

GUIDE UTILISATEUR

VALIDATION

PSYCHOMÉTRIQUE

*Ce guide a été réalisé par les psychologues du département de
Recherche et Innovation de Central Test.*

*Tout droit de reproduction, modification ou réutilisation doit être
accordé par écrit par l'éditeur Central Test.*

Copyright © 2022 par Central Test - Empowering Talent.

Tous droits réservés.

SOMMAIRE

Introduction.....	2
Fidélité	3
Validité.....	5
Désirabilité sociale.....	7
Considérations socio-légales	8
Conclusion	10

Introduction

Les révisions régulières des propriétés psychométriques sont cruciales pour garantir l'efficacité à long terme de toute évaluation psychométrique. Central Test prend donc le plus grand soin à évaluer régulièrement les propriétés psychométriques de toutes ses évaluations. Cela garantit également que toutes nos évaluations sont conformes aux standards des organismes de référence tels que l'Equal Employment Opportunity Commission (EEOC). Ces révisions sont effectuées en mettant régulièrement à jour les normes et l'analyse statistique des rapports et des manuels, ainsi que par la formation des utilisateurs sur les obligations éthiques à remplir lors de l'administration des tests et de l'interprétation des résultats.

Les évaluations de Central Test font l'objet d'un examen du contenu, d'analyses statistiques et d'études de validité indépendantes. Chaque évaluation est obligatoirement évaluée par rapport à quatre critères distincts avant publication : *fidélité, validité, désirabilité sociale et considérations socio-légales.*

Fidélité

La fidélité, ou fiabilité, fait référence à la cohérence avec laquelle une évaluation mesure une caractéristique. Lorsqu'une évaluation donne des scores similaires pour un individu qui répète l'évaluation, on dit alors qu'elle mesure la caractéristique de manière fiable. Il faut cependant noter que la mesure en sciences du comportement est toujours influencée par certaines variables externes. Ceux-ci pourraient être :

- **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 et d'autres facteurs similaires 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. Par conséquent, des outils d'évaluation fiables produisent des informations fiables, reproductibles et cohérentes sur les individus. 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** de chaque facteur. Elle indique à quel point les items d'un test mesurent la même construit. 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. Par conséquent, on s'attendrait à un coefficient de fiabilité test-retest plus élevé sur une évaluation en lecture que sur une évaluation qui mesure l'anxiété. Pour les construits qui sont supposés varier avec le temps, le coefficient de fidélité test-retest acceptable peut être plus faible que celui des construits stables dans le temps.
- **Test-retest reliability** indicates the repeatability of assessment scores over time. These estimates also reflect the stability of the characteristic or construct being measured by the assessment, with some constructs being more stable than others. For example, an individual's reading ability is more stable over a particular period than their anxiety level. Therefore, one would expect a higher test-retest reliability coefficient on a reading assessment than on an assessment that measures anxiety. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than those constructs that are stable over time.

La fiabilité test-retest est rapportée comme la corrélation entre deux administrations de l'évaluation au même échantillon. En tant que développeur d'évaluation soucieux de la qualité, Central Test rapporte des

estimations de fiabilité qui sont pertinentes pour une évaluation particulière tout en tenant compte du fait que le niveau acceptable de fiabilité diffère selon le type d'évaluation et l'estimation de fiabilité utilisée.

Validité

La validité fait référence aux caractéristiques mesurées par le test et à l'exactitude avec laquelle il évalue ces caractéristiques. Par conséquent, les études de validité nous indiquent si les caractéristiques mesurées par une évaluation sont liées aux exigences d'une situation d'évaluation. 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. Par conséquent, s'il a été démontré qu'une évaluation est un prédicteur valide de la performance dans un travail spécifique, on peut conclure que les individus obtenant un score élevé à l'évaluation sont plus susceptibles de mieux performer au travail que ceux qui obtiennent un score faible à l'évaluation, toutes choses égales par ailleurs. Par conséquent, la validité décrit le degré auquel on peut tirer des conclusions ou des prédictions spécifiques sur des individus en fonction de leurs scores d'évaluation. En d'autres termes, il indique l'utilité de l'évaluation.

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. Dans toute la documentation d'évaluation de Central Test, le contexte d'évaluation et le groupe cible sur lequel le test est validé sont mentionnés.

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 trois 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**)
 - En vérifiant que les facteurs du test sont conformes au modèle théorique proposé pour la construction du test (**CFA - Analyse Factorielle Confirmatoire**).

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. Lorsque c'est fait involontairement, on parle d'auto-illusion ou aveuglement. Lorsque fait intentionnellement, on parle de gestion des impressions (Tett & Simonet, 2011)[1]

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.

Pour minimiser le risque de truquer les résultats (intentionnellement ou non), la désirabilité sociale est contrôlée dans les évaluations psychométriques, soit au moyen d'une échelle de désirabilité sociale soit dans le format de l'évaluation.

- **Les évaluations ipsatives** protègent activement contre la désirabilité sociale grâce à l'utilisation d'un format de questionnaire qui comprend des questions présentées dans un format à choix forcé, dans lequel chaque

question est composée de deux déclarations générales qui obligent l'individu à choisir l'option qui représente le mieux sa préférence. La recherche a montré que les mesures ipsatives étaient plus résistantes aux réponses biaisées par rapport aux échelles de Likert plus couramment utilisées (Bowen, Martin, & Hunt[2], 2002 ; Kluger, Reilly, & Russell, 1991)[3].

- Dans **les évaluations normatives**, la désirabilité sociale est mesurée en ajoutant des questions faisant l'objet d'une échelle spécifique : l'échelle de désirabilité sociale. Elles sont construites de manière similaire aux autres questions de l'évaluation. Elles sont conçues de manière à s'assurer qu'elles ne soient pas facilement détectées, et le test inclura un score de désirabilité sociale dans le rapport.

Considérations socio-légales

Enfin, un bon test psychométrique ne doit pas être biaisé (positivement ou négativement) envers un groupe socioculturel particulier. Il ne doit pas montrer de discrimination sur la base de l'âge, du genre, du pays d'origine ou d'un handicap. Par conséquent, lors de l'élaboration des évaluations, tous les items sont examinés par rapport à la pertinence du contenu via des analyses statistiques comme :

- **l'ANOVA et tests**, qui sont utilisés pour comparer les moyennes des groupes.
- **Le fonctionnement différentiel des éléments (DIF)**, qui fournit des informations indiquant si un item est perçu de manière similaire dans différents groupes.

Des mesures sont prises au cours du développement pour s'assurer que les évaluations ne sont pas biaisées en faveur de la population cible concernant des critères démographiques tels que :

- **l'âge** - Des normes sont calculées et des analyses statistiques sont effectuées pour s'assurer que les items ne favorisent pas un certain groupe

d'âge par rapport à l'autre.

- **le handicap** – Les évaluations offrent une accessibilité tout en incorporant des polices adaptées aux dyslexiques.
- **le pays d'origine** – Des normes sont calculées et des analyses statistiques sont effectuées pour s'assurer qu'il n'y a pas de biais culturel.
- **le genre** – Les normes sont calculées et des analyses statistiques sont effectuées pour annuler tout biais basé sur le sexe.

Tous les détails des normes et des analyses statistiques effectuées sont mentionnés dans les manuels techniques des tests pour une meilleure compréhension.

Conclusion

Le degré d'analyse théorique, empirique et statistique d'un test détermine la qualité de ses propriétés psychométriques et, en fin de compte, la qualité du développement d'une évaluation psychologique. Par conséquent, Central Test mesure ces critères avant qu'une évaluation ne soit publiée pour garantir sa pertinence et sa modernité.

NOUS CONTACTER

Service client : customer@centraltest.com

Formation et consulting : consulting@centraltest.com



central test

EMPOWERING TALENT

www.centraltest.fr

Copyright © 2022 par Central Test – Empowering Talent.

Tous droits réservés.