**Fondements et limites des Modèles de Réponse à l’Item**

Réginald Burton (Université du Luxembourg)

De plus en plus de dispositifs d’évaluation locale (évaluation formative, monitoring, épreuves d’orientation scolaire) ont recourt aux Modèles de Réponse à l’Item (IRT models) pour estimer les compétences scolaires des élèves au départ de tests papier-crayon ou assistés par ordinateur. Ces modèles de mesure - dont les propriétés laissaient entrevoir des progrès et des avantages considérables - se sont d’abord imposés comme standards de référence dans les enquêtes internationales pour être ensuite transposés dans des contextes locaux.

Cependant, la technicité des MRI, leur nature stochastique et leurs composantes statistiques relativement complexes n’ont pas toujours permis une approche cartésienne et prudente de leur exploitation. On a cru pendant longtemps que la transposition de ces modèles des enquêtes internationales aux évaluations locales était naturelle et ne souffrait d’aucun obstacle fondamental. Mais les objectifs, les caractéristiques et les exigences des évaluations locales sont parfois bien éloignés des préoccupations internationales ce qui peut conduire à des erreurs de mesure.

Nous nous proposons donc, au départ de considérations théoriques et résultats empiriques, de déterminer, dans quelles conditions, l’application des MRI aux politiques d’évaluations locales est pertinente ? Dans cette perspective, plusieurs critères d’analyse seront envisagés : la nature théorique des modèles (les MRI sont-ils adaptés pour atteindre les objectifs des évaluations locales ?), les caractéristiques psychométriques des évaluations (les conditions d’application des MRI sont-elles toujours remplies dans le cadre des évaluations locales ?) et les méthodes validation des modèles (les méthodes pour évaluer l’adéquation des modèles sont-elles adaptées aux évaluations locales ?).

REFERENCES

Burton, R. (2012). La Robustesse des Modèles de Réponse à l’Item. Communication présentée au 4e colloque international Méthodes Quantitatives et Sciences Humaines. Le 8 juin 2012, Québec (Canada).

Burton, R. (2012). Quels modèles de mesure pour l’évaluation des compétences ? Communication présentée au 24e colloque international de l’ADMEE-Europe. L'évaluation des compétences en milieu scolaire et en milieu professionnel. Du 11 au 13 janvier 2010, Luxembourg.

Burton, R. & Dierendonck, C. (2011). Les enquêtes internationales. Québec : Presses Université Laval.

Burton, R. & Dierendonck, C. (2010). Modèles de mesure et évaluation des compétences complexes. Conférence présentée au 16e congrès de l’Association Mondiale des Sciences de l’Education (AMSE). Evaluation des performances complexes. Du 31 avril au 4 juin 2010. Monterrey (Mexique).

Burton, R. (2004). Influence des distributions du trait latent et de la difficulté des items sur les estimations du modèle de Birnbaum : une étude du type Monte-Carlo. Mesure et évaluation en éducation, 27, 3, 41–62.

Fagnant, A. & Burton, R. (2010). Développement de compétences et résolution de problèmes en mathématiques à l’école primaire : pratiques déclarées des enseignants et pratiques projetées des futurs enseignants. Scientia Paedagogica Experimentalis (à paraître dans le numéro 46(2) de 2009).

Hambleton, R. K., Swaminathan, H. & Rogers, H.J. (1991). Fundamentals of item response theory. Newbury Park CA : Sage publications.

Vrignaud, P. & Chartier, P. (1999). Quand les modèles de mesure deviennent réducteurs : apports et limites des Modèles de Réponse à l’Item pour les comparaisons internationales. Paris : Service de Recherche de l’INETOP.

Wu, M.L., Adams, R.J. & Wilson, M.R. (1997). ConQuest. Generalized item response modelling software. Draft release. Hawthorn Australia : ACER.