essentielle : « *Il ne s'agit pas seulement d'un problème sémantique sous-jacent au mouvement nouveau de création de tests, mais d'un problème essentiel. Le problème réside dans la distinction entre la mesure des différences individuelles et celle de la mesure (des résultats) de l'apprentissage.[...] Le test qui a été développé pour mesurer les différences inter-individuelles a été baptisé du terme de « test psychométrique ». Cette dimension du test est donc logiquement appelée « dimension psychométrique ». Un test peut ainsi être évalué en fonction de ses qualités psychométriques, c'est-à-dire, la mesure dans laquelle il reflète correctement des différences inter-individuelles stables, ce qui constitue traditionnellement l'intérêt premier du psychologue. L'autre dimension d'un test peut être appelée « dimension éduométrique » (cf. Carver, 1972, 1973). Un test peut être évalué en fonction de ses qualités éduométriques, c'est-à-dire la mesure dans laquelle il reflète les différences intra-individuelles liées aux développements des individus eux-mêmes, ce qui constitue plus généralement l'intérêt premier des tests dans le domaine de l'éducation. »*

Carver consacre la suite de son article à mettre en évidence les distinctions qui existent, tant du point de vue de la fidélité que de la validité, entre ces deux approches. Il décrit également les modalités différentes qui président à la conception de l'une ou l'autre des deux formes de tests, selon la dimension qui est privilégiée. Le texte est particulièrement didactique et évite de noyer le lecteur sous les formules. Il est traduit et reproduit dans le chapitre 3 de cette partie. Pour être complet, il existe également une série de réponses à cet article, dont une de Cronbach (1974). Mais, la question initiale persiste : faut-il utiliser deux termes

¹ Le texte de ce chapitre a été publié par l'auteur dans le *Bulletin de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation* (ADMEE-Europe) (2002/2).

« psychométrie », d'une part, « édumétrie », d'autre part, ou peut-on se contenter de l'appellation historique « psychométrie » en reconnaissant que la seconde est un domaine particulier de la première ?

A titre d'expérience, une interrogation électronique des collègues de la « *division D : Measurement and Research Methodology Forum* » de l'*American Educational Research Association* (AERA) a permis de collecter un certain nombre d'avis. En huit jours (du 8 août au 15 août 2002), 20 réponses de longueur et de qualité très variable ont ainsi été récoltées. La tendance est généralement à l'inclusion de l'édumétrie dans la psychométrie : « *In today's common usage (at least in the US), psychometrics includes the measurement of any psychological/behavioral attribute: achievement, performance, aptitudes, attitudes, opinions, values, beliefs, psychopathology, self-reports of just about anything. Thus it also includes most measurement that occurs in educational settings. So in my opinion (and that's all this is), edumetrics would be a subset of psychometrics, unless edumetrics would also include the measurement of things that are of concern to education but do not fit in the category of psychological/behavioral attributes.* » (E. Witt, Senior Psychometrician, CAT*ASI)

Certains correspondants se font plus ironiques : « *This seems to be a distinction that isn't. The educational dimension is still the psychological behavior of interest. To create "edumetrics" provides no benefit. What next - busometrics? industrometrics? orgometrics? argrometrics? medometrics? or [heaven forbid] legometrics? In the broadest sense Psychometrics involves studying the ways we measure people's behavior [and ways we analyze those measures]. Educational behaviors are still behaviors. I didn't read where the "edumetrics" provided any benefit to us or the consumer. On the contrary, it's more likely to cause confusion, even among the Div. D members!* » (S.K. Rock)

D'autres (P.R. Swank, Medical School, UT Health Science Center at Houston) précisent, à la défense de l'édumétrie, qu'il existe bien des psychométriciens et des biométriciens, alors que d'autres, beaucoup plus radicaux, déplorent cette multitude d'appellations : « *We have a discipline ... then where it is primarily applied ... so, we have statistics ... and ... applied to education??? biology??? etc ... quite frankly ... I have never heard of an educational statistic ... or a biological statistic ... is the mean or correlation coefficient different when applied to these different "content" areas?* » (D. Roberts, Pennsylvania State University).

Mais au fait, qu'en est-il dans le domaine francophone ? Devons-nous plutôt nous référer au concept de « mesure et évaluation en éducation », comme l'indique le titre de la revue de l'ADMEE¹ ? Faut-il disposer d'une appellation plus brève, facilitant la reconnaissance d'un champ, de pratiques, de professionnels... ? Naturellement, l'appellation « psychométricien » ne possède pas, dans le domaine francophone, le prestige et la reconnaissance qu'elle connaît en langue anglaise. Par ailleurs, comme le montrent les deux chapitres suivants (Cardinet et Carver), la nature même des instruments, leur conception et les critères de qualité qui leur sont appliqués, dans le domaine de l'éducation, ne correspondent pas nécessairement à la théorie classique des tests, du moins lorsqu'il ne s'agit pas uniquement de mesurer les compétences individuelles. Il est donc peut-être aussi utile de se démarquer de la théorie classique à travers un nom consacré.

¹ *Association pour le Développement des Méthodologies d'Évaluation en Éducation*. Le nom de la revue est d'ailleurs calqué sur celui de nombreux départements d'universités canadiennes.

Quelques références pour poursuivre la discussion (certaines m'ont été aimablement suggérées par William Pepper, à travers les échanges du forum AERA).

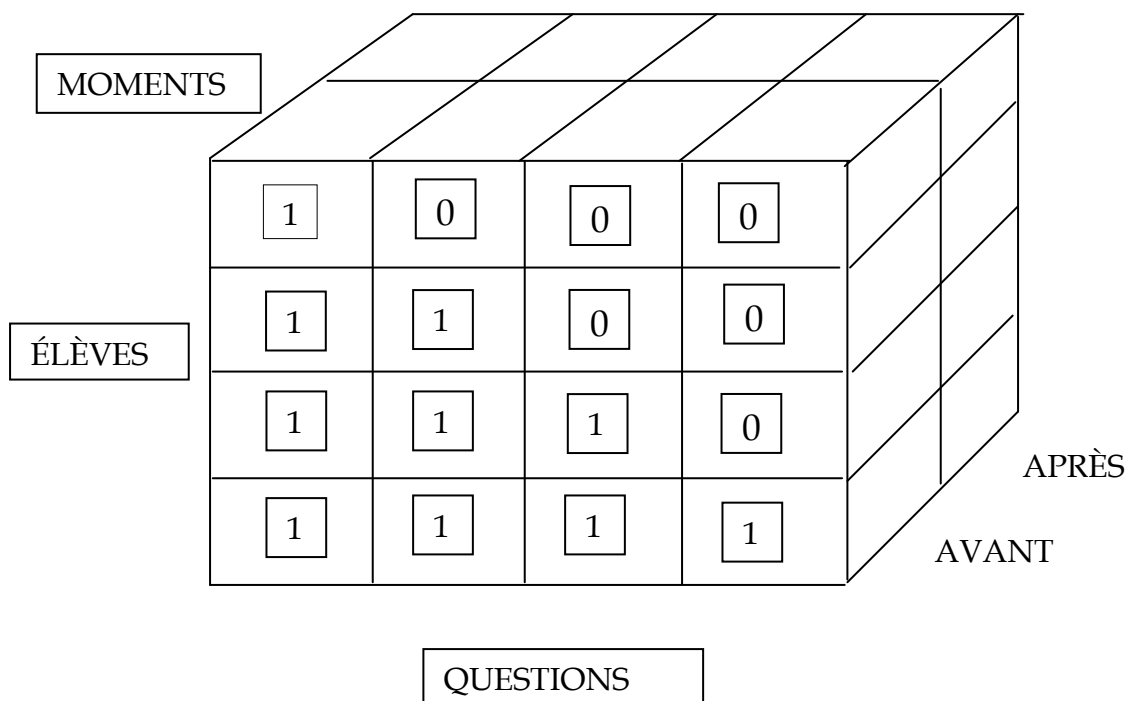
- Carver, R.P. (1970). Special problems in measuring change with psychometric devices. *Evaluative Research : Strategies and Methods*, Washington : American Institute for Research.
- Carver, R.P. (1972). Reading tests in 1970 versus 1980 : psychometric versus edumetric. *Reading Teacher*, 26, 229-302.
- Carver, R.P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, Vol 29(7), Jul. 1974. pp. 512-518.
- Cronbach, L.J. (1975). Dissent from Carver. *American Psychologist*, Vol 30(5), May 1975. pp. 602-603.
- Glastone, R. (1975). Where is fashion leading us? *American Psychologist*, Vol 30(5), May 1975. pp. 604-605.
- King, D.J. (1975). Control groups. *American Psychologist*, Vol 30(5), May 1975. pp. 602.
- Haladyna, T.M. (1975). On the psychometric-edumetric dimensions of tests. *American Psychologist*, Vol 30(5), May 1975. pp. 603-604.
- Cziko, G.A. (1983). Psychometric and edumetric approaches to languagetesting. In J.W. OLLER, Jr. (Ed.). *Issues in language testing research* (pp. 289-307). Rowley, MA: Newbury House.
- Hoz, R., Bowman, D., Chacham, T. (1997). Psychometric and edumetric validity of dimensions of geomorphological knowledge which are tapped by concept mapping. *Journal of Research in Science Teaching*, Vol 34(9), Nov 1997. pp. 925-947.
- Willoughby, T.L., Hutcheson, S.J. (1978). Edumetric validity of the Quarterly Profile Examination. *Educational & Psychological Measurement*, Vol 38(4), Win 1978. pp. 1057-1061.

POURQUOI FAUT-IL PARLER D'EDUMETRIE ?¹

Jean Cardinet

Dire que l'éduométrie n'est qu'un chapitre de la psychométrie, affirmer même qu'elle n'a pas de spécificité, comme le font certains, c'est s'imaginer qu'on ne peut rien mesurer en dehors des tests. La définition de l'éduométrie proposée par Viviane de Landsheere (1988) ouvre la voie au contraire à bien d'autres mesures en évoquant tout ce qui est « étude quantitative des variables relatives aux apprentissages suscités par l'éducation ». Les lignes suivantes essaient d'exemplifier de telles recherches. On verra que les objets d'étude en éducation débordent de beaucoup les « psychological/behavioral attributes » que vise la psychométrie et qu'il faut bien donner un nom (éduométrie) à l'activité de mesure visant tout cet ensemble.

Figure 1 - Représentation schématique du Cube des Données observées²



Pour fixer les idées, partons d'une classification simple, organisant des scores de performance à l'intérieur du « Cube des données » de la figure 1. Sur la face qui nous regarde, ce parallélogramme présente un tableau de résultats habituel, dont les lignes correspondent aux élèves et les colonnes aux questions, chaque case comportant un 1 pour une réponse correcte, ou un zéro en cas d'échec. Supposons que ce tableau a été rempli à la fin d'une année scolaire et que les mêmes questions ont été posées un an plus tard aux mêmes élèves, ce qui permet de

¹ MD : Ce texte a été publié, en réponse au précédent, dans le *Bulletin de l'Association pour le Développement des Méthodologie de l'Evaluation en Education* (ADMEE-Europe) (2002/3 et 2003/1).

² MD : les « 0 » et les « 1 » indiquent respectivement l'échec ou la réussite à une question particulière, pour un élève donné à un moment donné, disons lors d'un test ou d'une épreuve particulière.

construire un second tableau superposable au premier, et représenté par la tranche arrière du parallélogramme.

La psychométrie va « jouer » avec ces tableaux, par exemple en reclassant les élèves par ordre de réussite, de façon que pour chaque question (en colonne) on n'observe que des zéros en haut de la colonne (pour les élèves faibles) et que des 1 en bas de la colonne (pour les élèves forts). Lorsque les questions sont à leur tour reclassées par ordre de difficulté croissante, les données s'organisent pour former un triangle de zéros en haut à droite du tableau et un triangle de 1 en bas à gauche. On peut alors admettre qu'un seul facteur classe élèves et questions (*modèle de Guttman*). Le modèle de la *Théorie de Réponse aux Items* vise le même but, mais est plus souple, en admettant que ce soit seulement en probabilité que les 1 et les zéros se placent dans ces deux angles du tableau. Le modèle factoriel, enfin, admet que plusieurs sources de variation interviennent à la fois et qu'il faut plusieurs couches de composantes de scores pour rendre compte (approximativement) des réussites et des échecs observés.

De même, la psychométrie va étudier pour chaque question si les réussites et les échecs des élèves dans le premier tableau prédisent des résultats semblables dans le second tableau à l'arrière (validité prédictive de chaque question, mesurée par un *coefficient Phi*).

Dans de telles études, on parle bien toujours des comportements des Élèves observés, mais c'est justement parce qu'on reste en psychométrie pure, en ne s'intéressant qu'aux résultats des élèves ! Or il y a bien d'autres informations à tirer de ces données, par des recherches qu'on dira d'édu-métrie.

Prenons la dimension des Questions. Si on totalise les deux années, on peut classer les questions d'après leur taux de bonnes réponses et chercher à comprendre leur difficulté intrinsèque, en étudiant la structure de leur formulation verbale, le nombre d'étapes logiques dans la résolution du problème, le type d'illustration fournie, etc. Ou bien, comparant au contraire la difficulté de chaque question avant et après l'année d'étude, on pourra en tirer des enseignements sur l'effet du curriculum¹ suivi et les obstacles épistémologiques qui résistent. Ou bien encore, on pourra classer les questions par niveaux d'objectifs, pour savoir si l'apprentissage a suivi un certain ordre attendu a priori, etc. Cette fois, les objets d'étude ne sont plus les comportements des élèves. Ces comportements n'interviennent plus que comme des instruments pour réaliser des études totalement étrangères à la psychométrie, qui portent notamment sur la construction des questions ou des curriculums.

Prenons la dimension des deux Moments d'observation. La question la plus simple qui vienne à l'esprit est de savoir si un progrès a été accompli. Ici l'objet d'étude est la différence (Après - Avant). Le test de « t » pour échantillons appariés sera d'autant plus sensible que les résultats des élèves seront plus homogènes dans le cadre de chaque année. Or c'est là le contraire de ce que cherche à réaliser la psychométrie, qui différencie au maximum les sujets les uns par rapport aux autres. C'est bien la preuve que les objets d'étude de la psychométrie et de l'édu-métrie sont entièrement différents, puisque bien mesurer l'un oblige à ne pas bien mesurer l'autre.

¹ MD: Le terme "curriculum" a été introduit pour désigner l'ensemble des comportements qui interviennent dans la mise en place d'un programme de formation cohérent : le plan d'études indiquant les domaines de contenus à aborder et les objectifs à atteindre, les propositions de démarches à suivre dans l'organisation et l'animation des situations d'apprentissage, le matériel didactique (manuels, brochures, fiches, moyens audiovisuels, jeux, outils de travail, etc.) destiné à l'élève et aux enseignants. (Allal, L. (1996). L'évaluation des curricula in *Encyclopaedia universalis* ; voir aussi Demeuse, M., Strauven, S. (2006). *Développer un curriculum d'enseignement ou de formation. Des options politiques au pilotage*. Bruxelles : De Boeck université, Coll. « Perspectives en éducation et formation »)

Prenons même la dimension des Élèves qui intéresse spécifiquement les psychométriciens. Bien des études peuvent être effectuées qui supposent la création d'instruments tout différents des mesures psychométriques ordinaires. Supposons qu'on veuille voir en quoi diffère l'apprentissage des élèves suivant un même programme dans deux régions différentes d'un pays. Dans cette situation simple, fonction discriminante et équation de régression se confondent et on trouvera facilement la pondération optimale à donner aux questions pour différencier au mieux les deux régions. L'accent mis dans chaque région sur tel ou tel genre de questions pourra alors en être déduit. Or ces coefficients de pondération n'auront sûrement pas grand rapport avec les poids bêta pour estimer le facteur général des questions, le trait latent qui est l'objet d'étude privilégié de la psychométrie. Différencier les régions ne suppose pas du tout que l'on différencie d'abord les élèves, bien au contraire, comme l'exemple du test de « t » au paragraphe précédent l'a déjà montré.

Notre argumentation se veut concrète : nous montrons simplement que des instruments psychométriques ne sont pas appropriés pour ces quelques exemples de recherches en éducation. Mais le nombre d'études non-psychométriques possibles devient infini si l'on songe que chacune des trois dimensions Élèves, Questions et Moments peut être structurée à volonté pour créer de nouvelles dimensions super ordonnées. Par exemple, au lieu de grouper les élèves par région, on peut vouloir les différencier par niveau socio-économique d'origine, ou par sexe, ou par leur choix de langue étrangère, ou par établissement, par type d'enseignant... A chaque fois, les instruments à utiliser devront être choisis en fonction de la différenciation visée, la différenciation psychométrique des élèves représentant plutôt une source de perturbation dans nos résultats.

De même, la dimension des Questions pourra être structurée par toutes sortes de typologies d'objectifs éducatifs, ou d'autres principes classificatoires, issus des questions que l'on se posera. Les Élèves seront considérés comme les outils pour effectuer ces études. La dimension des Moments, à son tour, pourra comporter bien plus de deux niveaux et donner lieu à des regroupements divers, pour mettre en évidence par exemple des stades de développement successifs, comportant des paliers temporaires.

Enfin, ce seront souvent des interactions entre facteurs super ordonnés qui seront les objets d'étude, entre niveaux socio-économiques et progrès d'une année à l'autre, selon les types de compétences visées, par exemple.

En résumé, la recherche en éducation a besoin de mesures de toutes sortes (qu'il est commode d'appeler « éduométriques ») pour analyser le fonctionnement du système de l'enseignement. Il est rare que des épreuves construites pour différencier les élèves puissent servir d'instruments à ces recherches. Les tests psychométriques ne sont conçus en fait qu'en vue de classer les élèves pour fonder la sélection scolaire. Et là-dessus, il y aurait encore beaucoup à dire...

Bibliographie

de Landsheere, V. (1988). *Faire réussir, faire échouer. La compétence minimale et son évaluation*. Paris : Presses universitaires de France.