

Contribution 3

Utilisation des degrés de certitude pour étudier la structure et la dimensionnalité d'un test par QCM (7724)

Oliver Prosperi

Institut de recherche et de documentation pédagogique, Suisse

Mots-clés : degrés de certitude, dimensionnalité, ACP

Les degrés de certitude sont une technique d'attribution du score à un item, souvent une question à choix multiple, où on demande à la personne examinée d'exprimer l'estimation de sa probabilité subjective que la réponse donnée soit correcte. Ceci signifie que l'étudiant est appelé à accompagner chaque réponse d'un pourcentage (entre 0 % et 100 %) qui représente la probabilité que la réponse soit correcte. Cependant, pour des raisons de praticabilité et du fait que les étudiants ne seront pas en mesure d'apprécier la différence de quelques points de pourcentage (Leclercq 1982), le continuum 0%-100% est partagé en intervalles dont la taille et le nombre dépendent du contexte évaluatif. On définit ainsi une échelle d'intervalles de certitude qui sont par la suite numérotés et appelés degrés de certitude (DC). Le type de DC utilisés dans cette communication a été présenté par Leclercq, Boxus, de Borgniez, Wuidare et Lambert (1993) et correspondent à une échelle avec six degrés, numérotés de 0 à 5 et dont les valeurs seuil qui partagent le continuum sont (0%, 25%, 50%, 70%, 85%, 95% et 100%). Le score attribué à l'étudiant pour une question est calculé en fonction de la réponse (correcte/incorrecte) et du degré de certitude exprimé. Ceci est fait selon un schéma d'attribution du score prédéfini qui correspond à un tableau à double entrée : une entrée a deux options (correct/incorrect) et l'autre a un nombre d'options égales au nombre de degrés de certitude.

Lors d'un test avec degrés de certitude l'examineur relève donc deux types d'informations : la réponse donnée et le DC choisi pour chaque réponse. Le traitement de ces informations à niveau du groupe d'étudiants qui passent un test a permis le développement d'indices de qualité des items (Gilles, 2010), mais aussi de nouveaux modèles psychométriques de la famille des modèles de Rasch (Prosperi, 2017).

Or, un prérequis fondamental pour l'application du modèle de Rasch est celui de l'unidimensionnalité du test. En d'autres termes, il est requis que le test mesure et quantifie une seule caractéristique. En effet, l'application d'un modèle unidimensionnel à des données multidimensionnelles peut introduire de l'instabilité dans la détermination des paramètres des items et ne plus garantir la propriété d'invariance des paramètres des items (Alagumalai et Keeves, 1996 et Reckase, 1979).

Il existe une multitude d'instruments pour déterminer l'unidimensionnalité d'un test et Hattie (1985) en fait un long recensement au bout duquel il suggère que pratiquement tous les indices présentent des failles et ne permettent pas d'affirmer à 100% l'unidimensionnalité. Dans la pratique, les chercheurs utilisent principalement trois méthodes pour vérifier l'unidimensionnalité. Pour l'application des modèles de Rasch, on valide l'unidimensionnalité a posteriori en vérifiant que le modèle s'adapte bien aux données (si après avoir appliqué le modèle les données statistiques de « fit » sont bonnes, on peut affirmer que les données de départ sont unidimensionnelles) ou en effectuant une analyse en composantes principales (APC) sur les résidus de Rasch. La procédure DIMTEST (Nandakumar and Stout, 1993) représente la troisième méthode souvent citée dans la littérature. Dans le cas de réponses multinomiales (par exemple une échelle de Likert) il existe des méthodes basées sur l'analyse en composantes principales. Les « rules of thumb » de ces méthodes suggèrent qu'on peut affirmer

l'unidimensionnalité s'il existe un facteur qui explique plus de 20% de la variance totale et que le rapport entre « la première valeur propre moins la deuxième » et « la deuxième moins la suivante » soit « grand ».

Sur un autre plan, Reckase (1979) affirme que souvent dans la construction de tests conçus pour la mesure de compétence dans un domaine on ne se base pas sur une méthodologie orientée vers la mesure d'un facteur unique, mais plutôt sur une table de spécification qui dirige la création des questions sur la base des différents chapitres qui y figurent. Ceci, conclut-il, mène souvent à des structures qui sont factoriellement complexes et mesurent rarement un seul trait.

Les données qui sont à la base des élaborations qui font l'objet de cette communication proviennent d'un test de maîtrise de la langue française et qui est construit sur la base d'une table de spécification qui possède quatre chapitres (Orthographe, Vocabulaire, Syntaxe et Compréhension). On est donc dans la situation décrite par Reckase où on veut mesurer un trait (la maîtrise du Français) avec un instrument qui a des fortes chances d'être multifactoriel à cause des chapitres multiples de la table de spécification. Chaque chapitre se décline en 15 questions à choix multiple et le test est fait passer avec la technique des degrés de certitude.

Sur la base du succès obtenu dans l'exploitation des DC pour obtenir de l'information au sujet des items surgit la question de recherche qui est l'objet de cette communication : « Est-ce que l'expression des degrés de certitude contient de l'information relative à la structure et à la dimensionnalité d'un test ? »

Afin de répondre à cette question, les degrés de certitudes issus du test de maîtrise du français ont été utilisés pour effectuer une analyse par composantes principales avec rotation Varimax. L'application de la méthodologie suggérée par Hattie et par Fields (2009) montre que l'analyse des DC permet de recomposer la structure de la table de spécification et de donner des indices permettant de s'exprimer sur la dimensionnalité du test.