**L’évaluation standardisée à grande échelle: enjeux théoriques et pratiques**

Reinaldo Dos Santos (Direction de l’évaluation, de la prospective et de la performance)

La théorie classique des tests est un champ de la psychométrie qui se concentre sur la détermination des caractéristiques d’un test en tant que tel, et notamment sur les deux questions fondamentales qui se posent à l’évaluateur lors de la conception du test : est-ce que ce test mesure bien ce que je souhaite mesurer ? Est-ce qu’il le mesure de manière cohérente ? (Novick, 1966)

La qualité d’un test, selon la théorie classique, se décrit à l’aide de deux indicateurs : sa validité et sa fidélité. La fidélité d’un test représente sa cohérence interne. Elle permet de vérifier que l’ensemble des items du test mesure bien la même dimension psychométrique (Cronbach, 1951). La validité d’un test représente la précision du test dans sa mesure du niveau réel du sujet (Cronbach & Meehl, 1955).

Si la théorie classique des tests présente une utilité certaine pour analyser un test dans son ensemble, elle ne donne que peu de précisions sur le contenu détaillé de celui-ci. C’est pourquoi il est intéressant de lui substituer la théorie de réponse à l’item, afin de répondre aux deux questions suivantes : quelle est la difficulté réelle de tel item ? Quel est le niveau réel de tel sujet ? (Lord et al., 1968 ; Rasch, 1993)

Le postulat fondamental des modèles de réponse à l’item est que la performance d’un sujet à tel item est fonction d’un trait latent, et non directement observable. Le recours aux méthodes probabiliste permet d’approcher une mesure de ce trait, par la combinaison de performances à différents items de difficulté connue.

Enfin, le modèle de réponse à l’item permet de construire des tests d’un plus haut niveau de complexité. Nous prendrons ici l’exemple des tests dits adaptatifs (Yan et al., 2014 ; Jodoin et al., 2006).

Faire passer tel item à tel sujet est porteur d’information si la différence entre la difficulté réelle de l’item et le niveau réel du sujet n’est pas trop important. Par conséquent, comment réaliser un test qui soit optimal en termes d’items pour un public particulièrement hétérogène ? Cette difficulté se rencontre souvent dans le cadre de l’évaluation professionnelle, de la formation continue, mais également dans le cadre de l’évaluation fine d’une population fortement disparate (Journée Défense Citoyenneté, sujets en situation de handicap).

Un test adaptatif a pour principe de chercher à estimer le niveau réel du sujet en cours de test, de façon à lui proposer des items suffisamment proches de ce niveau supposé pour être porteurs d’information. On peut distinguer les tests dit "full adaptative" qui réajustent leur estimation après chaque item, et les tests "multistage" qui fonctionnent par paliers.

Nous prendrons comme exemple le HandiPanel réalisée par la Direction de l’évaluation, de la prospective et de la performance du ministère de l’éducation nationale.

REFERENCES

Allen, M. & Yen, W. (2002). Introduction to measurement theory. Long Grove, IL : Waveland Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297–334.

Cronbach, L. J. & Meehl, P. E. (1955). *Construct validity in psychological tests.* Psychological bulletin, 52 (4), 281-303.

Jodoin, M. G., Zenisky, A. & Hambleton, R. K. (2006). *Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes*. Applied Measurement in Education, Volume 19, Numéro 3, Pages 203 à 220.

Laveault, D. & Grégoire, J. (2002). *Introduction aux théories des tests : en psychologie et en sciences de l’éducation.* De Boeck Supérieur.

Lord, F. M., Novick, M. R. & Birnbaum, A. (1968). *Statistical theories of mental test scores.*

Novick, M. R. (1966). *The axioms and principal results of classical test theory. Journal of mathematical psychology*, Volume 3, Numéro 1, Pages 1 à 18.

Rasch, G. (1993). Probabilistic models for some intelligence and attainment tests. Chicago: MESA Press.

Yan, D., von Davier, A. A. & Lewis, C. (2014). *Computerized multistage testing : Theory and applications.* Boca Raton: CRC Press.