
Validation Psychométrique Central Test



L'art de l'évaluation

Propriétés psychométriques des tests

L'évaluation périodique de la validité et l'ajustement des tests sont des critères essentiels pour justifier de leur efficacité et de leur utilité. Central Test apporte une attention particulière à ces éléments en révisant régulièrement les propriétés psychométriques de tous ses outils. Ces révisions impliquent une mise à jour du contenu, des analyses statistiques et des études de validité. Avant la publication en ligne, un test est évalué selon quatre critères présentés ci-dessous.

Fiabilité ou fidélité du test

La fidélité fait référence à la façon dont un test mesure une dimension donnée (stabilité des résultats, cohérence de la mesure). Si une personne réalise le test à nouveau, aura-t-elle le même score, ou bien un score très différent ? Un test qui fournit des scores similaires lorsqu'une personne le passe à plusieurs reprises est considéré comme un test fiable. Il faut toutefois noter que la mesure dans les sciences comportementales est influencée par des variables externes qui sont :

- **L'état psychologique ou physique de la personne qui passe le test, au moment où elle le passe.** Par exemple, différents niveaux de fatigue, d'anxiété, ou de motivation peuvent influencer la performance de l'individu et donc ses résultats.
- **Les facteurs environnementaux.** Des différences dans l'environnement de passation du test, comme la lumière, le bruit,... peuvent influencer la performance d'un individu à un test.

Ces facteurs sont communément appelés sources « d'erreurs de mesure aléatoires ». S'il n'y avait pas d'erreur de mesure aléatoire, un individu aurait le même score au test, son « score vrai », à chaque passation, quelles que soient les conditions de passation. Un indice de la fiabilité du test est la part du score au test non affectée par l'erreur de mesure aléatoire.

Il existe divers types d'estimation de la fidélité, chacun influencé par différentes sources d'erreurs de mesure. Pour les outils développés par Central Test, deux types de fidélité sont particulièrement importantes :

- **La consistance interne.** Elle indique à quel point les items d'un test mesurent la même dimension. Un coefficient de consistance interne élevé indique que les items du test sont similaires dans leur contenu (c'est-à-dire homogènes). Il est important de noter que la longueur du test peut affecter la consistance interne. Par exemple, un test très long peut, de façon erronée, gonfler le coefficient de fidélité. La consistance interne est communément mesurée par un Alpha de Cronbach qui varie entre 0 (faible) et 1 (élevé).

- **La fidélité test-retest.** Elle indique le degré de répétabilité dans le temps des scores à un test. Cette estimation reflète également la stabilité des caractéristiques ou du construit évalués par le test. Certains construits sont plus stables que d'autres. Par exemple, la capacité de lecture d'un individu est plus stable, sur un même laps de temps, que son niveau d'anxiété. Pour les construits qui sont supposés varier avec le temps, un coefficient de fidélité test-retest acceptable peut être plus faible que celui des construits stables dans le temps.

Le niveau de fidélité acceptable dépend du type de test et du type de coefficient utilisé.

Pour les outils de Central Test, les mesures de fidélité sont fournies dans les documents techniques des tests.

Validité du test

La validité est le critère le plus important dans le choix d'un test. Elle fait référence aux *caractéristiques mesurées* par le test et à *l'exactitude* avec laquelle il évalue ces caractéristiques. La validité donne du sens aux scores du test. Les preuves de la validité indiquent qu'il existe un lien entre la performance au test et un critère de performance. Elle nous informe sur le degré avec lequel il est possible de tirer des conclusions spécifiques, ou des prédictions, sur la personne en fonction de son score au test. En d'autres mots, elle renseigne sur l'utilité du test. Par exemple, si un test est un prédicteur valide de la performance pour un poste particulier, nous pouvons conclure qu'une personne présentant un score élevé au test sera plus encline à être compétente à ce poste qu'une personne qui a obtenu un score faible, toutes choses étant égales par ailleurs.

Il est important de comprendre la différence entre la fidélité et la validité. La validité nous dit à quel point le test est bon pour évaluer une situation particulière ; la fidélité nous dit à quel point ce test est crédible, stable. La validité d'un test est établie en référence à un objectif particulier.

Il existe différents types de validité conceptualisés par les chercheurs. Cependant, tous peuvent être classés selon trois catégories :

- **La validité de critère.** Elle est évaluée en examinant la corrélation, ou autre relation statistique, entre la performance au test et un critère externe de performance. En d'autres mots, les individus qui présentent un score élevé au test tendent à être plus performants sur ce critère externe que les individus qui ont un score faible au test. Si le critère est obtenu en même temps que la passation du test, on parle de validité concourante ou concomitante ; si le critère est obtenu plus tard, on parle de validité prédictive.

- **La validité de contenu.** Elle est étudiée en examinant si le contenu du test est représentatif du construit que l'on veut mesurer. Par exemple, dans le cadre d'un test d'évaluation du potentiel commercial, les items du test doivent être pertinents et évaluer directement les exigences et qualifications pour le poste de commercial. Le contenu du test doit également correspondre au niveau de compréhension de la population cible.

- **La validité de construit.** Elle nécessite la démonstration que le test mesure bien le construit ou la caractéristique qu'il prétend mesurer. Par exemple un test pour les intérêts au travail doit évaluer les intérêts au travail et non pas le niveau de motivation. La validité de construit est, en général, évaluée de deux façons distinctes :

- En demandant à des experts de juger de la concordance entre les items du test (le contenu) et le construit évalué (**validité de façade**) ;
- En administrant le test avec d'autres tests développés sur des construits théoriquement similaires et en examinant la corrélation entre les deux (**validité convergente**) ou en administrant le test en même temps que des tests théoriquement opposés et en examinant leur corrélation (**validité divergente**).

Les trois types de validité, de critère, de contenu et de construit, sont utilisés en tant que supports de validation en fonction de la situation. En général, ces trois méthodes se chevauchent et, selon la situation, toutes ne sont pas appropriées.

Chez Central test, les études de validité de nos tests sont menées en interne ainsi qu'auprès d'experts. Les résumés de ces études sont fournis dans les documents techniques des tests et utilisés pour leur amélioration constante.

Désirabilité sociale

La désirabilité sociale est un biais dans l'évaluation correspondant à l'inclinaison d'une personne à répondre d'une façon qui sera vue comme favorable par les autres. Étant des êtres sociaux par nature, les individus sont enclins à rechercher un certain degré d'acceptation sociale. La désirabilité sociale est l'un des principaux biais dans les tests de personnalité ; il est encore plus prééminent dans le contexte du recrutement. Un candidat est normalement toujours tenté de répondre de sorte à impressionner positivement le recruteur.

Nous avons développé une stratégie en deux points pour surmonter le problème. Premièrement, tous les items sont examinés afin de s'assurer qu'ils ne sont pas sensibles à la désirabilité sociale. Deuxièmement, nous incluons une échelle de désirabilité sociale, dans les tests, afin d'identifier la « tricherie » et ce, qu'elle soit délibérée ou non.

Considérations socio-légales

Enfin, un bon test psychométrique ne doit pas être biaisé (positivement ou négativement) envers un groupe socio-culturel particulier. Il ne doit pas montrer de discrimination sur la base de la religion, du genre, de l'origine ethnique ou de la culture de la personne qui passe le test.

Lors du développement de nos outils, tous les items sont examinés de sorte à ne pas être favorable à une population particulière. Dans les études de validation, les relations entre les variables démographiques et les scores au test sont examinées afin d'éviter un biais potentiel. Ceci nous permet de construire des tests sensibles mais non biaisés.