

d2-R

Test d'Attention Concentrée – Révisé



Rolf Brickenkamp
Lothar Schmidt-Atzert
Detlev Liepmann

Supplément pour la version informatisée

d2-R Version informatisée Copyright © 2015 by Hogrefe Verlag GmbH & Co. KG. Not to be reproduced in whole or in part in any form or by any means without written permission of Hogrefe Verlag GmbH & Co. KG. All rights reserved.
www.hogrefe.de



d2-R Adaptation française de la version informatisée Copyright © 2016 by Éditions Hogrefe France. Toute reproduction interdite.
www.hogrefe.fr

Vous trouverez rapidement dans ce manuel l'information dont vous avez besoin en cliquant sur l'un des titres de la table des matières.

Table des Matières

1	Introduction	5
1.1	La mesure de l'attention concentrée avec le d2-R	5
1.2	Populations cibles et applications	5
2	Version informatisée du d2-R	8
2.1	Conditions d'administration du test	8
2.2	Recueil des résultats	9
2.3	Choix du contenu du rapport	10
3	Interprétation	11
3.1	Interprétation des notes au test	11
3.2	Interprétation du rapport	11
4	Développement de la version informatisée du d2-R et données psychométriques	13
4.1	Collecte des données	13
4.2	Standardisation continue	13
4.2.1	Critères d'exclusion	14
4.3	Caractéristiques démographiques	15
4.4	Impact du pays, du sexe et de l'âge	15
4.4.1	Pays	16
4.4.2	Sexe	18
4.4.3	Âge	18
4.5	Corrélations entre indices	21
4.6	Consistance interne	21
4.7	Stabilité	22
	Références bibliographiques	23

Liste des Tableaux et des Figures

Tableau 1:	Distribution par pays au sein de l'échantillon normatif européen	15
Tableau 2:	Distribution par sexe et par tranche d'âge au sein de l'échantillon normatif	16
Tableau 3:	Distribution par niveau d'éducation et par origine (immigrée ou non) au sein de l'échantillon normatif	16
Tableau 4:	Moyennes (<i>M</i>) et écarts-types (ET) des indices CC, CCT et E% par pays	17
Tableau 5:	Moyennes (<i>M</i>) et écarts-types (ET) des indices CC, CCT et E% en présence et en l'absence de supervision	17
Tableau 6:	Corrélation entre l'âge (en années) et les résultats pour les indices CC, CCT et E%	18
Tableau 7:	Moyennes (<i>M</i>) et écarts-types (ET) pour les indices CC, CCT et E% par âge	18
Tableau 8:	Corrélations entre CCT, CC et E% pour l'échantillon total	21
Tableau 9:	Alpha de Cronbach et corrélations item-total pour l'échantillon total	21
Tableau 10:	Alpha de Cronbach par âge (en années)	22
Tableau 11:	Fidélité test-retest de CC, CCT et E% pour deux intervalles de temps	22
Figure 1:	Extrait de l'écran présentant les instructions du d2-R	8
Figure 2:	Extrait de l'écran présentant un exercice d'entraînement	9
Figure 3:	Extrait de la page de la plateforme HTS 5 permettant de choisir les notifications souhaitées	9
Figure 4:	Extrait de la page de la plateforme HTS 5 permettant de choisir le contenu du rapport	10
Figure 5:	Extrait de la page de la plateforme HTS 5 permettant de choisir le groupe normatif de comparaison	10
Figure 6:	Moyenne pour l'indice CC par âge	20
Figure 7:	Moyenne pour l'indice CCT par âge	20
Figure 8:	Moyenne pour l'indice E% par âge	20



1 Introduction

Dans la pratique, le test d2 est un instrument très apprécié par les psychologues. Selon des enquêtes réalisées en Allemagne auprès des utilisateurs, il vient au premier rang (Steck, 1997) ou au deuxième (Schorr, 1995), en matière de fréquence d'utilisation, sur la liste des tests de performance. Le test d2 joue également un rôle de premier plan dans la recherche. Depuis sa parution, le nombre de publications dans les revues spécialisées, faisant référence au d2, n'a cessé d'augmenter. Ceci a incité les auteurs du test à poursuivre son développement. Cela signifiait surtout optimiser le matériel et la cotation, clarifier son appartenance conceptuelle (son fondement théorique), actualiser la vue d'ensemble des recherches concernant sa validité, établir de nouvelles normes, donner des instructions très concrètes pour son administration, et améliorer encore sa cotation et son interprétation. Le d2-R est donc une version révisée du renommé test d2.

Cette révision s'accompagne d'une nouveauté concernant les modalités de la passation: celle-ci peut désormais se faire sur ordinateur ou sur écran tactile. Par ailleurs, la version papier-crayon du test existe toujours pour les utilisateurs qui souhaitent l'utiliser. Le présent manuel a été spécifiquement conçu pour les utilisateurs de la version informatisée du d2-R. Un manuel plus complet, disponible aux Éditions Hogrefe France, fournit davantage d'informations au sujet des fondements théoriques, du développement et de la construction du d2-R.

1.1 La mesure de l'attention concentrée avec le d2-R

Plusieurs définitions sont possibles pour l'attention. Schmidt-Atzert, Krumm et Bühner (2008) voient dans la définition «sélection de stimuli ou d'événements externes pertinents perçus directement», une proposition susceptible de faire l'unanimité. L'important est donc tout d'abord qu'il s'agisse d'un phénomène en lien avec la perception et, ensuite, que ce phénomène soit reconnaissable par une sélection de stimuli.

Le test d2 oblige les participants à repérer certains stimuli (d avec deux traits) dispersés parmi des stimuli similaires. Il s'agit là, clairement, d'une perfor-

mance mettant en jeu l'attention. Celle-ci comprend différents aspects qui se distinguent par certaines tâches et certains traitements (attention sélective ou partagée, attention soutenue, vigilance). Le d2 peut être classé parmi les tests sollicitant l'attention sélective (appelée parfois aussi focalisée). Il ne peut cependant pas être considéré comme un test d'attention pure, car il exige en outre de la concentration. Dans la plupart des autres tests d'attention, les stimuli sont présentés à l'écran et les participants doivent décider rapidement s'il s'agit d'un stimulus pertinent (cible) ou non pertinent (distracteur) (Schmidt-Atzert et al., 2008). Il manque le passage rapide, déterminé par le participant lui-même, à la tâche suivante (ce qui est typique d'un mode de traitement *self-paced*, c'est-à-dire à son propre rythme).

Quand l'attention est couplée avec la concentration, on parle aussi d'«attention concentrée» (Schmidt-Atzert et al., 2008). L'élément déterminant n'est alors pas la durée, mais la difficulté à maintenir l'attention assez longtemps sur quelque chose. Dans ce sens, le d2 est un test d'attention concentrée. Ce qui signifie aussi qu'il mesure la concentration dans la sélection des stimuli pertinents ou dans une tâche exigeant de l'attention sélective.

1.2 Populations cibles et applications

La version informatisée du d2-R a été étalonnée dans 10 pays européens (Allemagne, Danemark, Finlande, France, Italie, Norvège, Pays-Bas, République tchèque, Royaume-Uni, Suède) auprès d'adultes âgés de 18 à 55 ans.

Rappelons ici que la cotation et l'interprétation des résultats au d2-R sont réservées aux professionnels qualifiés (psychologues, psychiatres).

Le d2-R peut être utilisé dans de très nombreux domaines. Divers manuels le recommandent comme test diagnostique. D'après le DSM-5 (APA, 2013), l'inattention est déterminée au moyen de neuf critères, parmi lesquels six au moins doivent être satisfaits pour que l'on puisse poser le diagnostic de troubles de l'attention. Le critère «semble souvent ne pas écouter quand on lui parle personnellement»

ne peut certainement pas être vérifié par un test, car il s'agit là d'un constat touchant le comportement au quotidien. En revanche, on pourrait très bien analyser le critère « a souvent du mal à soutenir son attention au travail ou dans les jeux » par un test d'attention tel que le d2-R.

Ci-dessous sont présentés les principaux domaines d'applications du d2-R.

Psychologie du travail et des organisations

Certaines activités pratiquées dans le cadre d'une profession exigent de la concentration. Quand une analyse des exigences d'un poste montre que cette capacité est nécessaire à l'exercice de la profession, il est possible d'évaluer les aptitudes d'une personne au moyen d'un test. Pour choisir le test qui convient, il faut, outre les preuves de validité, prendre en compte la similitude entre l'activité professionnelle en question et la tâche proposée par le test. Le choix devrait se porter sur le d2-R quand l'attention visuelle concentrée est particulièrement sollicitée.

Les tests sont particulièrement utiles pour sélectionner du personnel au moment d'un recrutement ou pour attribuer en interne des tâches particulières à des collaborateurs. Pour les questions de placement, qui se posent le plus souvent dans l'accompagnement des demandeurs d'emploi, la capacité de concentration peut, suivant la profession souhaitée, s'avérer déterminante.

Lorsqu'il s'agit d'organiser ou de réorganiser des postes ou des conditions de travail, la question se pose parfois de savoir si ces changements auront un effet positif ou négatif sur la concentration du personnel. Comme le test d2-R ne prend que quelques minutes, il peut le cas échéant être administré en poste de travail dans différentes conditions (anciennes et nouvelles, par exemple). Toutefois, en raison des effets d'apprentissage, il est important d'être prudent et de ne pas se contenter d'une simple comparaison « avant-après ».

Psychologie clinique et médicale

Les troubles de l'attention ou de la concentration constituent un critère diagnostique dans différentes pathologies. La capacité de concentration est parfois

diminuée dans les dépressions. L'un des critères diagnostiques d'une dépression majeure au sens du DSM-5 est formulé comme suit: « diminution de l'aptitude à penser ou à se concentrer ou indécision, presque tous les jours (signalée par la personne ou observée par les autres) » (APA, 2013). La terminologie du système diagnostique de l'OMS, la CIM-10, diffère quelque peu de celle du DSM-5. Elle fait la distinction en fonction de la gravité de l'épisode dépressif, dont l'un des critères est: « diminution de l'aptitude à penser ou à se concentrer, grande difficulté à prendre des décisions » (OMS, Dilling, Mombour, Schmidt & Schulte-Markwort, 2006). Le test d2-R peut, au besoin, objectiver une baisse de la capacité de concentration par rapport à des personnes du même âge.

Le trouble de l'attention ou de la concentration fait explicitement partie du TDA/H. Même si, en pratique, on peut souvent se contenter d'entretiens ou de questionnaires diagnostiques pour déterminer la présence des critères pertinents, le test d2-R est à même de fournir des compléments d'information importants. Pour éviter tout malentendu, il est important de rappeler qu'un « mauvais » résultat au test ne justifie pas un diagnostic de TDA/H, pas plus qu'un « bon » résultat ne prouve que la personne n'a pas de TDA/H. Le critère du DSM-5 « a souvent du mal à soutenir son attention au travail ou dans les jeux » se rapporte clairement à une faculté qui peut aussi se manifester en-dehors du jeu et des devoirs (au sens large). Un test qui exige le maintien de l'attention représente donc une excellente possibilité de rendre objectif un éventuel déficit. Mais comme il faut bien sûr prendre aussi en compte les autres critères, l'utilisation du test d2-R ne peut être qu'un élément parmi d'autres en faveur d'un diagnostic.

Le test d2-R peut tout aussi bien être employé en dehors des diagnostics de la CIM ou du DSM pour objectiver des troubles allégués par les patients (« j'ai du mal à me concentrer »). Il peut aussi être utilisé pour évaluer des mesures thérapeutiques – mais, étant donné les effets d'apprentissage connus, il ne peut pas être utilisé pour poser un diagnostic dans un cas particulier. Cette remarque ne concerne pas seulement les interventions visant à améliorer la capacité de concentration, mais aussi les éventuels effets secondaires des traitements médicaux ou pharmacologiques.

Neuropsychologie

Parmi les patients examinés d'un point de vue neuropsychologique suite à un accident vasculaire cérébral, à cause d'une tumeur ou d'un traumatisme crânien, beaucoup présentent un trouble de l'attention (Sturm & Zimmermann, 2000). Pour garantir l'efficacité d'un traitement ou d'une thérapie, il est particulièrement important de repérer et de quantifier les déficits du patient. La mémoire, la capacité de concentration, la rapidité de réaction sont-elles atteintes et, si oui, dans quelle mesure ? En neuropsychologie, le construit «attention» est abordé de manière très différenciée et de nombreux tests permettent d'en évaluer l'une ou l'autre des composantes. Toutefois, le point de vue sur les diverses fonctions postulées de l'attention est assez inflation-

niste : on admet plus de composantes que ce que l'on peut réellement en distinguer sur le plan théorique et surtout empirique. Schmidt-Atzert et al. (2008) ont passé en revue les divers modèles proposés et les résultats empiriques concernant la structure de l'attention. Ils concluent que les composantes les plus fondées, tant théoriquement qu'empiriquement, sont la promptitude (vitesse de réaction), la sélectivité, l'attention soutenue (maintien de l'attention pendant un certain temps et vigilance) et l'attention concentrée. Le test d2-R correspond clairement à la catégorie de l'attention concentrée.

Comme en psychologie clinique et médicale, le test d2-R convient aussi en neuropsychologie, pour évaluer les interventions et rechercher des effets secondaires.

2 Version informatisée du d2-R

La version informatisée du d2-R reprend largement la version papier: le participant doit rechercher et marquer un ensemble de caractères spécifiques (les lettres «d» accompagnées de deux traits). Le marquage s'effectue par un clic de souris sur ces caractères appelés «caractères cibles». Dans la version pour écran tactile, le caractère cible doit être touché du doigt. Le caractère est alors barré d'une ligne oblique qui le signale comme étant marqué. Les consignes présentent au participant les trois variantes possibles pour le caractère cible («d» accompagné de deux traits, soit au-dessus, soit en dessous, soit un trait au-dessus et l'autre en dessous). Elles présentent également les caractères à ne pas marquer, autrement dit «distracteurs» (la lettre «d» accompagnée d'un ou de trois traits, et la lettre «p» quel que soit le nombre de traits). La tâche est ensuite réalisée une première fois sans limite de temps, en guise d'exercice. Il est ensuite demandé de réaliser la tâche en temps limité. Les caractères doivent être traités ligne par ligne. Au début, le participant est averti en cas d'omission d'un caractère cible ou de marquage d'un caractère distracteur, et est invité à se corriger. Pour faire disparaître la barre oblique de marquage, il suffit de cliquer à nouveau sur le caractère, ou de le toucher du doigt. Le test se compose de 14 séries de caractères proposées les unes à la suite des autres, espacées d'une courte pause (d'environ une seconde). L'écran af-

fiche 60 caractères à la fois, ordonnés en 6 lignes de 10 caractères. Tous les caractères cibles, soit un «d» accompagné de deux traits, doivent être marqués. Dans les consignes, il est demandé de traiter le test rapidement et si possible sans commettre d'erreurs. Le rythme de traitement par série est limité à 20 secondes.

2.1 Passation du test

Le d2-R informatisé peut être passé grâce à *Hogrefe Testsystem* (HTS 5). En se connectant à la plateforme via internet, le participant peut passer le test en ligne. Un lien peut également être envoyé à la personne à qui l'on souhaite faire passer le d2-R.

Lorsque le participant commence une session, le mode plein écran est activé. Ce n'est qu'à la fin de la passation, ou si une interruption est souhaitée (avec la touche ESC), que le mode plein écran est désactivé.

La session commence par une explication de la tâche: voir **Figure 1**.

Un écran avec un exercice est ensuite proposé afin de permettre au participant de s'entraîner et de s'assurer que la tâche a bien été comprise: voir **Figure 2**.

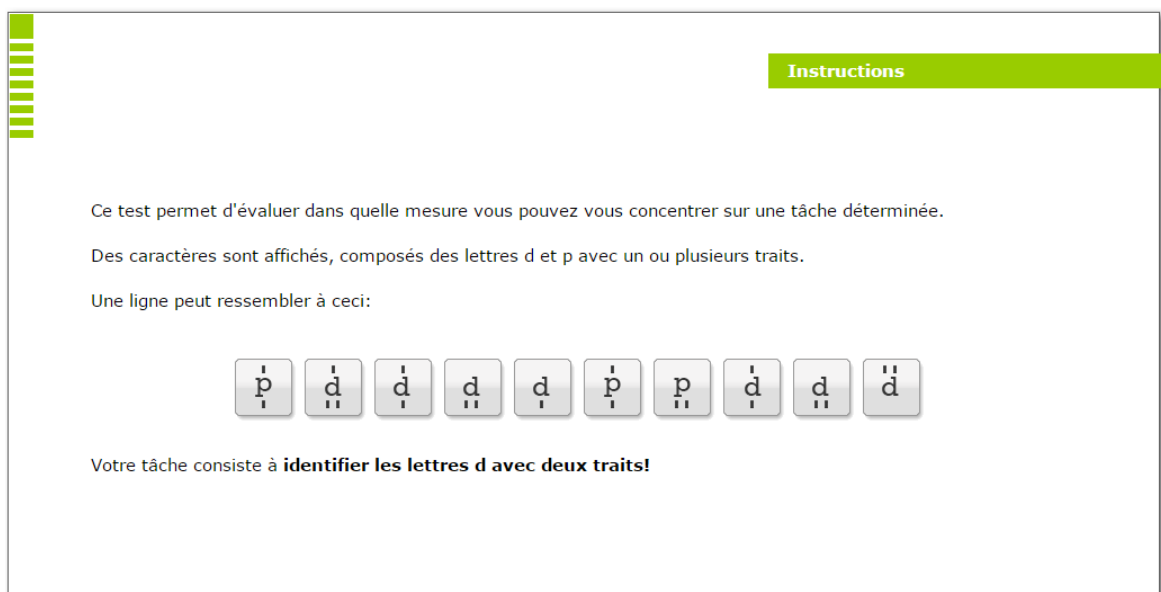


Figure 1: Extrait de l'écran présentant les instructions du d2-R

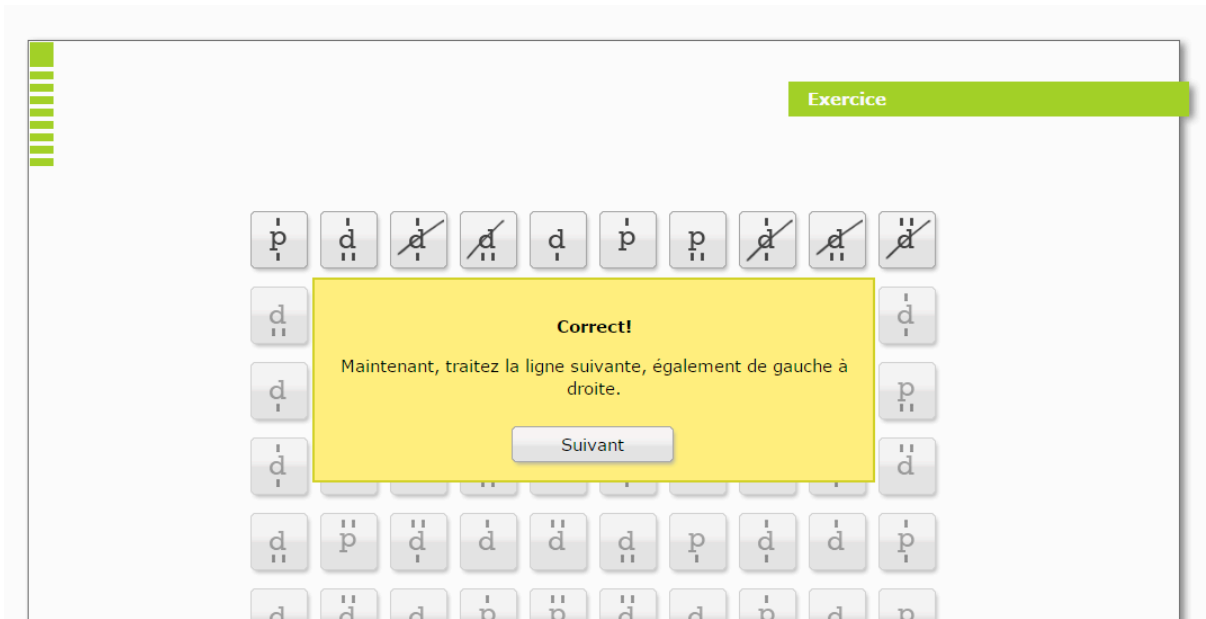


Figure 2: Extrait de l'écran présentant un exercice d'entraînement

Ensuite le test peut être démarré : 14 séries de caractères seront proposées à l'écran, chacune composée de 6 lignes de 10 caractères. Le temps de traitement de chaque série est limité à 20 secondes.

2.2 Recueil des résultats

Lorsque la session est terminée, il est possible d'accéder aux résultats sur la plateforme *Hogrefe Testsystem*. Si le participant a passé le test grâce à un

lien, l'administrateur reçoit une notification par mail lui indiquant que le test a été réalisé. Il peut alors obtenir les résultats.

Hogrefe Testsystem permet également des passations en groupe ou des passations multiples par le même participant à des moments différents. Dans ce cas, une fois toutes les passations réalisées, il est possible de comparer les résultats entre eux. Il est aussi possible de comparer les résultats avec un profil souhaité ou de créer un profil de groupe.

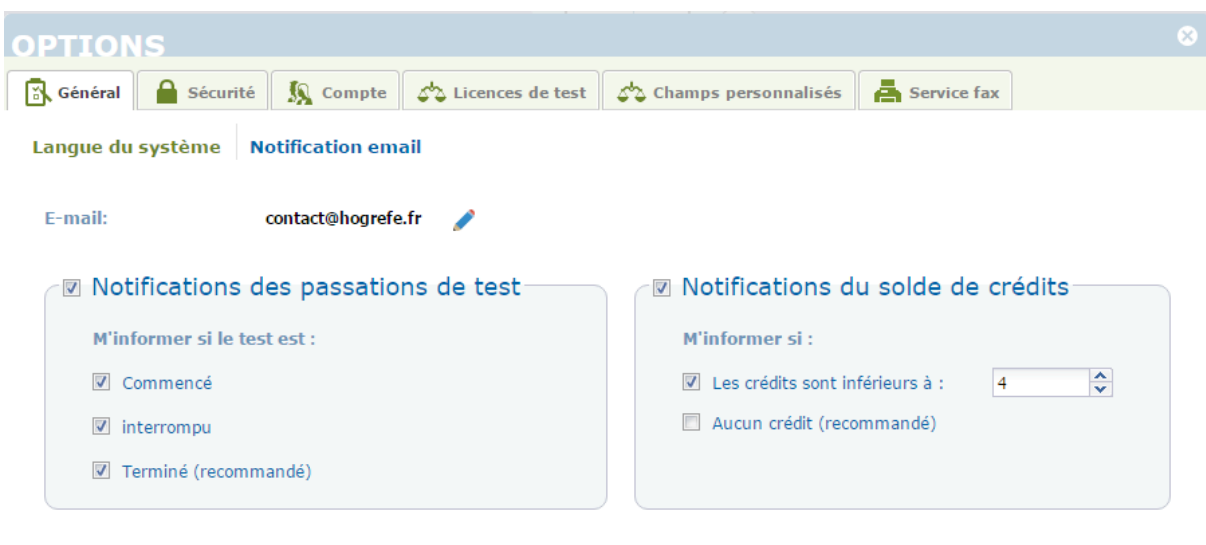


Figure 3: Extrait de la page de la plateforme HTS 5 permettant de choisir les notifications souhaitées



Figure 4: Extrait de la page de la plateforme HTS 5 permettant de choisir le contenu du rapport

2.3 Choix du contenu du rapport

Le rapport complet se compose de 4 parties : interprétation, feuille de profil, notes aux échelles, descriptions des échelles. L'administrateur du test peut choisir le contenu qu'il souhaite faire apparaître dans le rapport et un pdf sera créé en fonction de ses choix. Voir **Figure 4**.

L'administrateur du test peut également choisir les normes qu'il souhaite utiliser. Par défaut, le système

de normes proposé est celui des notes T et le groupe de référence est celui de la population européenne, des hommes et des femmes, de 18 à 55 ans.

Si on souhaite comparer les résultats du participant aux hommes et aux femmes du même groupe d'âge que lui, on sélectionnera « Population européenne, hommes et femmes, spécifique à l'âge ».

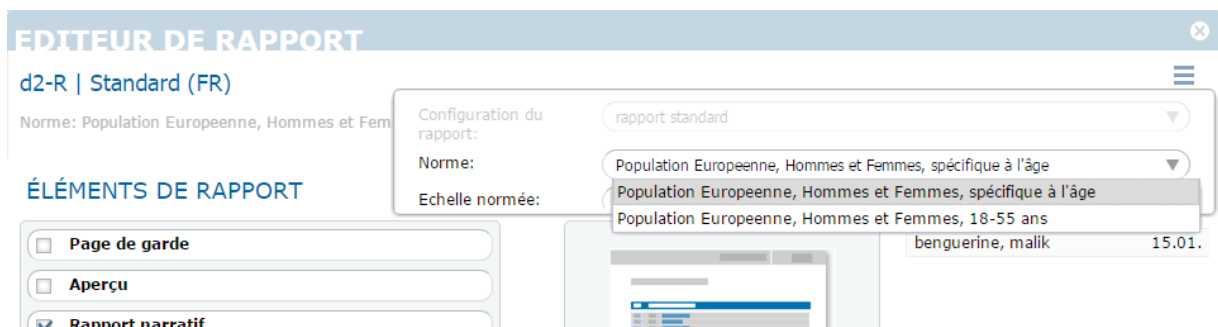


Figure 5: Extrait de la page de la plateforme HTS 5 permettant de choisir le groupe normatif de comparaison

3 Interprétation

3.1 Interprétation des notes au d2-R

Trois indices principaux sont calculés et transformés en notes standard :

- La capacité de concentration (CC) : cet indice est déterminé à partir du nombre de caractères cibles correctement barrés (CCT) moins le nombre d'erreurs de confusion et d'omission ($CC = CCT - EO - EC$). La capacité de concentration mesure également le rythme de traitement corrigé des erreurs.
- Le rythme de traitement (CCT) : le nombre de caractères cibles traités (CCT) est obtenu à partir du numéro du dernier caractère cible à avoir été traité. Il ne prend pas en compte que les bonnes réponses, mais aussi les caractères cibles ignorés jusque-là (« erreurs d'omission »). Cet indice mesure ainsi la vitesse de traitement sans prendre en compte la justesse des réponses.
- L'exactitude (E%) : le pourcentage d'erreurs (E%) est déterminé en divisant le nombre d'erreurs (d'omission et de confusion) par le nombre de caractères cibles traités (CCT). Pour obtenir un pourcentage, on multiplie ensuite le résultat par 100. Avec les valeurs normées, une note standard élevée correspond à une exactitude élevée dans le traitement du test, c'est-à-dire très peu d'erreurs.

Trois autres indices sont mesurés ; ils consistent en des notes brutes.

- Les erreurs d'omission (EO) : les caractères cibles figurant avant le dernier caractère traité (cf. CCT) et n'ayant pas été marqués sont considérés comme des erreurs d'omission. Ces erreurs surviennent fréquemment car elles sont faciles à commettre. Le pourcentage d'erreurs E% (cf. ci-dessus) dépend en général du nombre d'erreurs d'omission, car les erreurs de confusion sont rares.
- Les erreurs de confusion (EC) : lorsqu'un caractère autre qu'un caractère cible est marqué, il s'agit d'une erreur de confusion. Il est toutefois rare que le participant barre un « p », ou un « d » accompagné d'un ou de trois traits. Si le nombre d'erreurs de confusion est grand, cela peut in-

diquer que le test n'a pas été réalisé suivant les consignes. Si le participant marque les caractères au hasard, le nombre d'erreurs de confusion peut être très grand (le plus souvent largement supérieur à 100).

- Les erreurs de lettre (EL) : il s'agit d'une sous-catégorie d'erreur de confusion. Le participant peut marquer des caractères comportant la lettre « p ». Chez la plupart des participants, les caractères comportant un « p » sont reconnus plus facilement comme « faux » que ceux comportant un « d » accompagné d'un nombre de traits autre que deux. C'est pourquoi les erreurs de lettre sont caractéristiques d'un test réalisé sans respecter les consignes.

3.2 Interprétation du rapport

Le rapport propose des explications pour comprendre les représentations graphiques des résultats au d2-R ainsi que des consignes pour l'interprétation des résultats.

La plage de ± 0.5 écart-type autour de la valeur moyenne est considérée comme moyenne. Les valeurs situées de 0.5 à 1.5 écart-type sont considérées respectivement comme élevées ou faibles, et celles situées à plus de 1.5 écart-type comme très élevées ou très faibles.

L'échelle des notes standard est divisée selon les catégories suivantes :

- Une note standard (NS) inférieure à 85 est considérée comme très inférieure à la moyenne. Un résultat se trouvant dans cette plage peut donc être interprété comme très faible ; 6.7 % des personnes de l'échantillon normatif se situent dans cette catégorie.
- Une NS comprise entre 85 et 94 peut être considérée comme inférieure à la moyenne. Un résultat se trouvant dans cette plage peut donc être interprété comme faible ; 24.2 % des personnes du groupe de référence obtiennent un résultat se situant dans cette catégorie.
- Une NS comprise entre 95 et 105 est considérée

comme moyenne. Si le participant obtient un résultat se situant dans cette plage, on considérera ce résultat comme moyen. La plage des valeurs moyennes englobe 38.3 % de l'échantillon normatif.

- Une NS comprise entre 106 et 115 peut être considérée comme moyenne ou supérieure à la moyenne. Un résultat situé dans cette plage peut être vu comme élevé; 24.2 % des personnes de l'échantillon normatif obtiennent un résultat se situant dans cette plage.
- Une NS supérieure à 115 peut être considérée comme très supérieure à la moyenne. Le résultat correspondant peut être interprété comme très élevé; 6.7 % des personnes de l'échantillon normatif atteignent un tel résultat.

Feuille de profil

La feuille de profil fournit une vue d'ensemble des trois résultats CC, CCT et E%.

Notes aux échelles

Un premier tableau récapitulatif affiche les notes brutes obtenues pour CC, CCT, E% ainsi que le nombre d'erreurs d'omission (EO), d'erreurs de confusion (EC) et d'erreurs de lettre (EL). Les tableaux suivants détaillent les notes brutes obtenues pour chaque série de stimuli.

Description des échelles

Cette partie du rapport propose pour chacun des trois indices (CC, CCT et E%) une description de

l'indice et une représentation graphique des résultats sur des courbes de distribution normale.

Les résultats du test peuvent parfois indiquer un non-respect des consignes. Dans ce cas, un avertissement apparaît. Deux cas de figure sont possibles :

- Caractères marqués au hasard

Dans ce cas, la valeur CC est en général négative et le nombre d'erreurs de confusion est extrêmement élevé (pour les tests réalisés rapidement, $EC > 100$). La proportion de distracteurs est significativement supérieure à celle de caractères cibles. Un marquage au hasard produit par conséquent plus d'erreurs de confusion (EC) que de bonnes réponses. Le calcul de la valeur CC ($CC = CCT - EO - EC$) est ainsi négatif. Un marquage au hasard peut résulter d'une mauvaise compréhension ou d'un oubli des consignes, ou encore d'un « boycott » intentionnel de la part du participant.

- Simulation d'une capacité faible

Les participants cherchant à simuler une mauvaise performance se font la plupart du temps repérer par deux ou plus (mais rarement plus de 20) erreurs de lettre. Le nombre total d'erreurs de confusion est le plus souvent élevé car parfois, le participant marque aussi intentionnellement la lettre « d » accompagnée d'un mauvais nombre de traits. Si en plus, la valeur CC est négative, le résultat sera le fruit d'un marquage au hasard plutôt que d'une simulation.

4 Développement de la version informatisée du d2-R et données psychométriques

Ce chapitre présente les données psychométriques de la version informatisée du test d2-R. Il commence par une brève explication de la façon dont les données ont été recueillies et de la procédure normative, également appelée «standardisation continue». Sont également expliqués les critères d'exclusion applicables aux participants faisant partie de l'échantillon normatif. Ensuite sont présentées les caractéristiques démographiques de l'échantillon normatif, avant d'aborder l'impact du pays, du sexe et de l'âge sur les résultats au test d2-R, ainsi que les corrélations entre les différents indices du test d2-R. Pour finir, la fidélité du test d2-R est analysée à la lumière de la consistance interne des indices et de la stabilité dans le temps.

4.1 Collecte des données

Les données utilisées pour la version informatisée du test d2-R ont été collectées dans 10 pays d'Europe: la République tchèque, le Danemark, la Finlande, la France, l'Allemagne, l'Italie, les Pays-Bas, la Norvège, la Suède et le Royaume-Uni.

En Finlande, en Norvège et en Suède, la collecte des données a été supervisée, c'est-à-dire qu'un superviseur était présent pendant que les participants réalisaient le test. Les participants avaient été sélectionnés afin que l'échantillon soit représentatif du pays considéré en matière de sexe, d'âge et de niveau d'études. En outre, les participants issus de l'immigration ont été invités à signaler leurs origines.

Dans les autres pays, la collecte des données n'a fait l'objet d'aucune supervision (se référer à la partie 4.4.1 pour une comparaison entre les données dont la collecte a été supervisée et celles collectées sans supervision).

En France, en Allemagne, en Italie, aux Pays-Bas et au Royaume-Uni, les données ont été collectées par le biais d'une étude par panel. Dans chaque pays, l'entreprise de sondages a élaboré un échantillon afin qu'il soit représentatif de la population en matière de sexe, d'âge et de niveau d'études. Les participants issus de l'immigration ont été invités à préciser leurs origines.

Au Danemark et en République tchèque, les participants ont été contactés par le biais d'un registre public ou en collaboration avec des universités. Les participants ont été sélectionnés afin d'obtenir un échantillon représentatif de la population en matière de sexe, d'âge et de niveau d'études. De même, il a été demandé aux participants issus de l'immigration d'indiquer leurs origines.

4.2 Standardisation continue

La version informatisée du d2-R propose deux sortes de normes. La première consiste en un échantillon normatif classique, tous âges confondus. La deuxième consiste en des normes spécifiques à l'âge. Ces dernières ont été élaborées à l'aide d'une procédure de standardisation continue. Cela signifie que pour chaque âge, des normes spécifiques ont été calculées. Dans la procédure de standardisation classique, les données sont regroupées pour former des catégories par tranches d'âges. Par exemple, les 18–25 ans, les 26–35 ans, les 36–45 ans, les 46–55 ans. Les normes continues ont l'avantage de permettre d'éviter les grands écarts de notes standard entre tranches d'âge adjacentes. Par exemple, dans la version néerlandaise de la troisième édition de l'échelle d'intelligence de Wechsler destinée aux enfants, on observe parfois des différences entre tranches d'âge de 10 points de QI (Oosterveld, 2012). En modélisant les données normatives et en insérant l'âge comme variable continue, on évite ce type d'écart. La standardisation continue offre un second avantage: la détermination d'une note standard pour chaque association d'une note brute à un âge tient compte des données de tout l'échantillon. Par conséquent, l'utilisation de normes continues permet de réduire la taille de la norme suffisante (sous-groupes de 50 à 75 participants) par rapport aux normes classiques (Zhu & Chen, 2011).

Les normes continues sont particulièrement utiles lorsque les notes brutes varient en fonction de l'âge. Ainsi, la standardisation continue est principalement employée dans les tests cognitifs (tests d'intelligence, tests de connaissances, tests neuropsychologiques, etc.). La standardisation continue s'appuie

sur une analyse de régression, où les notes obtenues au test sont modélisées en fonction de l'âge. Il s'agit également du modèle utilisé pour le test SON-R d'intelligence non verbale (Tellegen & Laros, 2011). Les notes brutes obtenues au test ont été converties en notes standard normalement distribuées, sur le principe de la distribution des fréquences cumulées. La distribution normale utilise ici une note minimale de -2.5 et une note maximale de 2.5. Par conséquent, les notes ont été dérivées de la note brute obtenue au test et de l'âge à l'aide d'une analyse de régression non linéaire. Les prédicteurs utilisés pour cette analyse de régression étaient: la note obtenue au test (t), cette même note au carré (t^2), cette note à la puissance trois (t^3), l'âge en nombre d'années (a), l'âge au carré (a^2), l'âge à la puissance trois (a^3) et les interactions entre ces prédicteurs. L'objectif de ce calcul est d'élaborer un modèle de régression dans lequel la note standard prédite est normalement distribuée pour chaque âge. Pour finir, les valeurs z ont été converties en notes T . L'utilisation d'une note standard minimale de -2.5 et d'une note standard maximale de 2.5 dans la distribution de l'analyse de régression a conduit à une note T minimale de 25 et à une note T maximale de 75.

Remarque sur la sélection de la norme: les normes spécifiques à l'âge ne correspondent pas toujours au but diagnostique recherché. Dans les procédures de sélection par exemple, il n'est pas toujours souhaitable que les normes soient corrigées de l'âge, car la performance absolue est susceptible d'être plus importante que la performance relative. Il peut alors être judicieux d'utiliser un échantillon normatif classique, tous âges confondus. Pour un diagnostic clinique, l'emploi de normes continues spécifiques à l'âge peut revêtir une grande importance.

4.2.1 Critères d'exclusion

Pour l'échantillon normatif européen, 2 175 participants issus de 10 pays d'Europe (voir partie 4.3 Caractéristiques démographiques) ont répondu à la version informatisée du test d2-R. Les données de ces participants ont été soigneusement vérifiées afin d'éliminer les tests dont les résultats étaient irréalistes ou improbables. Les critères d'exclusion utilisés étaient les suivants :

- Lorsqu'un participant n'a pas sélectionné au moins un caractère cible sur chaque écran (c'est-

à-dire pour chaque série de 60 caractères), son protocole est éliminé. Il est en effet très improbable qu'un participant répondant au d2-R ne soit pas capable de sélectionner au moins un caractère cible sur chaque écran. Il est davantage probable que ce participant n'ait pas terminé le test, ou n'ait pas traité tous les écrans du test.

- Lorsqu'un participant a commis 10 «erreurs de lettre» (EL) ou plus, son protocole a été éliminé. Les erreurs de lettre (marquer un «p» accompagné d'1, 3 ou 4 traits, ou un «p» accompagné de 2 traits, c'est-à-dire une double erreur) indique une possible falsification. Lorsque leur nombre atteint 10 ou plus, la probabilité de falsification est très forte.
- Lorsqu'un participant a obtenu une note très élevée (les 2 % supérieurs de l'échantillon) concernant les caractères cibles traités (CCT), mais aussi une note très élevée (les 2 % supérieurs de l'échantillon) concernant les erreurs d'omission (EO), le protocole a été éliminé. Normalement, un nombre de CCT élevé ne doit pas s'accompagner d'un nombre d'EO élevé. Ce cas indique que de nombreux caractères ont été ignorés. Le participant a peut-être essayé de réaliser le test très rapidement (en donnant trop d'importance à la vitesse de traitement).
- Lorsqu'un participant a obtenu une note très élevée (les 2 % supérieurs de l'échantillon) concernant les caractères cibles traités (CCT), mais aussi une note très élevée (les 2 % supérieurs de l'échantillon) à propos des erreurs de confusion (EC), le protocole a été éliminé. Normalement, un nombre de CCT élevé ne doit pas s'accompagner d'un nombre d'EC élevé. Ce cas indique que de nombreux caractères ont été sélectionnés au hasard.
- Lorsqu'un participant a obtenu des notes très élevées (les 2 % supérieurs de l'échantillon) en matière d'erreurs d'omission (EO), mais aussi en matière d'erreurs de confusion (EC), le protocole a été éliminé. Ce cas indique que de nombreux caractères ont été soit ignorés, soit sélectionnés au hasard.

Lorsqu'un protocole a rempli un ou plusieurs de ces critères, il a été considéré comme non valide. Cette situation s'est appliquée à 75 des participants de

Tableau 1 : Distribution par pays au sein de l'échantillon normatif européen

Pays	Échantillon normatif	
	N	Pourcentage
République tchèque	192	9 %
Danemark	126	6 %
Finlande	192	9 %
France	223	11 %
Allemagne	248	12 %
Italie	185	9 %
Pays-Bas	263	12 %
Norvège	194	9 %
Suède	277	13 %
Royaume-Uni	200	10 %
Total	2100	100 %

l'échantillon, dont les protocoles ont par conséquent été éliminés. L'échantillon normatif final était donc constitué de 2 100 participants.

4.3 Caractéristiques démographiques

Le test d2-R a été réalisé par des adultes âgés de 18 à 55 ans, et issus de 10 pays d'Europe différents : la République tchèque, le Danemark, la Finlande, la France, l'Allemagne, l'Italie, les Pays-Bas, la Norvège, la Suède et le Royaume-Uni. Le **Tableau 1** montre, pour chaque pays, le nombre de participants ainsi que le pourcentage relatif compte tenu de la taille de l'échantillon normatif total. À l'exception du Danemark, chaque pays représente 9 % ou plus de l'échantillon normatif européen. Le **Tableau 2** montre la distribution selon le sexe et l'âge (nombre absolu et pourcentage relatif). Pour mieux appréhender la distribution par âge, les tranches d'âge suivantes ont été utilisées : 18–25 ans, 26–35 ans, 36–45 ans et 46–55 ans. Le **Tableau 2** met en avant la distribution homogène des hommes et des femmes pour chaque tranche d'âge. La distribution entre tranches d'âge s'avère elle aussi homogène. À partir de ces informations, il est possible de conclure que la distribution par pays, par sexe et par âge au sein de l'échantillon normatif européen est adéquate.

Le **Tableau 3** montre la distribution par niveau d'éducation et par origine (immigrée ou non) au sein de l'échantillon normatif. Le niveau d'éducation est déterminé sur la base de la Classification Internationale Type de l'Éducation (CITE, 2011¹). Cette dernière distingue trois niveaux :

- Niveau d'éducation faible : CITE 0, 1 et 2 (éducation de la petite enfance, enseignement primaire et premier cycle de l'enseignement secondaire)
- Niveau d'éducation moyen : CITE 3 et 4 (deuxième cycle de l'enseignement secondaire et enseignement post-secondaire non supérieur)
- Niveau d'éducation élevé : CITE 5, 6 et 7 (enseignement supérieur de cycle court, niveau licence ou équivalent, niveau master ou équivalent)

L'origine immigrée est indiquée par le lieu de naissance : un participant né dans le pays considéré n'est pas issu de l'immigration, tandis qu'un participant né hors du pays est issu de l'immigration. Le **Tableau 3** montre également la distribution par niveau d'éducation et par origine immigrée ou non, en Europe. On distingue ici « Pays européens de l'échantillon d'Hogrefe », pour lesquels les pourcentages sont calculés à partir des moyennes dans les dix pays participants où les données ont été collectées, et l'Union européenne, pour laquelle les pourcentages concernent l'ensemble des 28 États. Pour les « Pays européens de l'échantillon d'Hogrefe » comme pour l'Union européenne, les pourcentages concernant le niveau d'éducation et l'origine immigrée ne diffèrent pas beaucoup des pourcentages calculés pour l'échantillon normatif. Par conséquent, il est possible de conclure que la distribution par niveau d'éducation et par origine immigrée au sein de l'échantillon normatif européen est adéquate.

4.4 Impact du pays, du sexe et de l'âge

La question d'éventuelles différences de notes entre pays, sexes et âges a été traitée par le biais d'analyses multivariées de la variance.

1 Commission européenne (2014). Structures des systèmes éducatifs européens 2014/2015 : diagrammes (novembre 2014). Disponible à l'adresse : http://eacea.ec.europa.eu/education/eurydice/documents/facts_and_figures/education_structures_FR.pdf.

Tableau 2 : Distribution par sexe et par tranche d'âge au sein de l'échantillon normatif

Tranche d'âge	Hommes		Femmes		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
18-25 ans	265	13 %	296	14 %	561	27 %
26-35 ans	326	15 %	267	13 %	593	28 %
36-45 ans	229	11 %	250	12 %	479	23 %
46-55 ans	233	11 %	234	11 %	467	22 %
Total	1053	50 %	1047	50 %	2100	100 %

Tableau 3 : Distribution par niveau d'éducation et par origine (immigrée ou non) au sein de l'échantillon normatif

Niveau d'éducation	Échantillon normatif		Population européenne*	
	N	Pourcentage	Pays européens de l'échantillon Hogrefe (10 pays)**	Union européenne (28 États)
Faible	466	22 %	22 %	25 %
Moyen	939	45 %	49 %	48 %
Élevé	682	32 %	29 %	26 %
Inconnu	13	1 %	-	-
Origine immigrée				
Non	1 911	91 %	90 %	93 %
Oui	187	9 %	10 %	7 %
Inconnu	2	< 1 %	-	-

* Source : Eurostat (bureau des statistiques de l'Union européenne ; tableaux « edat_lfs_9903 » et « tps00178 »).

** Les pourcentages pour « Pays européens de l'échantillon Hogrefe » sont calculés à partir de la moyenne des 10 pays où ont été collectées les données.

4.4.1 Pays

Le **Tableau 4** montre les moyennes et les écarts-types pour chacun des trois indices CC, CCT et E% pour tous les pays, ainsi que les informations statistiques de l'analyse multivariée de la variance. Le sexe, l'âge, le niveau d'éducation et l'origine immigrée ont été pris en compte comme covariables. Bien que pour les trois indices, des différences significatives apparaissent, l'ampleur de l'impact reste faible à chaque fois ($\eta^2 \leq .02$), selon les critères de Cohen (1988). Le fait qu'il existe de petits écarts entre pays n'est que peu significatif, ce qui permet une standardisation européenne. En fait, dans les

échantillons normatifs de pays considérés séparément, il est relativement courant d'observer des différences régionales. Les petits écarts au sein de l'Europe peuvent être considérés comme une fluctuation normale au sein d'une grande région géographique.

Différences dans la méthode de collecte des données

Les pays peuvent également être catégorisés d'une autre manière. Dans trois des pays, la collecte des données a été supervisée : il s'agit de la Finlande, de la Norvège et de la Suède. Dans ces trois pays, les

Tableau 4: Moyennes (*M*) et écarts-types (ET) des indices CC, CCT et E% par pays

Pays	CC		CCT		E%	
	<i>M</i>	ET	<i>M</i>	ET	<i>M</i>	ET
République tchèque	182.7	40.2	196.8	39.1	7.4	6.8
Danemark	183.5	36.0	201.3	36.4	8.8	7.7
Finlande	183.4	36.7	198.8	36.4	7.6	7.9
France	194.4	38.5	213.3	39.5	8.7	8.5
Allemagne	185.2	36.4	204.5	38.3	9.2	8.0
Italie	183.1	40.2	206.1	40.7	10.7	11.0
Pays-Bas	191.1	41.1	210.6	40.6	9.1	10.1
Norvège	164.4	36.4	186.2	37.2	11.6	9.5
Suède	187.8	33.9	207.6	37.3	9.3	7.3
Royaume-Uni	195.8	40.4	219.3	42.4	10.3	9.8
<i>p</i> (sig.)	< .01		< .01		< .01	
<i>F</i>	4.18		3.96		2.60	
<i>df</i>	9, 1758		9, 1758		9, 1758	
η^2 partiel	.02		.02		.01	

Tableau 5: Moyennes (*M*) et écarts-types (ET) des indices CC, CCT et E% en présence et en l'absence de supervision

	CC		CCT		E%	
	<i>M</i>	ET	<i>M</i>	ET	<i>M</i>	ET
Pays avec supervision	179.7	36.8	198.8	38.0	9.5	8.3
Pays sans supervision	188.4	39.4	207.9	40.3	9.2	9.0
<i>p</i> (sig.)	< .01		< .01		n.s.	
<i>F</i>	23.38		24.31		0.54	
<i>df</i>	1, 2098		1, 2098		1, 2098	
η^2 partiel	.01		.01		-	

participants ont fait l'objet d'un encadrement lors de la réalisation du test d2-R informatisé. Dans les sept autres pays, la collecte des données n'a pas été supervisée. La mesure dans laquelle cette supervision a influencé les notes obtenues au d2-R a été examinée. Le **Tableau 5** montre les moyennes et les écarts-types des indices CC, CCT et E% pour les pays avec et sans supervision, ainsi que les données

statistiques. Les pays dans lesquels les participants ont été supervisés alors qu'ils répondaient au test d2-R informatisé (c'est-à-dire la Finlande, la Norvège et la Suède) ont obtenu des notes légèrement inférieures pour les indices CC et CCT par rapport aux pays où les participants n'ont pas été supervisés. À partir de ces informations, il est possible de conclure que cette supervision n'a pas incité les

participants à s'appliquer davantage. Au contraire, les participants supervisés ont obtenu en moyenne des notes légèrement inférieures. L'ampleur de l'effet est cependant très faible ($\eta^2 = .01$) selon les critères de Cohen (1988).

4.4.2 Sexe

Aucune différence significative n'a été observée entre les sexes. La performance au test d2-R informatisé est la même pour les hommes et pour les femmes. Aucun effet d'interaction entre le sexe et le pays ou l'âge n'a été observé non plus.

4.4.3 Âge

Une relation négative a été observée entre l'âge et la performance à la version informatisée du d2-R (voir le **Tableau 6**). Pour les indices CC et CCT, il existe une corrélation négative avec l'âge en années, tandis qu'il existe une corrélation positive entre E% et l'âge (une note brute élevée à l'indice E% indique de nombreuses erreurs et donc une mauvaise performance). Ces résultats indiquent que la performance de participants adultes décroît avec l'âge, même si les coefficients de corrélation ne sont pas très élevés. Le **Tableau 7** montre les moyennes et les écarts-types pour les trois indices CC, CCT et E% pour tous les âges, ainsi que les informations statis-

Tableau 6 : Corrélation entre l'âge (en années) et les résultats pour les indices CC, CCT et E%

	r (corrélation)
CC	-.17**
CCT	-.15**
E%	.05*

** La corrélation est significative au niveau 0.01 (bilatéral).

* La corrélation est significative au niveau 0.05 (bilatéral).

tiques de l'analyse multivariée de la variance. Des écarts significatifs ont été observés entre les âges pour les indices CC et CCT, tandis que la valeur η^2 partiel indique un impact faible à moyen selon les critères de Cohen (1988). L'évolution des notes moyennes pour les indices CC, CCT et E% en fonction de l'âge est représentée par les **Figures 6, 7 et 8**. On peut observer que la performance au test d2-R informatisé décroît à mesure que l'âge augmente. Par conséquent, il peut être judicieux d'utiliser des normes spécifiques à l'âge lorsqu'il s'agit de poser un diagnostic spécifique.

Tableau 7 : Moyennes (M) et écarts-types (ET) pour les indices CC, CCT et E% par âge

Âge	CC		CCT		E%	
	M	ET	M	ET	M	ET
18	187.3	38.2	211.7	38.1	11.4	8.9
19	198.0	37.8	217.4	38.5	8.6	8.9
20	189.2	34.2	206.4	35.9	8.2	6.5
21	186.6	41.4	207.3	40.3	9.9	10.2
22	197.5	42.0	215.9	39.8	8.6	8.7
23	190.6	39.6	209.6	41.8	8.9	7.2
24	187.8	31.7	202.8	32.7	7.3	5.6
25	197.3	43.2	220.1	44.2	10.0	10.4
26	194.9	45.3	213.7	40.9	9.0	10.0
27	192.0	39.5	210.2	41.9	8.5	6.4
28	195.0	38.2	211.4	40.8	7.4	7.7
29	190.8	37.6	207.7	39.7	7.9	9.0
30	181.8	38.8	201.8	37.5	9.7	10.6

Tableau 7 : Moyennes (M) et écarts-types (ET) pour les indices CC, CCT et E% par âge (suite)

Âge	CC		CCT		E%	
	M	ET	M	ET	M	ET
31	195.5	36.7	210.1	38.3	6.8	6.2
32	179.2	39.4	200.0	40.1	10.2	10.2
33	188.6	42.0	209.1	53.3	8.8	8.3
34	185.9	38.8	205.4	41.8	9.2	8.1
35	189.2	41.8	211.8	43.5	10.3	10.3
36	190.1	32.7	208.0	32.4	8.6	7.5
37	190.7	37.2	211.9	35.8	9.7	11.1
38	181.7	38.4	204.9	32.7	11.6	11.4
39	182.1	39.5	200.3	38.6	9.2	7.5
40	187.7	43.8	206.9	46.3	9.2	6.8
41	179.8	36.2	198.9	37.2	9.5	8.0
42	186.1	32.7	205.4	36.9	9.2	7.1
43	186.2	33.7	205.1	31.8	9.1	9.0
44	174.9	32.7	193.8	32.0	9.4	10.2
45	175.5	39.0	196.8	41.3	10.4	10.0
46	174.4	41.0	203.2	40.5	13.0	15.6
47	180.3	30.7	201.0	34.9	9.9	7.3
48	178.2	34.6	194.3	34.4	8.4	6.0
49	182.1	43.3	197.1	36.6	8.4	9.5
50	165.4	32.9	183.4	36.2	9.3	8.6
51	190.0	37.0	207.3	41.6	8.0	7.0
52	172.5	35.1	191.8	34.5	9.8	9.7
53	169.6	37.3	195.1	41.1	12.5	10.1
54	179.0	45.1	200.7	49.7	10.5	8.9
55	166.8	35.5	186.8	37.5	10.6	8.0
Total	185.7	38.8	205.0	39.8	9.3	8.8
p (sig.)	< .01		< .01		n.s.	
F	2.79		2.50		1.12	
df	37. 2062		37. 2062		37. 2062	
η² partiel	.05		.04		-	

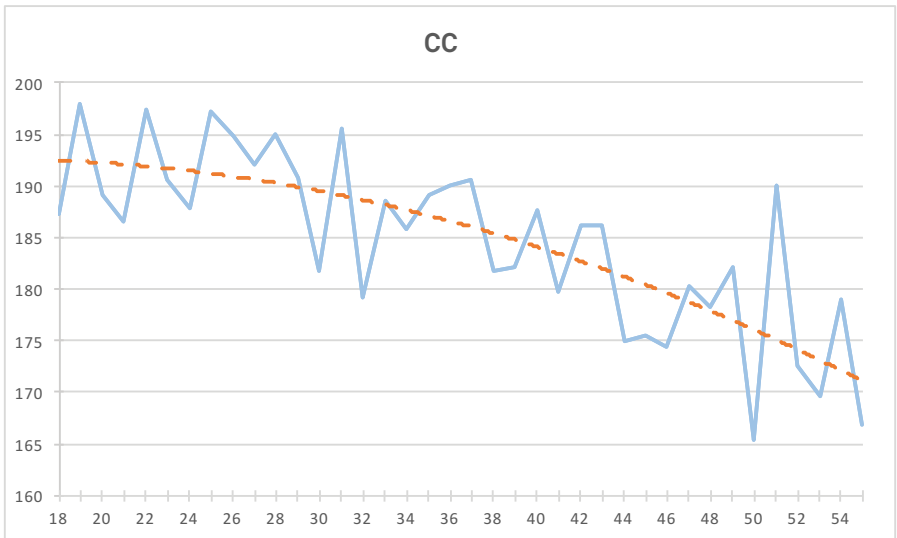


Figure 6: Moyenne pour l'indice CC par âge

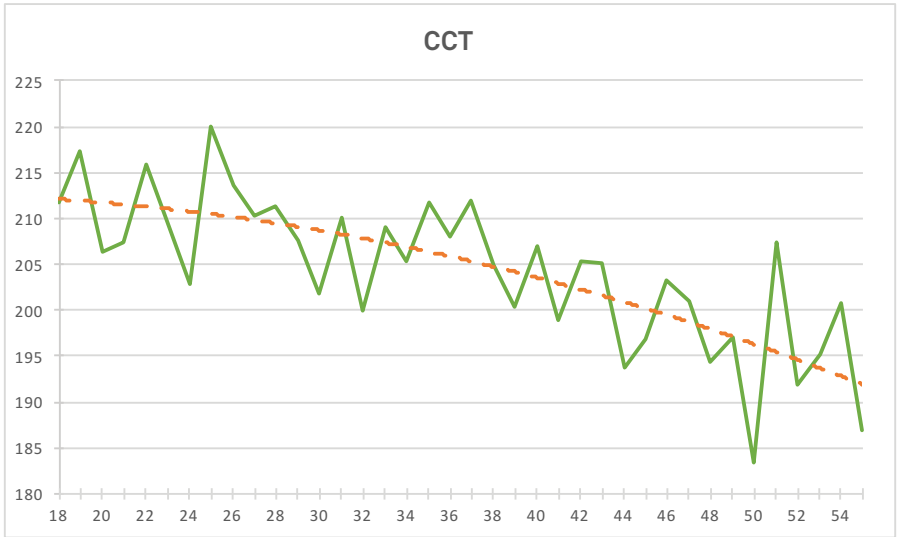
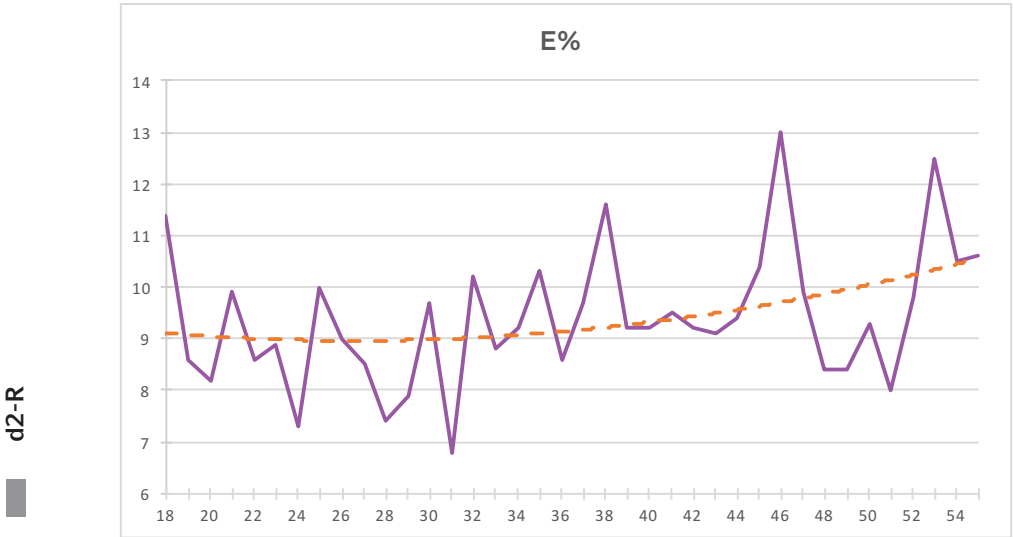


Figure 7: Moyenne pour l'indice CCT par âge



d2-R

Figure 8: Moyenne pour l'indice E% par âge

Tableau 8 : Corrélations entre CCT, CC et E% pour l'échantillon total

	CC	CCT	E%
CC	1		
CCT	.86*	1	
E%	-.39*	.12*	1

* $p < .01$

Tableau 9 : Alpha de Cronbach et corrélations item-total pour l'échantillon total

	α de Cronbach	Corrélations item-total (plage)
CC	.97	.82–.88
CCT	.97	.78–.85
E%	.92	.57–.71

4.5 Corrélations entre indices

Le d2-R informatisé se compose de trois indices qui visent à mesurer la capacité de concentration (CC), la vitesse de traitement (CCT) et l'exactitude du traitement (E%). Vitesse et précision sont deux concepts clairement différents ; la corrélation entre les indices correspondants doit être faible. Sur le plan arithmétique, l'indice CC dépend de l'indice CCT et du nombre d'erreurs. Ainsi, on peut s'attendre à des relations claires. Le **Tableau 8** montre la corrélation entre les trois indices pour l'échantillon normatif total. Comme prévu, on n'observe aucune corrélation forte entre la vitesse (CCT) et la précision (E%). Par ailleurs, il est possible de conclure qu'il existe une relation positive très forte entre CC et CCT. La corrélation entre CC et E% est également importante, bien que clairement plus faible. Le signe négatif confirme les attentes, car une note brute élevée pour l'indice E% indique une mauvaise performance

(un grand nombre d'erreurs). Globalement, les corrélations entre CC et CCT ou E% montrent qu'une forte capacité de concentration (note élevée pour l'indice CC) va en général de pair avec un traitement rapide et précis du test (la vitesse de traitement jouant un rôle plus important que la précision de traitement).

4.6 Consistance interne

Lorsqu'ils répondent au test d2-R, les participants réalisent la tâche sur 14 écrans successifs. Pour le calcul des notes brutes, seuls les écrans 2 à 14 sont pris en compte. Pour l'estimation de la cohérence, les notes de chaque écran sont considérées comme un item. Ainsi, les indices CC, CCT et E% sont constitués de 13 items. On s'attendait à ce que pour ces trois indices, la fiabilité estimée selon l'alpha de Cronbach soit d'un ordre élevé. Dans la mesure où les écrans sont absolument identiques en matière de types et de présentation des caractères et du temps de traitement nécessaire, on peut conclure que la vitesse de traitement (CCT) et l'exactitude (E%) des participants pendant le test restent relativement constantes. On peut affirmer la même chose pour la capacité de concentration (CC), qui dérive de la vitesse et de l'exactitude. Le **Tableau 9** montre les alphas de Cronbach pour l'échantillon normatif total, tandis que le **Tableau 10** présente les alphas de Cronbach pour chaque âge. Le **Tableau 9** montre également la plage de corrélations item-total pour CC, CCT et E% indiquant l'importance du rapport entre un écran (c'est-à-dire un item) et les autres écrans (c'est-à-dire le total). Pour estimer que la cohérence interne est bonne, les corrélations item-total doivent être d'au moins .30 (Evers et al., 2010). Il est possible de conclure que tous les indices présentent une cohérence interne très forte si l'on examine les 13 écrans ; globalement, les estimations pour E% sont légèrement inférieures par rapport à CC et à CCT, mais peuvent malgré tout être qualifiées d'élevées.

Tableau 10: Alpha de Cronbach par âge (en années)

Âge	α de Cronbach		
	CC	CCT	E%
18	.96	.95	.89
19	.97	.97	.94
20	.97	.97	.85
21	.97	.97	.93
22	.98	.96	.91
23	.97	.97	.88
24	.97	.96	.85
25	.97	.97	.95
26	.98	.97	.95
27	.97	.98	.84
28	.98	.98	.92
29	.97	.96	.94
30	.97	.97	.95
31	.97	.97	.90
32	.98	.97	.93
33	.98	.98	.92
34	.97	.97	.88
35	.98	.97	.95
36	.96	.95	.90
37	.97	.96	.95
38	.97	.94	.94
39	.97	.97	.88
40	.98	.98	.87
41	.97	.97	.89
42	.96	.96	.89
43	.97	.96	.92
44	.97	.95	.94
45	.97	.96	.92
46	.97	.95	.97
47	.97	.97	.89
48	.97	.97	.83
49	.98	.97	.94
50	.97	.97	.91
51	.97	.98	.86
52	.97	.96	.92
53	.97	.97	.92
54	.98	.98	.94
55	.98	.97	.92

4.7 Stabilité

Pour le test d2-R informatisé, la stabilité des notes sur la durée a été vérifiée par une répétition du test après une heure et après une semaine. Tous les participants étaient des étudiants néerlandais en Psychologie, qui réalisaient le test d2-R dans le cadre de leur module obligatoire au sujet des instruments psychométriques. Ils ont été séparés en deux groupes: l'un était composé de 113 étudiants (26 hommes, 87 femmes) avec une moyenne d'âge de 20.4 ans (ET = 1.9 ans). Ce groupe a réalisé le test d2-R au début d'un cours, et une deuxième fois à la fin du même cours, après un laps de temps d'une heure. Le deuxième groupe était composé de 81 étudiants (20 hommes, 61 femmes) avec une moyenne d'âge de 20.3 ans (ET = 1.8 ans). Ce groupe a réalisé le test d2-R au début d'un cours, et une deuxième fois au début du cours suivant, après un laps de temps d'une semaine. Les résultats de fidélité test-retest sont présentés dans le **Tableau 11**. Les indices mesurant la vitesse de traitement (CCT) et la capacité de concentration (CC) montrent une fidélité test-retest globale élevée. La stabilité du nombre d'erreurs (E%) est quelque peu inférieure à celle des indices CC et CCT. Les coefficients de retest pour CCT et CC s'avèrent légèrement inférieurs après une semaine par rapport au retest après une heure, mais peuvent tout de même être considérés comme adéquats.

Tableau 11: Fidélité test-retest de CC, CCT et E% pour deux intervalles de temps

	Intervalle de temps 1 heure	Intervalle de temps 1 semaine
	r	r
CC	.89	.84
CCT	.90	.78
E%	.68	.71

Références bibliographiques

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*, (DSM-5®). Washington, DC: American Psychiatric Publishing.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dilling, H., Mombour, W., Schmidt, M.H. & Schulte-Markwort, E. (2006). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Diagnostische Kriterien für Forschung und Praxis (4., überarb. Aufl.)*. Bern: Huber.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *Het COTAN-beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Oosterveld, P. (2012). *Continuous norming models: an empirical evaluation*. Article présenté à la 8^{ème} conférence du Comité international des tests.
- Schmidt-Atzert, L., Krumm, S. & Bühner, M. (2008). Aufmerksamkeitsdiagnostik: Ableitung eines Strukturmodells und systematische Einordnung von Tests. *Zeitschrift für Neuropsychologie*, 19 (2), 59–82.
- Schorr, A. (1995). Stand und Perspektiven diagnostischer Verfahren in der Praxis. Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen. *Diagnostica*, 41 (1), 3–20.
- Steck, P. (1997). Psychologische Testverfahren in der Praxis: Ergebnisse einer Umfrage unter Testanwendern. *Diagnostica*, 43, 267–284.
- Sturm, W. & Zimmermann, P. (2000). Aufmerksamkeitsstörungen. In W. Sturm, M. Herrmann & C.W. Wallesch (Hrsg.), *Lehrbuch der Klinischen Neuropsychologie: Grundlagen, Methoden, Diagnostik, Therapie*. Lisse: Swets & Zeitlinger.
- Tellegen, P. J. & Laros, J. A. (2011). *Test non verbal d'intelligence : SON-R 2 1/2-7*. Paris: Editions Hogrefe France.
- Zhu, J. & Chen, H. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*, 29 (6), 570–580.