

MANUEL TECHNIQUE & GUIDE UTILISATEUR

SMART

Ce manuel a été réalisé par les consultants psychologues du département de Science et Innovation de Central Test.

Le test **SMART** et le logo Central Test sont des marques déposées.

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

Sommaire

1. Introduction	3
2. Description générale et domaines d'application	5
2.1. Domaines d'application principaux	5
2.2. Structure du questionnaire	6
2.3. Mode d'administration, langues et besoins spéciaux	6
2.4. Sécurité du test	7
2.5. Instructions de l'administration	7
2.6. Aspects techniques et sécurité	8
2.6.1. Respect de la charte éthique	8
2.6.2. Confidentialité et sécurité des données de l'utilisateur	8
2.6.3. Architecture matérielle et disponibilité des serveurs	9
2.6.4. Rapports générés par ordinateur	9
2.7. Utilisation, conditions et disponibilité	9
3. Cadre conceptuel et développement	11
3.1. Histoire de l'intelligence et des tests d'aptitude	11
3.2. Les tests d'aptitudes au travail	12
3.3. Histoire des tests adaptatifs informatisés	12
3.4. Une banque d'items calibrés	12
3.5. Modèle de la réponse à un item	13
3.5.1. Paramètres des items	13
3.5.1.a. Difficulté des items et discrimination des items	14
3.5.2. Avantages et études de validation	14
4. Les modules de SMART	16
4.1. Raisonnement logique	16
4.1.1. Développement	16
4.1.2. Normes	17
4.1.3. Fidélité	19
4.1.3.a. Fidélité interne	19
4.1.4. Paramètres des items	20
4.2. Raisonnement numérique	22
4.2.1. Développement	22
4.2.2. Normes	23
4.2.3. Fidélité	25

4.2.3.a. Fidélité Interne	25
4.2.4. Paramètres des items	26
4.3. Raisonnement verbal	28
4.3.1. Développement	29
4.3.2. Normes	29
4.3.3. Fidélité	31
4.3.3.a. Fidélité interne	31
4.3.4. Paramètres des items	32
5. Score	34
5.1. Rang percentile et méthode d'évaluation	34
5.2. Score combiné	34
Références	35

1. Introduction

SMART est un outil d'évaluation adaptatif conçu pour mesurer les capacités de raisonnement d'un individu. Contrairement aux tests de raisonnement traditionnels qui impliquent généralement que les candidats reçoivent le même ensemble de questions - certaines plus difficiles et d'autres plus faciles - et qui comparent ensuite les performances des individus, une évaluation comme **SMART** fournit une représentation plus précise des compétences de la personne, en raison de sa nature adaptative. Cela signifie que les paramètres de mesure de **SMART** s'ajustent en fonction des réponses de l'utilisateur tout au long du test.

En tant que test adaptatif, **SMART** évalue chaque individu aux limites de ses capacités. Lorsque les individus obtiennent des bonnes réponses, ils sont confrontés à des questions de plus en plus difficiles. Lorsqu'ils obtiennent des mauvaises réponses, des questions moins difficiles leur sont présentées. De cette façon, un test adaptatif tel que **SMART** nous permet d'affiner le niveau de capacité de la personne évaluée.

Le développement et l'administration d'un test adaptatif est plus difficile qu'un test traditionnel sous forme fixe. Cela nécessite une coordination et une expertise beaucoup plus spécialisée de la technique et de l'analyse psychométrique. Cependant, les tests adaptatifs tels que **SMART** offrent des avantages distincts qui en valent la peine.

Le fait que **SMART** s'adapte aux candidats signifie que leur expérience est unique et proche de leur niveau de capacité. Ils seront rarement confrontés à des items qui risqueraient de les frustrer par leur niveau de difficulté trop élevé ou de leur faire perdre leur temps avec des items trop simples. Cette personnalisation signifie que les participants auront une expérience plus positive du processus d'évaluation.

Cette expérience sur mesure signifie également que l'intégrité et la sécurité des questions du test pourront être maintenues plus longtemps, par rapport aux tests sous forme fixe où chaque participant voit chaque question. De plus, nous pouvons continuer à ajouter de nouveaux items au fil du temps, améliorant ainsi la précision et la sécurité du test global.

SMART est une version mise à jour et sous format adaptatif du *Test de Raisonnement de Central Test*.

Voici les avantages dont bénéficie un utilisateur de **SMART** :

- Les items sont précis, représentatifs, cohérents, clairs, équitables et sans culture.
- Les items, de différents niveaux de difficulté, sont calibrés de manière à s'adapter au modèle de réponse du candidat dès lors que celui-ci commence à répondre.
- Les candidats sont moins susceptibles de se sentir frustrés par la passation d'items trop difficiles, ou d'avoir l'impression de perdre leur temps à passer des items bien trop simples pour leur niveau de compétence.
- Un score est obtenu en fonction du niveau de "difficulté" auquel le candidat appartient, selon le test.
- Tout cela a été développé à l'aide d'une technique psychométrique de pointe, la théorie de la réponse aux éléments.

- Des rapports faciles à utiliser et personnalisables qui peuvent facilement être adaptés par l'administrateur.

Dans ce manuel et guide d'utilisateur, vous trouverez des informations utiles à propos du développement, du contexte théorique, des propriétés psychométriques ainsi que des applications pratiques du test **SMART**.

2. Description générale et domaines d'application

Wechsler (1944) décrit l'intelligence comme "une capacité globale d'agir selon une intention, de penser rationnellement et d'agir efficacement sur son environnement". **SMART** évalue les capacités intellectuelles globales et spécifiques d'un individu dans une gamme de domaines spécifiques. De plus, les fondements techniques de **SMART** permettent l'ajout de tâches de raisonnement supplémentaires si nécessaire.

2.1. Domaines d'application principaux

SMART a été spécialement conçu afin d'aider les professionnels des ressources humaines ou les managers à recruter, sélectionner et développer des employés à travers le monde. Il peut être utilisé tout au long de la carrière d'un employé mais convient également aux débutants. Le tableau 1, ci-dessous, détaille les différentes applications de **SMART**.

Tableau 1. Applications de **SMART**.

Recrutement	Mobilité Interne	Sélection
Pour recruter des individus dotés de capacités intellectuelles élevées, ce qui conduira à un meilleur retour sur investissement pour votre organisation.	Pour identifier les capacités de raisonnement de vos employés, vous permettant ainsi de prédire où ils seront les plus susceptibles d'exceller, et valider leur potentiel de réussite sur les tâches et projets futurs.	Donner une image claire de l'aptitude et de la capacité d'apprentissage d'un étudiant, ce qui aide à sélectionner les candidats de votre université ou école de commerce.

SMART peut également être utilisé conjointement avec d'autres évaluations professionnelles afin d'acquérir un profil plus complet de personnes occupant des rôles spécifiques. Par exemple, combiner ce test avec **Profil Pro 2©** offre un profil plus précis et exhaustif de l'individu, comme le montre le tableau 2 ci-dessous.

Tableau 2. Exemple d'application de **SMART** avec une autre évaluation.

Test additionnel	Objectif	Avantage
Profil Pro 2©	Développement de carrière	Informations détaillées à propos des capacités intellectuelles et de la personnalité, adaptées à une carrière spécifique.

2.2. Structure du questionnaire

SMART, une évaluation adaptative par ordinateur, est composée d'items sélectionnés d'après une banque d'items. Les items sont choisis afin de correspondre au niveau estimé d'aptitude de raisonnement du candidat. Si celui-ci répond correctement à un item, un item légèrement plus difficile lui sera ensuite présenté, et vice-versa, s'il échoue, un item moins difficile se présentera alors à lui.

2.3. Mode d'administration, langues et besoins spéciaux

SMART est administré en ligne et fonctionne avec tout système d'exploitation informatique moderne comprenant Windows, MAC-OS ou encore Linux. Comme d'autres questionnaires modernes, l'évaluation peut se faire à distance et sans supervision. Cependant, si des groupes d'évaluations standardisées sont requis, cela peut facilement être mis en place. Les rapports sont uniquement générés en ligne.

Le questionnaire est actuellement disponible en anglais, français et espagnol.

SMART est conçu pour être utilisé facilement par des personnes d'horizons diverses et par des personnes ayant des besoins particuliers. Tous les candidats peuvent ajuster les spécifications de l'écran visuel en fonction des exigences individuelles. Par exemple, les répondants ayant une déficience visuelle ou nécessitant une taille de police plus grande pour plus de confort, peuvent augmenter la taille de la police en élargissant l'affichage de l'écran de l'ordinateur (en appuyant sur "CTRL" et "+"). De même, les répondants peuvent accéder aux instructions de test tout au long de l'évaluation.

Les administrateurs sont encouragés à contacter l'éditeur de test pour obtenir des conseils sur l'administration du questionnaire aux individus ou aux groupes ayant des exigences particulières.

2.4. Sécurité du test

Un test adaptatif comme **SMART** est très résistant à la compromission. La probabilité que deux personnes passent le même test est très faible, il n'est donc pas possible de pré-apprendre les réponses ou de copier à partir d'un autre candidat.

La sécurité du test est améliorée en rendant la sélection des items aléatoire et en utilisant les performances du candidat face à ces items pour estimer sa capacité de raisonnement. Cela permet également de s'assurer qu'un large éventail d'items est utilisé, ce qui améliore encore la sécurité du test.

La randomisation évite également que les candidats répondent aux mêmes items. Cela permet d'égaliser l'utilisation de la banque d'items.

De plus, en raison du chevauchement, les difficultés des items peuvent être comparées et contrôlées, garantissant ainsi une évaluation équitable des performances pour tous les candidats (Chae et al., 2000).

2.5. Instructions de l'administration

Dans la plupart des cas, l'évaluation sera effectuée à distance via un accès en ligne et sans supervision. Cependant, dans le cas d'une évaluation dans un cadre supervisé, les directives suivantes doivent être appliquées.

Bien que les instructions pour remplir **SMART** soient présentées à l'écran, il est important que les administrateurs communiquent avec les candidats. Avant de commencer l'évaluation, les administrateurs doivent vérifier que les candidats comprennent ce qu'ils doivent faire et qu'ils ont la possibilité de demander des éclaircissements.

Les instructions de test sont les mêmes pour chaque individu. Des instructions normalisées garantissent que tous les participants sont exposés à des conditions similaires au cours du processus d'évaluation, afin de garantir qu'elles soient équitables et objectives et qu'elles ne donnent pas directement ou indirectement un avantage à un individu ou à un groupe.

Autres conseils :

- Si les candidats ont des questions après le début de la session d'évaluation, l'administrateur doit essayer d'y répondre sans déranger les autres candidats.
- Si un candidat ne comprend pas une question, l'administrateur doit répondre soigneusement et éviter d'expliquer le sens des mots ou des questions dans ses propres mots afin de ne pas influencer le candidat.
- Si un administrateur soupçonne que le manque de compréhension d'un candidat est dû à la maîtrise de la langue, il doit demander au candidat de passer l'évaluation dans la langue avec laquelle il est le plus à l'aise.

Les candidats reçoivent une notification pour répondre au questionnaire par e-mail avec un lien Web intégré et des informations de connexion. Les liens Web vers le questionnaire sont protégés par un mot de passe et sont propres à chaque compte candidat. Il est possible de fixer des dates d'expiration afin de s'assurer que les candidats remplissent le questionnaire dans le délai spécifié. Il est possible de configurer l'invitation de façon à ce que l'évaluation soit effectuée avec la même adresse IP que celle utilisée pour accéder au questionnaire. Le temps mis pour finir le questionnaire est également enregistré à des fins de suivi pour permettre aux administrateurs de voir si un candidat s'est précipité dans le questionnaire.

2.6. Aspects techniques et sécurité

Ce chapitre détaille les directives régissant la sécurité de l'administration de **SMART**. Ces directives garantissent que chaque administration de **SMART** est un processus agréable, objectif et équitable. Les spécifications de sécurité garantissent que les informations sur les candidats sont protégées pendant l'administration et après la fin de l'évaluation.

2.6.1. Respect de la charte éthique

En tant qu'éditeur d'outils d'évaluation en ligne, Central Test respecte les conventions légales du droit du travail et les dispositions réglementaires résultant de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés : Déclaration N ° 1834067.

De plus, depuis le 25 mai 2018, Central Test est conforme aux réglementations RGPD (Règlement Général de Protection des Données) édictées par le Parlement européen et le Conseil.

Les résultats des tests et les données personnelles ne peuvent être partagés avec des tiers sans le consentement explicite du candidat. Les candidats ont le droit d'accéder à leurs résultats et l'évaluateur doit leur accorder l'accès sur demande. De plus, si le candidat le demande, toutes les données doivent être définitivement supprimées ; pour ce faire, le candidat doit contacter directement l'organisation ou Central Test.

2.6.2. Confidentialité et sécurité des données de l'utilisateur

Central Test garantit la sécurité et la confidentialité des données générées lors de l'administration des tests en ligne. Les codes d'accès sont cryptés dans la base de données et ne sont accessibles à personne. Les mots de passe sont générés par calcul et ne sont pas visibles pour les entreprises. Un module de réécriture d'URL (générateur de lien sécurisé) est installé et a un passage de paramètre transparent pour l'utilisateur.

La navigation et l'échange de données via l'application Central Test peuvent être sécurisés et cryptés à l'aide du protocole SSL. La navigation sécurisée est garantie par le protocole https, qui est représenté par une icône en forme de cadenas sur le navigateur internet. L'application Central Test possède une signature numérique «GeoTrust SSL True BusinessID», émise et certifiée par GeoTrust ©. Ce certificat est une preuve irréfutable de notre attitude sérieuse

envers la sécurité et protège toutes les évaluations de Central Test des attaques de phishing. Les invitations à passer des tests se font via des liens sécurisés à usage limité (en nombre et en durée). Les tests sont chronométrés et les adresses IP sont suivies. L'accès aux questionnaires peut être limité en termes de nombre et de durée.

2.6.3. Architecture matérielle et disponibilité des serveurs

L'architecture de Central Test comprend un réseau de 10 serveurs physiques dédiés fournissant un service de haute qualité.

Chaque serveur est hébergé dans un centre de données spécialisé comprenant :

- Des lignes dédiées
- Une bande passante garantie de 100 Mbps (par serveur)
- Des sources d'énergie redondantes
- Un temps de restauration de l'équipement de 4 heures
- Un temps d'intervention sur incident de 2 heures
- Un suivi et une assistance 24 heures sur 24, 7 jours sur 7, 365 jours par an.

L'accès à l'administration du serveur est limité à une adresse IP et est sécurisé par mot de passe. Les échanges de données sont sécurisés.

L'application de Central Test a été testée et est compatible avec les navigateurs suivants :

- Internet Explorer 8 ou supérieur
- Firefox 21 ou supérieur
- Google Chrome 30 ou supérieur
- Safari 7 ou supérieur

2.6.4. Rapports générés par ordinateur

Les rapports peuvent être générés quelques minutes après que les candidats ont rempli le questionnaire. Le rapport **SMART** est simple d'utilisation et personnalisable. Le contenu du rapport peut être personnalisé en fonction des besoins de l'administrateur. Des rapports séparés peuvent être configurés pour le candidat et l'entreprise, le recruteur, le coach ou le psychologue. Les rapports sont disponibles sous format PDF.

2.7. Utilisation, conditions et disponibilité

Les conditions d'utilisation et la disponibilité du test **SMART** dépendent des pays. Cependant, les points suivants s'appliquent de manière générale :

- **SMART** est uniquement disponible pour les psychologues, les praticiens des RH et les coaches ;
- La formation est obligatoire avant l'utilisation et la distribution ;

- Ce manuel est disponible pour les personnes qui ont été formées à l'utilisation de **SMART** ;
- **SMART** peut être commandé en ligne ou en contactant un représentant du service client de Central Test ou un partenaire local dans le monde entier (customer@centraltest.com) ;
- Pour plus d'informations sur les prix de **SMART** et les produits et services associés, contactez un représentant du service client de Central Test ou un de ses partenaires locaux ;
- Pour plus d'informations sur la formation, veuillez contacter le service client de Central Test ou l'un de ses partenaires.

3. Cadre conceptuel et développement

3.1. Histoire de l'intelligence et des tests d'aptitude

L'intelligence humaine nous permet d'apprendre de notre environnement et de nos expériences, de résoudre des énigmes, de comprendre des notions abstraites et de comprendre les informations qui nous sont données. La fiabilité des mesures de l'intelligence fait partie des questions les plus étudiées en psychologie. L'une des origines de ces études date de 1869 avec le travail de Galton, au cours duquel il a utilisé un tableau de tests primitifs pour mesurer la motricité d'un grand échantillon de personnes. Son travail a mené au développement du test d'aptitude tel qu'il est maintenant utilisé en psychométrie

À la suite de Galton, en 1904, Charles Spearman a conçu une théorie à deux facteurs pour définir l'intelligence, en utilisant une technique statistique appelée « analyse factorielle ». Cette technique est dérivée des constatations selon lesquelles une personne qui réussit bien à un test de capacité mentale a également tendance à bien performer aux autres, tandis que celles qui réussissent mal à un test ont également tendance à ne pas réussir aussi bien aux autres. Par conséquent, la théorie de Spearman examine les tendances des scores d'un individu tout au long d'une évaluation et propose deux facteurs sous-jacents qui pourraient expliquer les différences individuelles au sein des scores : le facteur « général » (d'intelligence générale) et le facteur « spécifique » (intelligence liée spécifiquement au test en question). Cette idée est maintenant communément appelée « capacité mentale générale » et est une construction largement utilisée à ce jour, avec des mesures telles que les scores de QI.

Thurstone (1938) a cependant déclaré qu'il existerait en fait sept facteurs sous-jacents à la capacité mentale, nommés de la façon suivante : 1. Compréhension verbale, 2. Maîtrise verbale, 3. Nombre, 4. Visualisation spatiale, 5. Raisonnement inductif, 6. Mémoire et 7. Vitesse de perception.

Bien qu'il y ait encore un débat sur la définition exacte de la capacité mentale, il existe un large consensus parmi les chercheurs, selon lequel l'intelligence peut être définie comme la capacité de traiter et de comprendre de nouvelles informations. Cela implique d'appliquer des règles afin de trouver des modèles parmi les informations, et donc de résoudre des énigmes. Appelée « intelligence fluide » par Raymond Cattell en 1963, cette capacité ne serait pas affectée par des facteurs environnementaux tels que la qualité de l'éducation. Elle diffère de ce que Cattell appelait « intelligence cristallisée » qui réfère à des connaissances acquises précédemment, qui, en revanche, sont affectées par des facteurs environnementaux. L'intelligence fluide est largement mesurée par des tests mesurant les capacités de raisonnement logique et abstrait. L'intelligence cristallisée, d'autre part, est souvent mesurée à l'aide de tests d'aptitude verbaux et numériques.

Les tests d'aptitude permettent une mesure objective de l'aptitude d'un individu sur des aspects tels que les capacités de raisonnement verbal, numérique et logique, permettant ainsi une méthode d'évaluation de l'intelligence de l'individu.

3.2. Les tests d'aptitudes au travail

Dans le processus de recrutement, les évaluations d'aptitudes sont essentielles. Ils permettent au recruteur de déchiffrer objectivement si un candidat est capable de comprendre et d'exécuter un ensemble de compétences spécifiques au poste. En raison de leur nature objective, les tests d'aptitude permettent également de comparer les candidats entre eux, en examinant leurs scores.

3.3. Histoire des tests adaptatifs informatisés

Les tests adaptatifs informatisés ont été développés en grande partie dans les années 1970 et 1980, car ils sont devenus accessibles pour traiter et analyser psychométriquement de grands ensembles de données sur ordinateur (Weiss, 1983). Le but principal des tests adaptatifs informatisés est de prendre en compte un large éventail de niveaux de capacité de l'auditoire participant à l'évaluation, et de s'adapter à ces niveaux en sélectionnant dans une banque d'articles, des éléments de test qui sont appropriés au niveau de trait de l'utilisateur (Meijer, 1999).

Le test de dépistage adaptatif informatisé ou *Computerised Adaptive Screening Test*, en anglais, (Gade, Johnson & Pliske, 1984) a ensuite été mis en œuvre pour être utilisé par l'armée, dans le but d'estimer avec précision la capacité d'un candidat et d'aider à déterminer si les candidats avaient besoin de tests supplémentaires. MicroCAT a également été introduit en 1984 par la *Assessment Systems Corporation*, offrant aux utilisateurs un moyen de s'auto-évaluer et administrant des banques de tests informatisés (Rocklin, Roos, Wise et Yoes, 1996).

Offrant des avantages tels qu'une réduction de la fatigue de l'utilisateur, des coûts, du temps et des données plus précises, les tests adaptatifs informatisés continuent d'être largement utilisés de nos jours.

Thompson et Weiss (2011) présentent les cinq composantes proposées pour les constituer : 1. Une banque d'articles calibrés ; 2. Un point de départ ; 3. Un algorithme de sélection d'articles ; 4. Un algorithme de notation et 5. Un critère de terminaison.

3.4. Une banque d'items calibrés

La banque d'items à partir de laquelle les questions du test sont choisies est appelée la "banque d'items calibrés". Cela signifie que les paramètres définis de chaque item sont stockés sur l'ordinateur. Un point important lors de la création des items est de constituer une banque aussi grande que possible (Flaughner, 2000), afin d'assurer la fiabilité des paramètres générés.

Les paramètres des items ont été déterminés par une procédure de prétest et d'étalonnage des scores tests (*test equation*), par laquelle de plus petites listes contenant à la fois des items « uniques » et « communs » ont été administrés à un large échantillon de participants. Les paramètres de difficulté des items ont été calculés pour chacune des listes. Une fois les paramètres calculés, des éléments d'ancrage ont été utilisés pour dériver des constantes

d'équation pour chaque liste via des procédures d'équation moyenne / sigma (Kolen et Brennan 1995).

3.5. Modèle de la réponse à un item

Le Théorie des Réponses aux Items ou "item response theory" (IRT) a d'abord été développée dans le champ de la psychométrie pour la notation d'évaluation mesurant les capacités et a pris de plus en plus d'ampleur durant les dernières décennies. Ce modèle est désormais très largement utilisé au détriment de la théorie classique des tests ou "classical test theory" (CTT) pour l'évaluation des compétences et de la personnalité (Crocker & Algina, 1986). La théorie se base sur une large famille de modèles statistiques donnant une explication à la relation entre les traits latents (caractéristiques non-observables) et les résultats observables. En psychométrie, le modèle IRT a pour objectif d'établir un lien entre les items d'un test, les réponses aux items du test et les traits sous-jacents (comme la capacité de raisonnement numérique).

Dans le cadre de l'évaluation des compétences, le modèle IRT repose sur quatre hypothèses. La première, la monotonie, indique que lorsque le niveau d'un trait augmente, la probabilité d'une réponse correcte augmente. La seconde, l'unidimensionnalité, suppose qu'il existe un trait latent dominant mesuré, dont dépendent les réponses observées pour chaque item du test. Troisièmement, l'indépendance locale suppose que les réponses données à un item sont indépendantes du niveau de compétence. Finalement, l'hypothèse de l'invariance nous permet d'estimer le paramètre de l'item quelque soit sa position sur la courbe de réponses aux items. Ainsi, nous pouvons estimer le paramètre de chaque item depuis n'importe quel groupe de participants ayant répondu à l'item. Ces quatre hypothèses, si elles sont respectées, nous permettent de conclure que les différences observées dans les réponses correctes des participants seront liées aux différences de niveau du trait latent.

3.5.1. Paramètres des items

Le paramètre d'un item fait référence à ses caractéristiques et à sa position sur le continuum du trait latent, permettant au modèle IRT de prédire la réponse du participant, avec l'hypothèse que chaque réponse à un item délivre une indication du niveau du trait latent (θ). Ce niveau de compétence θ peut donc être défini comme la probabilité qu'un individu choisisse la bonne réponse à un item. Plus θ est élevé, plus il est probable qu'un individu choisisse la bonne réponse. En d'autres termes, pour chaque niveau de compétence, il existe une probabilité qu'un individu avec cette compétence puisse donner une réponse correcte à l'item.

Selon le modèle logistique à 2 paramètres (2PL) de la Théorie des Réponses aux Items, cette probabilité est donnée par la formule suivante :

$$P(\theta) = \frac{1}{1+e^{-L}} = \frac{1}{1+e^{-a(\theta-b)}}$$

Avec :

b paramètre de difficulté

a paramètre discriminant

$L = a(\theta - b)$ déviation logistique (logit)

θ le niveau de compétence.

Cette relation peut être représentée graphiquement en utilisant la courbe caractéristique d'un item (ICC). En application statistique, θ est compris entre -3 et 3 (voir figure 1).

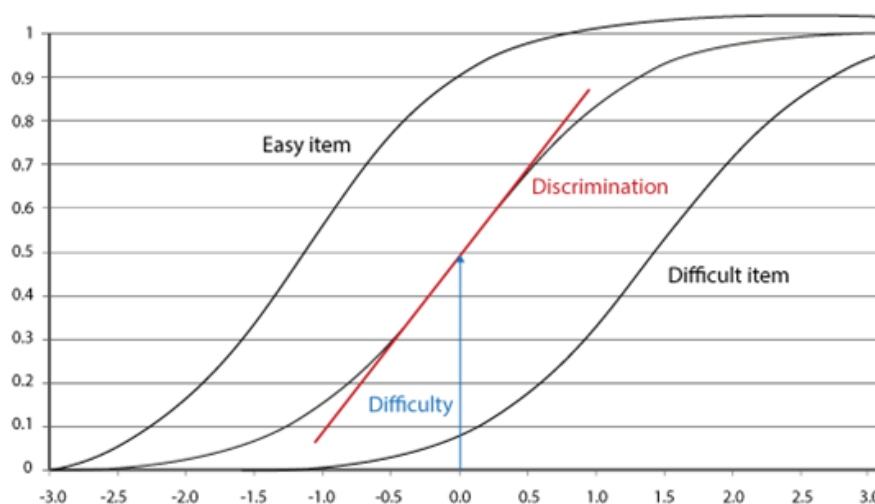


Figure 1. Courbe caractéristique d'un item pour un item facile, modéré, et difficile, d'un test de compétence.

3.5.1.a. Difficulté des items et discrimination des items

Sous le modèle 2PL de l'IRT, les courbes caractéristiques des items ont deux propriétés : la difficulté des items, et la discrimination des items. La difficulté des items indique où, sur l'échelle de compétence, fonctionne l'item. Elle est calculée au niveau de probabilité médian (la compétence pour laquelle 50% des candidats ont sélectionné la bonne réponse). Sur l'ICC, comme montré ci-dessus, les items difficiles sont orientés vers la droite de l'échelle, indiquant le plus haut niveau de compétence des individus qui ont répondu correctement à l'item. Les items les plus simples tendent vers la gauche de l'échelle.

La discrimination des items indique la fréquence avec laquelle la probabilité de choisir la bonne réponse change à un niveau de compétence donné (Baker, 2001). C'est particulièrement important lorsque nous comparons des individus qui montrent un niveau similaire sur un trait latent - en effet, des items plus discriminants sont nécessaires afin de placer plus précisément les individus le long du continuum du trait latent mesuré.

3.5.2. Avantages et études de validation

L'IRT a de nombreux avantages pour la précision et la fiabilité de la mesure en psychométrie. La principale caractéristique différenciant l'IRT des théories précédentes est que la fiabilité et l'erreur de mesure sont calculées et peuvent en conséquence être évaluées pour chaque item plutôt que pour le test dans sa globalité. Cela permet une plus grande précision dans la sélection des items et donc un nombre moindre d'items nécessaires à la construction du test (Thomas, 2011). Nous parlons dans ce cas de fonction d'*information sur les items* et de *fonction d'information sur le test*, respectivement pour les items et les tests.

L'un des avantages clé de l'IRT est qu'il permet d'effectuer des calculs pour des tests adaptatifs comme c'est le cas pour **SMART**. Les tests adaptatifs construits grâce à l'IRT fonctionnent en réalisant une estimation initiale du niveau de compétence du candidat, puis en choisissant et en présentant des items qui donneront le plus d'informations possible sur le niveau de compétence réel du candidat. Le test calcule une nouvelle estimation du niveau de compétence du candidat sur la base de ses réponses, et le test s'achève lorsque les critères de complétion sont satisfaits (par exemple une limite de temps ou de questions.)

Ceci permet d'obtenir des tests plus efficaces dans la mesure où le maximum d'informations relatives aux compétences de l'individu peut être obtenu dans un laps de temps plus court, réduisant le temps et la charge de travail associés à la passation des tests tout en renforçant leur fiabilité (Vispoel et al., 1999 ; Thomas, 2011).

Un autre avantage de l'IRT est l'hypothèse d'invariance des paramètres d'un item: les paramètres des items ne dépendent pas des compétences de l'individu, et les paramètres de l'individu ne dépendent pas des items qui leur sont présentés (Bortolotti et al., 2012). Cette invariance signifie que les paramètres se comportent de manière identique dans des populations différentes, permettant l'obtention d'une échelle de mesure uniforme, utilisable pour différents groupes d'individus et permettant d'évaluer la compétence du candidat en utilisant différentes formes d'un test ou différents sous-groupes d'items (Magno, 2009). En somme, ceci permet à l'IRT d'inférer les réponses des individus de manière plus efficace que d'autres théories existantes, comme la CTT.

Des études complémentaires mettent en exergue les avantages de l'IRT qui facilite une mise à l'échelle significative des paramètres individuels et des items. L'IRT permet une plus grande précision que d'autres techniques lors de la création d'échelles pour les paramètres des items et des individus, qui participent à notre compréhension de la relation entre les items et les traits latents (Ryan & Brockmann, 2009). La séparation des paramètres des items et des individus permet également une plus grande précision dans l'identification de la mesure DIF (Differential item functioning), car les différences de capacités au sein des groupes peuvent être écartées avec précision des biais du test, et les biais des tests peuvent être examinés plus en détail (Thomas, 2011).

4. Les modules de SMART

4.1. Raisonnement logique

Chaque fois qu'une personne découvre des liens entre divers items, des facultés logiques sont utilisées. C'est la capacité intellectuelle la plus fortement liée à la capacité mentale générale. Dans le milieu professionnel, de nombreuses activités reposent sur un sens de la logique très développé comme en droit, en informatique, en médecine, etc.

Il existe des preuves substantielles que les mesures du raisonnement et de la capacité mentale générale sont fortement corrélées à la performance au travail (Schmidt & Hunter, 2004). L'une des façons les plus efficaces et les mieux comprises d'évaluer la capacité mentale générale consiste à effectuer des tests de raisonnement abstrait, où les individus reçoivent des informations figuratives et doivent décider de la réponse qui correspond le plus précisément et le plus systématiquement au modèle. Une des raisons de la popularité et de l'utilisation généralisée de ces tests est qu'ils sont au mieux « sans culture », ou « culture free » en anglais (Cattell, 1971), et au pire moins influencés par des facteurs culturels que d'autres évaluations d'aptitude ou de capacité qui tendent à évaluer un domaine spécifique de capacités (ex : raisonnement numérique ou verbal).

Des niveaux plus élevés d'intelligence fluide sont associés à de meilleurs résultats concernant les tâches suivantes :

- Identifier des modèles et des tendances lorsque de nouvelles informations sont fournies ;
- Penser clairement et de manière rigoureuse à des idées complexes ;
- Tirer des conclusions logiques sur des problèmes ;
- Acquérir de nouvelles compétences et comprendre de nouvelles informations.

De plus, des niveaux inférieurs d'intelligence fluide signifient probablement qu'un individu aura des difficultés à :

- Tirer des conclusions correctes ou logiques ;
- Soutenir les décisions avec des preuves ou des justifications cohérentes ;
- Comprendre les relations complexes ;
- Identifier les problèmes et comprendre leurs causes.

4.1.1. Développement

En développant les items de **SMART – LOGICAL**, l'équipe de recherche et d'innovation a suivi les lignes directrices établies par Thompson et Weiss (2011) qui sont les suivantes :

- Le développement d'une banque d'items ;
- Le prétest et le calibrage des items ;
- L'intégration dans le système des tests adaptatifs informatisés.

La première étape a nécessité le développement d'un grand nombre d'items mesurant le raisonnement abstrait. Comme l'exemple ci-dessous le présente (Figure 1), ces items se composent de matrices 3 x 3 carrés, chaque carré contenant un modèle de forme à l'exception du carré en bas à droite. L'agencement des formes varie à travers chaque carré selon un ensemble de règles. Lors de la réponse à un item, les individus sont tenus d'identifier lequel des modèles conviendra le mieux, compte tenu des informations fournies.

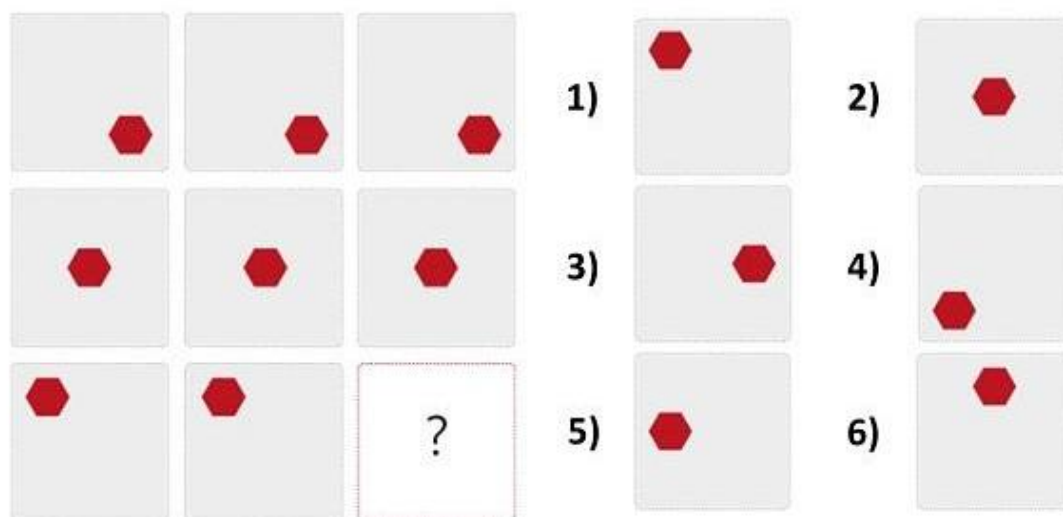


Figure 2. Exemple d'item de **SMART – LOGICAL**.

Une fois créés, il était nécessaire de tester les items afin de calculer les paramètres qui permettront leur sélection dans le test adaptatif. Étant donné le volume important d'articles, il n'était pas possible de tester l'intégralité de la banque à des particuliers. Au lieu de cela, les items d'ancrage ont été sélectionnés et administrés parallèlement à l'un de nos nombreux tests alternatifs. Ces items d'ancrage ont été sélectionnés avec un premier test préliminaire, afin de déterminer ceux qui présentaient une difficulté « moyenne ». La recherche indique que la sélection des items d'ancrage dans cette plage de difficulté est la plus appropriée lors de la réalisation d'études de liaison (Sinharay, Haberman, Holland et Lewis, 2012).

10 autres formes d'évaluation, ainsi que des items d'ancrage, ont été administrés aux participants (N = 2137) du *Testing Center* du département *Science et Innovation* de Central Test. Les tests ont été administrés avec des instructions en français et en anglais, et l'analyse différentielle du fonctionnement des items n'a montré aucune différence dans les réponses des participants francophones ou anglophones. Parce qu'il n'y avait pas de différence dans la réponse aux items et parce que les items s'appuient sur des capacités de raisonnement non verbal, les paramètres IRT ont été calculés sur la base des deux langues regroupées.

4.1.2. Normes

Les normes aident à comparer le score d'un individu à un test avec les scores d'autres personnes ayant effectué la même évaluation, généralement dans des conditions similaires.

Ces personnes qui ont passé le test en même temps sont représentatives de la population à laquelle une évaluation est destinée et sont donc désignées comme le groupe de référence.

Avec les normes, les créateurs de tests et les psychologues sont capables d'établir des comparaisons entre les individus et de prendre des décisions plus objectives. Les normes aident également les gestionnaires des ressources humaines, les recruteurs et les coachs à déterminer l'aptitude d'un candidat à assumer un rôle par rapport aux spécifications du poste, à évaluer les capacités d'un candidat dans un éventail de compétences souhaitées, à identifier les déficits de compétences d'un employé pour adapter la formation au développement et à établir des paramètres objectifs de performance. De plus, les groupes de référence permettent d'attribuer du sens aux résultats des tests et de quantifier les performances, le niveau d'un attribut ou d'un trait d'un individu par rapport à ses pairs.

Afin de fournir des données normatives pour **SMART – LOGICAL**, les performances aux items communs aux administrations de test (items d'ancrage) ont été analysées séparément des performances sur les items uniques dépendants de la liste.

Du fait de la nature adaptative de **SMART - LOGICAL**, des normes ont été calculées pour les scores non convertis (de -3 à 3) dans les tableaux 3 à 5. Veuillez noter que les scores non convertis ne sont pas affichés sur le rapport, et sont détaillés dans le manuel afin de visualiser les différences au sein des paramètres estimés, comme mentionné dans le Chapitre 4.1.4.

*Tableau 3. Normes de **SMART – LOGICAL** en fonction du genre.*

Genre	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
Femmes	337	-0.76	1.01	4.7
Hommes	327	-0.69	1.02	5.4

*Tableau 4. Normes de **SMART – LOGICAL** en fonction du groupe d'âge.*

Groupe d'âge	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
18-25	178	-0.61	1.04	5.7
26-33	170	-0.63	1.00	5.7
34-41	137	-0.79	0.97	4.9

42-49	79	-0.78	0.97	4.6
50+	100	-0.97	1.06	3.9

Tableau 5. Normes de **SMART – LOGICAL** en fonction du niveau d'études.

Niveau d'études	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
Niveau 1 - Diplôme de 2e ou 3e cycle universitaire / enseignement supérieur	50	-0.55	1.01	6.3
Niveau 2 - Diplôme de 1er cycle universitaire / enseignement supérieur	133	-0.62	5.1	0.87
Niveau 3 - Diplôme d'études professionnelles ou techniques/ BTS / DUT	53	-0.67	5.3	1.01
Niveau 4 - BAC / Diplôme d'études secondaires	43	-0.91	4.5	1.17

4.1.3. Fidélité

La fidélité mesure à quel point une évaluation est exempte d'erreurs et le degré auquel les items d'échelle mesurent le même construit. Lors de l'élaboration et de l'utilisation des évaluations, une préoccupation majeure est l'influence des erreurs de mesure aléatoires sur les performances d'un test. Si un test n'est pas fidèle, il est impossible qu'il soit valide, car la précision des résultats (validité) dépend de la cohérence (fidélité) avec laquelle les mêmes résultats sont produits. La fidélité permet aux utilisateurs de tests d'être sûrs que les résultats reflètent fidèlement l'attitude, les aptitudes ou le comportement du candidat au travail. Dans le cadre d'une évaluation basée sur le travail, la fidélité aide à garantir aux utilisateurs du test que les scores obtenus reflètent la manière dont un candidat ou un employé est susceptible de se comporter sur le lieu de travail, d'interagir avec les autres et de travailler sur des tâches spécifiques. Si une évaluation est très fidèle, elle permet également de se prémunir contre les biais qui peuvent être causés par des erreurs de jugement humain lors du processus de recrutement et de sélection.

4.1.3.a. Fidélité interne

La fidélité interne indique la mesure dans laquelle les éléments d'une échelle sont cohérents en interne et mesurent tous le même construit. Elle est mesurée en calculant l'alpha (α) de Cronbach, pour chaque facteur. Celui-ci fait la moyenne de toutes les corrélations inter-items

possibles sans s'appuyer sur un seul indicateur de fidélité, qui peut contenir de grandes quantités d'erreurs. L'alpha de Cronbach varie de 0 à 1, avec une valeur minimale recommandée de 0,7 (1 étant égal à une fidélité parfaite (Nunnally, 1978)). Des valeurs supérieures à 0,9 sont considérées comme excellentes, de 0,7 à 0,9 sont considérées comme bonnes, de 0,6 à 0,7 sont acceptables, de 0,5 à 0,6 sont médiocres et des valeurs inférieures à 0,5 sont inacceptables.

La fidélité de **SMART – LOGICAL** a été calculée sur un échantillon (n=664) et montre une fidélité globale de **0,80** ce qui indique que l'échelle a une **bonne** cohérence interne et est très fiable.

4.1.4. Paramètres des items

De part les spécificités de la mesure des capacités de raisonnement logique, les paramètres des items pour **SMART – LOGICAL** ont été calculés en utilisant un modèle 1PL (ou modèle Rasch), ce qui signifie que les items ne sont caractérisés que par leur niveau de difficulté. Le niveau de difficulté des items montre comment les items fonctionnent dans l'échelle de compétence, entre -3 et 3 (comme mentionné au chapitre 3.5.1.a.). Des scores plus faibles indiquent qu'un item est facile, tandis qu'un item avec un score supérieur indique que l'item est difficile.

Table 6. Évaluation du niveau de difficulté des items de **SMART – LOGICAL**.

Question No.	Difficulté des Items	Question No.	Difficulté des Items	Question No.	Difficulté des Items
1	-1.19	35	0.519510854	68	0.326047841
2	-0.75	36	1.028810313	69	-0.857268049
3	1.04	37	1.419273231	70	-0.812881657
4	1.96	38	1.113693556	71	-0.049435725
5	0.85	39	-1.551640279	72	0.190250789
6	-3.65	40	0.366721016	73	-0.129331229
7	-0.03	41	1.478691502	74	0.39442819
8	-1.09	42	-0.22512786	75	0.021582502
9	1.13	43	-0.321358501	76	1.379806079

10	0.81	44	1.26207295	77	0.65186926
11	-1.5	45	-0.959980025	78	1.770406324
12	0.35	46	-0.942483545	79	-3.558688655
13	1.04	47	0.973381028	80	-0.987533607
14	1.04	48	-0.426337381	81	0.10564456
16	-0.501700281	49	0.579710225	82	-0.016791395
17	-0.234173829	50	-0.870144022	83	0.770296885
18	-0.121985316	51	-1.185987525	84	0.980187093
19	2.121784928	52	0.744167217	85	0.735315183
20	0.99126992	53	-0.062988402	86	0.875241989
21	-0.208284172	54	-0.856106533	87	0.131880836
22	0.741003239	55	0.659942283	88	0.770296885
23	0.775522781	56	1.102123188	89	-2.754087094
24	0.174006803	57	0.603792327	90	0.135341465
25	-1.571650387	58	0.807335918	91	-0.269883272
26	-2.25877077	59	-0.625138854	92	-1.106760446
27	-0.540969812	60	1.411519718	93	0.135341465
28	0.573279458	61	-1.71061073	94	0.276289199
29	2.272509595	62	0.348428705	95	0.478901567
30	-0.039557641	63	-0.815376193	96	1.606483444
31	1.761812013	64	1.411519718	97	1.113166373

32	1.186116557	65	1.288424969	98	0.619849302
33	-2.035474765	66	0.986283313	99	1.452916318
34	-1.016875847	67	2.351515981		

4.2. Raisonnement numérique

Peu de professions sont exemptes d'exigences numériques. Les emplois diffèrent dans la proportion dans laquelle les individus doivent manipuler les nombres et la complexité des inférences qu'ils sont censés faire à partir des données numériques.

SMART – NUMERICAL est un test de raisonnement, mesurant la capacité d'une personne à manipuler et à raisonner à l'aide d'informations numériques. Plutôt que d'évaluer si un individu peut effectuer des calculs numériques mentalement, il se concentre plutôt sur sa capacité à suivre les étapes appropriées lors de la transformation des nombres.

Des niveaux supérieurs d'intelligence numérique sont associés à :

- Une manipulation précise et efficace des ensembles de données numériques ;
- Une amélioration de la confiance face aux nombres ;
- Une meilleure identification des erreurs dans la pensée mathématique ;
- Une compréhension et interprétation des résultats numériques.

A l'inverse, des niveaux inférieurs d'intelligence numérique suggèrent qu'un individu a probablement des difficultés à :

- Effectuer des calculs et manipuler des séries de données ;
- Approcher en toute confiance des tâches qui nécessitent une analyse des données ou des transformations mathématiques ;
- Travailler avec précision et efficacité avec des données numériques ;
- Comprendre les implications des résultats numériques.

4.2.1. Développement

En développant les items de **SMART – NUMERICAL**, l'équipe *Science et innovation* a suivi les lignes directrices énoncées par Thompson et Weiss (2011). En résumé, ce sont :

- Le développement d'une banque d'items ;
- Le prétest et calibrage des items ;
- L'intégration dans le système des tests adaptatifs informatisés.

La première étape a nécessité le développement d'un grand volume d'items mesurant le raisonnement numérique. Comme il est possible de voir sur l'exemple présenté ci-dessous

(Figure 2), ces items consistent en des problèmes mathématiques nécessitant la conversion de nombres d'une échelle à une autre (comme le temps, la distance, les devises, etc.).

0 Après avoir commandé 200 cartes de visite, Adrien a remarqué que 50 d'entre elles n'étaient pas correctement imprimées. Il avait initialement payé 29€ pour les 200 cartes mais on lui promet un remboursement pour chaque carte mal imprimée. Quelle somme va-t-il récupérer ?
a) 9,25€
b) 14,50€
c) 9,67€
d) 7,25€

Figure 3. Exemple d'item de **SMART – NUMERICAL**.

Une fois créés, il était nécessaire de tester les items afin de calculer les paramètres qui permettront plus tard leur sélection dans le test adaptatif. Étant donné le volume important d'items, il n'était pas possible d'administrer l'intégralité de la banque à des particuliers. Au lieu de cela, les items d'ancrage ont été sélectionnés et administrés parallèlement à l'un des nombreux tests alternatifs. Ces items d'ancrage ont été sélectionnés avec un premier test préliminaire, afin de déterminer ceux qui présentaient une difficulté « moyenne ». La recherche indique que la sélection des items d'ancrage dans cette plage de difficulté est la plus appropriée lors de la réalisation d'études de liaison (Sinharay, Haberman, Holland et Lewis, 2012).

Une fois que les items de difficulté "moyenne" ont été sélectionnés, 10 autres formes d'évaluation, ainsi que des items d'ancrage, ont été administrés aux participants (N = 2019) du *Testing Center* du département *Science et Innovation* de Central Test. Les tests ont été administrés avec des instructions en français et en anglais. Les données qui en sont issues ont ensuite été utilisées pour calculer les paramètres IRT de chaque item.

4.2.2. Normes

Les normes aident à comparer le score d'un individu à un test avec les scores d'autres personnes ayant effectué la même évaluation, généralement dans des conditions similaires. Ces personnes qui ont passé le test en même temps sont représentatives de la population à laquelle une évaluation est destinée et sont donc désignées comme le groupe de référence.

Avec les normes, les créateurs de tests et les psychologues sont capables d'établir des comparaisons entre les individus et de prendre des décisions plus objectives. Les normes aident également les gestionnaires des ressources humaines, les recruteurs et les coaches à déterminer l'aptitude d'un candidat à assumer un rôle par rapport aux spécifications du poste, à évaluer les capacités d'un candidat dans un éventail de compétences souhaitées, à identifier les déficits de compétences d'un employé pour adapter la formation au développement et à établir des paramètres objectifs de performance. De plus, les groupes de référence permettent

d'attribuer du sens aux résultats des tests et de quantifier les performances, le niveau d'un attribut ou d'un trait d'un individu par rapport à ses pairs.

Afin de fournir des données normatives pour **SMART – NUMERICAL**, les performances sur le sous-ensemble d'items communs aux administrations de test (items d'ancrage) ont été analysées séparément des performances sur les items uniques dépendants de la liste.

Du fait de la nature adaptative de **SMART – NUMERICAL**, des normes ont été calculées pour les scores non convertis (de -3 à 3) dans les tableaux 7 à 9. Veuillez noter que les scores non convertis ne sont pas affichés sur le rapport, et sont détaillés dans le manuel afin de visualiser les différences au sein des paramètres estimés, comme mentionné au Chapitre 4.1.4.

*Tableau 7. Normes de **SMART – NUMERICAL** en fonction du genre.*

Genre	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
Femmes	180	-0.54	1.24	0.52
Hommes	173	-0.62	1.42	0.49

*Tableau 8. Normes de **SMART – NUMERICAL** en fonction du groupe d'âge.*

Groupe d'âge	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
18-25	77	-0.80	1.43	4.2
26-33	67	-0.45	1.44	5.3
34-41	78	-0.47	1.30	5.4
42-49	59	-0.39	1.31	5.5
50+	72	-0.73	1.12	4.6

Tableau 9. Normes de **SMART – NUMERICAL** en fonction du niveau d'études.

Niveau d'études	n	Moyenne du score brut	Ecart-type	Moyenne du rang percentile
Niveau 1 - Diplôme de 2e ou 3e cycle universitaire / enseignement supérieur	73	-0.10	1.43	6.4
Niveau 2 - Diplôme de 1er cycle universitaire / enseignement supérieur	37	-0.27	1.23	6.2
Niveau 3 - Diplôme d'études professionnelles ou techniques/ BTS / DUT	60	-0.66	1.24	5.7
Niveau 4 - BAC / Diplôme d'études secondaires	76	-0.71	1.14	4.4

4.2.3. Fidélité

La fidélité mesure à quel point une évaluation est exempte d'erreurs et le degré auquel les items d'échelle mesurent le même construit. Lors de l'élaboration et de l'utilisation des évaluations, une préoccupation majeure est l'influence des erreurs de mesure aléatoires sur les performances d'un test. Si un test n'est pas fidèle, il est impossible qu'il soit valide, car la précision des résultats (validité) dépend de la cohérence (fidélité) avec laquelle les mêmes résultats sont produits. La fidélité permet aux utilisateurs de tests d'être sûrs que les résultats reflètent fidèlement l'attitude, les aptitudes ou le comportement du candidat au travail. Dans le cadre d'une évaluation basée sur le travail, la fidélité aide à garantir aux utilisateurs du test que les scores obtenus reflètent la manière dont un candidat ou un employé est susceptible de se comporter sur le lieu de travail, d'interagir avec les autres et de travailler sur des tâches spécifiques. Si une évaluation est très fidèle, elle permet également de se prémunir contre les biais qui peuvent être causés par des erreurs de jugement humain lors du processus de recrutement et de sélection.

4.2.3.a. Fidélité Interne

La fidélité interne indique la mesure dans laquelle les éléments d'une échelle sont cohérents en interne et mesurent tous la même construction. Elle est mesurée en calculant l'alpha (α) de Cronbach, pour chaque facteur. Celui-ci fait la moyenne de toutes les corrélations inter-items possibles sans s'appuyer sur un seul indicateur de fidélité, qui peut contenir de grandes quantités d'erreurs. L'alpha de Cronbach varie de 0 à 1, avec une valeur minimale recommandée de 0,7 (1 étant égal à une fidélité parfaite (Nunnally, 1978)). Des valeurs supérieures à 0,9 sont considérées comme excellentes, de 0,7 à 0,9 sont considérées comme

bonnes, de 0,6 à 0,7 sont acceptables, de 0,5 à 0,6 sont médiocres et des valeurs inférieures à 0,5 sont inacceptables.

La fidélité de **SMART – NUMERICAL** a été calculée sur l'échantillon anglophone (n = 291) et montre une fidélité globale de 0,79, ce qui indique que l'échelle a une bonne cohérence interne et est très fiable.

4.2.4. Paramètres des items

Selon le modèle 2PL, les paramètres des items représentent la discrimination et la difficulté des items. *La difficulté des items* montrent où les items de **SMART - NUMERICAL** fonctionnent sur l'échelle de compétence, entre -3 et 3. Des scores faibles indiquent qu'un item est facile, tandis que des scores forts indiquent qu'un item est difficile. *La discrimination des items* indique à quel point un item est en capacité de discriminer le niveau de compétence des candidats (comme mentionné au chapitre 3.5.1.a).

Table 10. Estimations des paramètres des items de **SMART – NUMERICAL**.

Question Numéro	Discrimination des items	Difficulté des items	Question Numéro	Discrimination des items	Difficulté des items
1	1.200470885	0.214463401	44	1.316653698	0.365350443
2	0.354090958	2.874429583	45	0.544254317	0.761824995
3	0.693992425	1.235416552	47	1.079591016	0.022704525
5	1.394108274	-1.663154958	49	1.22694701	0.549936967
6	1.059298181	-1.694216293	51	0.720241415	-1.848958503
7	1.733554835	-2.16342045	52	1.220652633	0.180278357
8	1.662591768	-0.940224244	53	0.862048985	2.531467834
9	1.123863122	-0.688112156	54	1.998005374	-0.98708771
10	2.18143024	-2.75464666	55	1.986282595	0.42518552
11	1.997682901	-1.922729418	56	0.770446964	-0.085624793
12	0.418602374	-0.589808755	57	0.97934297	1.094296796

13	1.086931642	-0.606619654	58	1.760015213	0.668320765
15	1.264340383	0.542478906	59	1.793652667	0.223925003
16	2.084487046	-1.069491494	60	0.638253834	0.551861979
17	0.842634593	0.276327304	62	1.13828527	1.764005911
18	1.653578765	-1.243420318	63	0.889514842	0.276085604
19	1.230073096	-0.427840458	64	0.622502694	3.068095974
20	1.061408145	-0.14325936	66	1.614723888	-0.364755181
21	1.182652926	1.314225834	67	1.238486888	-0.609627588
22	1.253577275	0.66165344	68	1.507085951	0.95198374
23	0.499678197	1.451503963	69	0.617735417	2.462698676
24	1.796098902	-1.008929376	70	2.383152784	-0.184513238
25	0.480533485	1.459023153	72	0.935946596	0.566906425
26	2.020718872	-0.320261381	74	1.41927911	-0.63347253
27	0.830352517	-1.105926454	75	0.979584683	0.005501288
29	1.728604675	-0.328032542	76	1.082823019	-0.086300184
30	0.7101766	0.317925246	77	0.861566778	0.639008427
32	1.392600031	-0.629143612	78	1.523214592	-0.921816024
33	1.132830733	-0.039567218	80	1.362978796	-0.714749965
34	0.979578365	1.174983081	81	1.058308427	-0.2450991
35	0.842863368	0.383319397	82	1.959730296	0.719501217
36	1.11821153	1.336359545	83	0.456942619	0.383834889

37	1.768498247	-0.713288282	84	1.068588906	0.49984615
38	1.743354344	-0.586805527	85	1.422881541	-0.060737635
39	1.37361501	0.849992612	86	0.509093157	0.688041235
40	1.391181942	0.18638908	87	0.870592497	0.980271824
41	0.646444066	1.690537112	88	1.757727149	-0.491504074
43	0.425898019	1.109006287	89	1.792077801	-0.878048831

4.3. Raisonnement verbal

SMART – VERBAL est un test d'habileté cognitive, mesurant la capacité d'une personne à manipuler et à raisonner à l'aide d'informations verbales. Plus précisément, l'évaluation teste le raisonnement « analogique », c'est-à-dire la capacité de raisonner par analogie. Les analogies exigent qu'un individu comprenne les relations logiques entre différents concepts et significations et distingue les similarités ou les différences.

Des niveaux plus élevés de raisonnement verbal sont associés à :

- De meilleures compétences en communication et efficacité accrue pour convaincre ou persuader les autres ;
- Une amélioration de la confiance lors du traitement d'arguments complexes ou de l'apprentissage de nouveaux concepts ;
- Une meilleure identification des erreurs dans le raisonnement verbal ou dans les arguments.
- Une compréhension et interprétation des résultats des arguments présentés verbalement.

De plus, des niveaux inférieurs d'intelligence verbale suggèrent qu'un individu pourrait avoir des difficultés à :

1. Communiquer efficacement ou convaincre et persuader les autres ;
2. Aborder avec confiance les tâches qui nécessitent la compréhension d'arguments complexes ou l'apprentissage de nouveaux concepts ;
3. Travailler avec précision et efficacité avec des informations verbales ;
4. Comprendre les implications des arguments présentés verbalement.

4.3.1. Développement

En développant les items de **SMART – VERBAL**, l'équipe de recherche et d'innovation a suivi les lignes directrices énoncées par Thompson et Weiss (2011). En résumé, ce sont :

- Le développement d'une banque d'items ;
- Le prétest et calibrage des items ;
- L'intégration dans le système des tests adaptatifs informatisés.

La première étape a nécessité le développement d'un grand volume d'items mesurant le raisonnement verbal. A l'instar de l'exemple présenté ci-dessous (Figure 3), ces items consistent en des analogies nécessitant la liaison d'éléments d'une structure verbale à un autre élément d'une autre structure verbale tout en appliquant deux principes.

0 A : Arbre, Jungle, Feuille B : Nature, Vache, Herbe

(A) est à forêt ce que (B) est à pré.

Figure 4. Exemple d'item de **SMART – VERBAL**.

Une fois créés, il était nécessaire de tester les items afin de calculer les paramètres qui permettront plus tard leur sélection dans le test adaptatif. Étant donné le volume important d'articles, il n'était pas possible d'administrer l'intégralité de la banque à des particuliers. Au lieu de cela, les items d'ancrage ont été sélectionnés et administrés parallèlement à l'un des nombreux tests alternatifs. Ces items d'ancrage ont été sélectionnés avec un premier test préliminaire, afin de déterminer ceux qui présentaient une difficulté « moyenne ». La recherche indique que la sélection des items d'ancrage dans cette plage de difficulté est la plus appropriée lors de la réalisation d'études de liaison (Sinharay, Haberman, Holland et Lewis, 2012).

Une fois les items de difficulté "moyenne" sélectionnés, 10 autres formes d'évaluation, ainsi que des items d'ancrage, ont été administrés aux participants (N = 2023) du *Testing Center* du département *Science et Innovation* de Central Test. Les données qui en sont issues ont ensuite été utilisées pour calculer les paramètres IRT de chaque items.

4.3.2. Normes

Les normes aident à comparer le score d'un individu à un test avec les scores d'autres personnes ayant effectué la même évaluation, généralement dans des conditions similaires. Ces personnes qui ont passé le test en même temps sont représentatives de la population à laquelle une évaluation est destinée et sont donc désignées comme le groupe de référence.

Avec les normes, les créateurs de tests et les psychologues sont capables d'établir des comparaisons entre les individus et de prendre des décisions plus objectives. Les normes aident également les gestionnaires des ressources humaines, les recruteurs et les coaches à

déterminer l'aptitude d'un candidat à assumer un rôle par rapport aux spécifications du poste, à évaluer les capacités d'un candidat dans un éventail de compétences souhaitées, à identifier les déficits de compétences d'un employé pour adapter la formation au développement et à établir des paramètres objectifs de performance. De plus, les groupes de référence permettent d'attribuer du sens aux résultats des tests et de quantifier les performances, le niveau d'un attribut ou d'un trait d'un individu par rapport à ses pairs.

Afin de fournir des données normatives pour **SMART – VERBAL**, les performances sur le sous-ensemble d'items communs aux administrations de test (items d'ancrage) ont été analysées séparément des performances sur les items uniques dépendants de la liste.

Du fait de la nature adaptative de **SMART – VERBAL**, des normes ont été calculées pour les scores non convertis (de -3 à 3) dans les tableaux 11 à 13. Veuillez noter que les scores non convertis ne sont pas affichés sur ce rapport, et sont détaillés dans le manuel afin de visualiser les différences au sein des paramètres estimés, comme mentionné au Chapitre 4.1.4.

*Tableau 11. Normes de **SMART – VERBAL** en fonction du genre.*

Genre	n	Moyenne du score brut	Écart-type	Moyenne du rang percentile
Hommes	165	-0.30	0.88	5.5
Femmes	209	-0.38	0.91	4.8

*Tableau 12. Normes de **SMART – VERBAL** en fonction du groupe d'âge.*

Groupe d'âge	n	Moyenne du score brut	Écart-type	Moyenne du rang percentile
18-25	102	-0.33	0.95	5.1
26-33	85	-0.26	0.95	5.6
34-41	64	-0.30	0.96	5.7
42-49	59	-0.48	0.75	4.0
50+	64	-0.40	0.79	4.6

Tableau 13. Normes de **SMART – VERBAL** en fonction du niveau d'études.

Niveau d'études	n	Moyenne du score brut	Écart-type	Moyenne du rang percentile
Niveau 1 - Diplôme de 2e ou 3e cycle universitaire / enseignement supérieur	74	-0.13	0.99	6.5
Niveau 2 - Diplôme de 1er cycle universitaire / enseignement supérieur	35	0.05	0.88	6.1
Niveau 3 - Diplôme d'études professionnelles ou techniques/ BTS / DUT	63	-0.33	0.79	5.8
Niveau 4 - BAC / Diplôme d'études secondaires	80	-0.39	0.75	4.8

4.3.3. Fidélité

La fidélité mesure à quel point une évaluation est exempte d'erreurs et le degré auquel les items d'échelle mesurent le même construit. Lors de l'élaboration et de l'utilisation des évaluations, une préoccupation majeure est l'influence des erreurs de mesures aléatoires sur les performances d'un test. Si un test n'est pas fidèle, il est impossible qu'il soit valide, car la précision des résultats (validité) dépend de la cohérence (fidélité) avec laquelle les mêmes résultats sont produits. La fidélité permet aux utilisateurs de tests d'être sûrs que les résultats reflètent fidèlement l'attitude, les aptitudes ou le comportement du candidat au travail. Dans le cadre d'une évaluation basée sur le travail, la fidélité aide à garantir aux utilisateurs du test que les scores obtenus reflètent la manière dont un candidat ou un employé est susceptible de se comporter sur le lieu de travail, d'interagir avec les autres et de travailler sur des tâches spécifiques. Si une évaluation est très fidèle, elle permet également de se prémunir contre les biais qui peuvent être causés par des erreurs de jugement humain lors du processus de recrutement et de sélection.

4.3.3.a. Fidélité interne

La fidélité interne indique la mesure dans laquelle les éléments d'une échelle sont cohérents en interne et mesurent tous la même construction. Elle est mesurée en calculant l'alpha (α) de Cronbach, pour chaque facteur. Celui-ci fait la moyenne de toutes les corrélations inter-items possibles sans s'appuyer sur un seul indicateur de fidélité, qui peut contenir de grandes quantités d'erreurs. L'alpha de Cronbach varie de 0 à 1, avec une valeur minimale recommandée de 0,7 (1 étant égal à une fidélité parfaite (Nunnally, 1978)). Des valeurs supérieures à 0,9 sont considérées comme excellentes, de 0,7 à 0,9 sont considérées comme bonnes, de 0,6 à 0,7 sont acceptables, de 0,5 à 0,6 sont médiocres et des valeurs inférieures à 0,5 sont inacceptables.

La fidélité de SMART - Verbal a été calculée sur l'échantillon francophone et montre une fidélité globale de 0,79 ce qui indique que l'échelle a une bonne cohérence interne et est très fiable.

4.3.4. Paramètres des items

Selon le modèle 2PL, les paramètres des items représentent la discrimination et la difficulté des items. La *difficulté des items* montre où les items de **SMART - VERBAL** fonctionnent sur l'échelle de compétence, entre -3 et 3. Des scores faibles indiquent qu'un item est facile, tandis que des scores forts indiquent qu'un item est difficile. La *discrimination des items* indique à quel point un item est en capacité de discriminer le niveau de compétence des candidats (comme mentionné au chapitre 3.5.1.a).

Table 14. Estimations des paramètres des items de **SMART – VERBAL**

Question Numéro	Discrimination de l'item	Difficulté de l'item	Question Numéro	Discrimination de l'item	Difficulté de l'item
1	1.027082473	0.252038292	49	0.339549466	-2.163052987
2	1.693056616	0.152188886	50	0.945007805	1.454578595
3	1.585186399	0.144500242	51	0.459870261	-3.85843976
4	0.958461024	0.375435271	52	0.459148322	3.799808528
5	0.831342235	0.902966652	53	1.367693294	1.904804066
7	1.407273171	0.022354129	54	0.847711126	1.068686733
8	0.831582352	-1.320025566	57	-0.250315813	2.375646415
9	1.619390138	-0.399940094	61	0.97562687	-2.043021455
10	0.937646638	-0.559981	62	1.130343245	-0.346029617
11	0.599837091	0.963461134	63	1.673422174	1.380770677
12	1.323162151	-1.187850656	64	1.235840531	3.351233929
13	0.902776918	3.456863664	66	0.870745381	-3.013347338
14	0.93947746	-0.212287812	67	0.499702784	-2.194416826
15	0.69961467	0.290043603	68	1.240667329	-1.137991805
16	0.601184535	1.745697945	69	0.296131478	-0.570020124
18	1.229916347	-1.40814677	70	1.540765061	0.531093208

19	1.830064869	-1.056455887	72	0.57531922	1.142851097
20	1.438411092	-0.675765086	73	1.565114291	0.434415473
21	0.755820555	-0.701658575	75	0.607225751	-2.953924656
22	0.205529859	0.925360402	76	0.693901386	-3.065734143
23	0.949633613	0.161182263	77	0.452559712	0.276480019
25	0.854406474	-1.261444779	78	1.473883523	-0.878448473
27	1.373671825	-1.879138312	79	1.385319224	-1.995434419
28	1.399861176	-0.660310869	80	0.242322872	-2.2028254
29	0.656097221	-2.002108382	81	0.788679783	2.985353285
30	1.584803696	0.139887537	82	0.527796836	1.624799524
31	1.35787989	-0.11871718	83	1.07074316	1.42628429
33	0.305062409	1.417964386	84	0.4230321	-3.10040447
34	0.482334222	-0.852574154	85	1.00089639	-0.574240244
35	0.382763112	3.995698078	87	0.467251122	-1.039008603
37	1.139766928	-2.31720733	88	1.01451942	0.262235675
38	0.92246214	-0.691054692	89	0.38372786	-2.983360573
39	1.63323708	-0.06632227	90	0.751511921	3.047297387
40	0.976739066	1.108839465	91	0.700126485	1.768805371
41	0.491183878	-1.285840385	94	1.357934822	-0.869338958
42	1.264247671	0.979491686	95	2.071849249	-0.677915171
44	0.956138882	-0.560430667	96	1.004803199	1.308177482
45	0.695139339	1.793256833	97	0.657365074	0.00395581
48	1.21107769	-0.948139057	98	0.82166786	-0.518626212

5. Score

5.1. Rang percentile et méthode d'évaluation

Afin de délivrer un score simple à interpréter, qui permet d'indiquer efficacement où le score de l'individu se situe comparé à la population normalisée, les scores du rapport de **SMART** sont générés en rang percentiles.

Les individus reçoivent un score entre 0 et 10, indiquant leurs rangs percentiles et illustrant la performance des candidats par rapport aux scores du groupe auquel ils sont comparés, permettant de ce fait d'établir des comparaisons significatives. Ceci est représenté visuellement sur un graphique affichant le rang percentile du candidat comme un pourcentage entre 0 et 100. Par exemple, un score de 7,4 équivaut à un rang percentile de 74%. Un rang percentile de 74% indique qu'un individu a eu un score supérieur à ceux de 74% du groupe de comparaison, mais cela ne signifie pas que la personne a obtenu 74% de réussite au test. Les commentaires de chaque module de **SMART** suivent cette méthode d'évaluation, donnant une interprétation du niveau de compétence en raisonnement à travers 4 niveaux. Par exemple, les rangs percentiles pour les scores faibles seront accompagnés de commentaires indiquant que l'individu pourrait avoir des difficultés à réaliser des tâches requérant la compétence de raisonnement associée, tandis que des rangs percentiles élevés sont accompagnés de commentaires indiquant que l'individu a un niveau avancé de la compétence en raisonnement.

5.2. Score combiné

Quand un individu passe les trois modules de **SMART** adaptatif, ils reçoivent un score combiné sur leur talent map, également appelé "score d'intelligence générale". Pour calculer ce score combiné, une pondération de chaque module de **SMART** est calculée. Les pondérations sont évaluées grâce à une analyse factorielle conduite sur un échantillon de participants ayant passé tous les modules. Cette analyse factorielle nous permet de déterminer le facteur g de chaque module, c'est-à-dire le niveau de corrélation de chaque module de raisonnement avec le facteur g d'intelligence.

Dans le cas où nous n'avons pas assez de données pour conduire une analyse factorielle, la littérature peut être utilisée pour déterminer le facteur g de chaque module.

Table 15. Facteur g actuel, pour chacune des compétences en raisonnement mesurées par SMART.

Capacité de raisonnement	Charge g
Numerical	0.44
Verbal	0.56

Références

- Baker, F. (2001). *The Basics of Item Response Theory (2nd ed.)*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Bortolotti, S., Tezza, R., de Andrade, D., Tezza, R., & de Sousa Junior, A. (2012). Relevance and advantages of using the item response theory. *Qual Quant* 47, 2341–2360.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22.
<http://dx.doi.org/10.1037/h0046743>
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cattell, R. B. (1987). *Intelligence: Its Structure, Growth, and Action*. Elsevier Science Publishers.
- Chae, Sunhee, Kang, Unson, Jeon, Eunhwa, & Linacre, J. M. (2000). Development of Computerized Middle School Achievement Test. *MESA Memorandum No. 69*. Seoul, South Korea: Komesa Press.
- Crocker, L., and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Harcourt.
- Flaugher, R. (2000). Item Pools. In Wainer, H. (Ed.) *Computerized adaptive testing: A Primer*. Mahwah, NJ: Erlbaum.
- Galton, F. (1869). *Hereditary Genius*. London: MacMillan.
- Gade, P. A., Johnson, R. M., & Pliske, R. M. (1984). Cross-Validation of the Computerized Adaptive Screening Test (CAST). *Research Report 1372*. U. S. Army Research Institute for the Behavioral and Social Sciences.
- Hally, T. (2012). *The g-factor*. https://www.researchgate.net/publication/275659645_The_g-Factor
- Kolen, M. F., & Brennan, R. L. (2014). *Test Equating, Scaling and Linking*. New York, NY: Springer.
- Magno, C. (2009). Demonstrating the difference between Classical Test Theory and Item Response Theory using derived test data. *International Journal of Educational and Psychological Assessment*, 1(1), 1-11.
- Meijer, R. R. (1999). Computerized Adaptive Testing: Overview and Introduction. *Applied Psychological Measurement*. 23 (3), 187-194.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

- Rocklin, T. R., Roos, L. L., Wise, S. L., & Yoes, M. E. (1996). Conducting Self-Adapted Testing Using MicroCAT. *Educational and Psychological Measurement*, 56 (5), 821-827.
- Ryan, J., & Brockmann, F. (2009). A Practitioner's Introduction to Equating with Primers on Classical Test Theory and Item Response Theory. Retrieved from <https://eric.ed.gov/?id=ED544690>
- Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, 86 (1), 162-173.
- Sinharay, S., Haberman, S., Holland, P., & Lewis, C. (2012). A Note on the Choice of an Anchor Test in Equating. ETS research report series.
- Spearman, C. (1927). *The Abilities of Man*. London: Macmillan.
- Thomas, M. (2011). The Value of Item Response Theory in Clinical Assessment: A Review. *Assessment*, 18(3), 291-307.
- Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*. 16 (1)
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Vinney, C. (2020). Fluid Versus Crystallized Intelligence: What's the Difference? <https://www.thoughtco.com/fluid-crystallized-intelligence-4172807>
- Vispoel, W., Rocklin, T., Wang, T., & Bleiler, T. (1999). Can Examinees Use a Review Option to Obtain Positively Biased Ability Estimates on a Computerized Adaptive Test? *Journal of Educational Measurement*, 36(2), 141-157.
- Wechsler, D. (1944). *The measurement of adult intelligence (3rd ed.)*. Baltimore: Williams & Wilkins.
- Weiss, D. J. (1983). New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing.

NOUS CONTACTER

***Service Client* : customer@centraltest.com**

***Formation et Consulting* : consulting@centraltest.com**



www.centraltest.com

© 2021 par Central Test – Empowering Talent.
Tous droits réservés.