

Rezensionen / recensions / recensioni

Cardinet, Jean; Johnson, Sandra & Pini, Gianreto (2010). *Applying Generalizability Theory using EduG*. New York: Routledge. 240 p.

Mesurer et évaluer les performances, les comportements ou les attitudes des individus de manière fiable est l'un des principaux défis des sciences sociales. Comment s'assurer qu'un test est fidèle, qu'il mesure de manière précise ce pour quoi il est conçu? Dans quelle mesure un test est-il stable d'une utilisation à l'autre ou d'un contexte à l'autre?

La théorie de la généralisabilité (TG) est actuellement le modèle le plus avancé pour pouvoir répondre à ces questions. Elle émane des modèles de mesure classiques dont elle reprend le principe selon lequel le score observé d'un individu est la somme de deux termes théoriques: le score «vrai» (ou score univers), qui correspond à ses compétences réelles, et l'erreur de la mesure, qui correspond à l'écart entre valeur observée et valeur vraie. Pour qu'une mesure soit fiable, il faut que l'erreur de mesure soit minimale. Mais pour pouvoir minimiser l'erreur, il est indispensable de pouvoir la contrôler en identifiant les facteurs qui en sont responsables. Que contient-elle donc cette erreur de la mesure? Quels sont les facteurs qui la déterminent? Les modèles classiques ne permettent pas d'y répondre, la raison étant que la source d'erreur y est unique et indifférenciée.

L'élaboration de la TG à partir des années 1960 a permis de répondre à ces questions grâce aux nouvelles possibilités de décomposition de la variance, notamment pour ce qui est de la variance d'erreur. Il devient alors possible de définir un plan de mesure dans lequel plusieurs sources d'erreur sont identifiées et isolées. L'analyse de généralisabilité permet d'estimer les effets de ces différents facteurs, ainsi que de leurs interactions. A partir des analyses de fiabilité, l'évaluateur est également à même d'optimiser son plan d'analyse en intervenant directement sur les facteurs qui affaiblissent les qualités métriques de l'instrument.

L'innovation de la TG va plus loin encore: alors qu'à l'origine le seul objet d'intérêt était l'individu, l'application du principe de symétrie permet de considérer comme objets de mesure d'autres éléments jouant un rôle primordial dans une démarche de recherche ou d'évaluation (p. ex. les items, les méthodes d'enseignement, les classes ou les juges). La TG devient ainsi l'un des outils de référence pour contrôler la fiabilité de divers instruments ou dispositifs de mesure.

L'ouvrage de Cardinet, Johnson et Pini est la plus récente contribution à la divulgation de la TG. Le livre est remarquable par sa concision, par son caractère éminemment pédagogique et extrêmement concret qui ne peut que réjouir le lecteur avide d'exemples pratiques. Les auteurs situent au cœur de l'ouvrage le logiciel EduG, un programme d'analyse de la généralisabilité auquel ils consacrent un chapitre et auquel ils se réfèrent dans une série d'exemples

pratiques et d'exercices. Par ailleurs, ces exemples pratiques et ces exercices font appel à une même série de données, disponibles comme EduG en libre accès sur Internet (<http://www.irdp.ch/edumetrie/logiciels.htm>), ce qui multiplie les possibilités de mise en relation entre les volets pratiques du livre. L'avant-propos de Richard Shavelson – l'un des auteurs qui a considérablement contribué au développement de la TG – laisse comprendre que ce livre était particulièrement attendu outre-Atlantique par le milieu académique et étudiant.

Le livre est organisé en six chapitres. Les chapitres 1 et 2 sont consacrés à l'illustration conceptuelle de la TG. Ils contiennent un aperçu des origines du modèle, les objectifs visés par la théorie ainsi que les développements théoriques du modèle. On y trouve également une explication des principes de décomposition de la variance, de l'identification des sources de variation et de la distinction entre coefficients de généralisabilité relatif et absolu. Cette partie théorique de l'ouvrage se termine par l'illustration des différentes procédures d'optimisation d'un plan d'analyse.

Le chapitre 3 est entièrement réservé à la présentation d'EduG. Très simple d'utilisation, disponible en français et en anglais, EduG est spécialement conçu pour l'application du principe de symétrie. Dans ce chapitre sont précisées les démarches de création et de traitement des données, de définition des plans d'analyse (plans d'observation, d'estimation et de mesure) et d'interprétation des résultats observés.

Dans le chapitre 4, les auteurs présentent des exemples d'application de la TG. Ainsi, il est proposé d'analyser une étude qui mesure la dépression chez des patients psychiatriques, d'évaluer une méthode de recrutement de futurs managers ou encore d'estimer la difficulté relative d'une série de problèmes ou exercices en vue d'une évaluation en mathématique. Tous les exemples comportent une description du contexte, la justification de l'intérêt d'une étude de généralisabilité dans un tel contexte, une explication sur la mise en place de l'étude de généralisabilité, une illustration de la répartition de la variance par des diagrammes et une aide pour l'interprétation des résultats.

Le chapitre 5 contient cinq exercices pratiques. Ces exercices se composent d'une mise en contexte et d'une série de questions de difficulté variable suivies par les réponses correspondantes. Le lecteur appréciera la présence dans les exercices de tableaux de résultats, de graphiques et de diagrammes.

Le chapitre 6 est le chapitre conclusif. Ici, le lecteur peut être surpris par un retour aux origines du modèle, déjà abordé dans le chapitre 1. Les raisons de cette reprise ne sont pas énoncées, mais l'intention des auteurs est visiblement d'approfondir certains aspects théoriques et d'établir un lien avec les développements futurs. Le chapitre est enrichissant pour le lecteur qui s'intéresse aux coefficients de généralisabilité comme estimateurs de la taille de l'effet dans la recherche expérimentale. Ou encore comme moyen d'évaluer les effets d'interaction dans le but d'optimiser le plan d'analyse. Le chapitre se conclut par une mise en relation de la TG avec la Théorie de réponse aux items et par la discus-

sion des avantages que constituerait l'utilisation conjointe des deux théories.

L'ouvrage est complété par une section importante d'annexes regroupant les aspects plus techniques de l'ouvrage. On y voit la volonté des auteurs d'alléger le texte sans pour autant renoncer complètement à l'illustration mathématique. Les annexes sont consacrés à une introduction à l'analyse de la variance, au problème des sommes des carrés pour les plans non équilibrés, à la relation entre le coefficient de généralisabilité et d'autres indices statistiques (rhô carré ou oméga carré), ainsi qu'à la définition des intervalles de confiance.

Comme son titre l'indique, le livre est écrit en anglais, mais dans un anglais très accessible, même pour les lecteurs moins familiers avec cette langue. Les auteurs parviennent à conjuguer brillamment théorie, pratique et technique en à peine 200 pages. Grâce à sa structure groupée, l'ouvrage permet une lecture à plusieurs niveaux, laissant le choix d'aborder le livre soit par une lecture globale, soit par une lecture plus approfondie, complétée par la réalisation d'exercices et par l'utilisation du programme EduG.

Autant pour son contenu que pour sa forme, cet ouvrage est vivement conseillé aux personnes s'intéressant aux problèmes de la mesure, que ce soient des chercheurs ou des étudiants de niveau avancé ayant des connaissances en statistique. Ce livre s'adapte également à l'enseignement universitaire, tout particulièrement lorsqu'il est accompagné de travaux pratiques.

Alexander Tomei, Section des Sciences de l'éducation, Université de Genève