



HAL
open science

La mesure de la personnalité à l'aide du Big Five Inventory français (BFI-Fr) est-elle invariante selon le genre et l'activité des répondants ?

B. Lignier, F. Lheureux

► **To cite this version:**

B. Lignier, F. Lheureux. La mesure de la personnalité à l'aide du Big Five Inventory français (BFI-Fr) est-elle invariante selon le genre et l'activité des répondants ?. *L'Encéphale*, 2022, 48 (1), pp.52-59. 10.1016/j.encep.2021.02.006 . hal-03648345

HAL Id: hal-03648345

<https://hal.science/hal-03648345v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

La mesure de la personnalité à l'aide du Big Five Inventory français (BFI-Fr) est-elle invariante selon le genre et l'activité des répondants ?

Is personality assessment with the French Big Five Inventory (BFI-Fr) invariant according to gender and activity of the respondents?

Baptiste Lignier^{1, 2*}, Florent Lheureux³

1 : Laboratoire Psychologie : Dynamiques Relationnelles et Processus Identitaires (Psy-DREPI, EA 7458), Département de Psychologie, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

2 : CSAPA La Santoline, SEDAP, Dijon, France

3 : Laboratoire de psychologie EA3188, Univ. Bourgogne Franche-Comté, F-25000 Besançon, France

* **Auteur correspondant** : Baptiste Lignier

Laboratoire Psy-DREPI, EA 7458, Département de psychologie, Université de Bourgogne Franche-Comté, Esplanade Erasme, Boulevard du Docteur Petitjean, 21 000 Dijon, France

e-mail : baptiste.lignier@u-bourgogne.fr

Résumé

Objectifs : Le BFI-Fr est un outil efficace et robuste pour évaluer les cinq grandes dimensions de la personnalité. Toutefois, l'incohérence de certains items pouvant être due à un manque d'invariance entre groupes, nous avons examiné cette invariance selon le genre et l'activité des répondants. **Méthodes** : L'échantillon est composé de 1904 participants (1 232 femmes et 672 hommes, 1 159 étudiants et 745 professionnels) ayant rempli le BFI-Fr. La méthodologie de la théorie de la réponse à l'item (*Item Response Theory*) a été utilisée pour tester l'invariance via le test du fonctionnement différentiel (*Differential Item Functioning, DIF*). **Résultats** : La structure factorielle en cinq dimensions est retrouvée et les variances expliquées sont comparables pour chaque sous-groupe. Toutefois, huit items varient dans leur mesure selon le genre et six selon l'activité, mais toutes ces variations sont faibles (sauf une valeur DIF modérée). Les femmes ont des scores A et N supérieurs aux hommes et les professionnels présentent des scores C plus élevés et des scores N plus faibles que les étudiants. Lorsque l'on supprime les items non-invariants, l'effet du genre devient significatif en faveur des femmes sur C mais la différence reste minime. Pour N, la différence selon l'activité devient non significative. **Conclusions** : Le BFI-Fr permet bien une évaluation robuste des cinq grandes dimensions pour chaque sous-groupe, tout en nécessitant un contrôle de l'invariance intergroupe lors de son utilisation. Cela, afin de s'assurer que les différences intergroupes sur les scores E-A-C-N-O ne sont pas le reflet d'un artefact psychométrique.

Mots-clés : Personnalité, Modèle en cinq facteurs, Psychométrie, Invariance, Fonctionnement différentiel de l'item

Abstract

Objectives: The BFI-Fr is an effective tool to assess the five dimensions of the five-factor model of personality. The psychometric qualities are excellent in several validation studies. However, because the inconsistency of certain items may be due to a lack of invariance between groups, we examined this invariance according to the gender and activity of the respondents. **Methods**: The sample consists of 1,904 participants (1,232 women and 672 men; 1,159 students and 745 professionals). Recruitment was done by "snowball" from different studies conducted over four years. Respondents completed the BFI-Fr. The 45 items were corrected for the acquiescence bias. The Item Response Theory methodology was used to test the invariance using the differential item functioning (DIF) test. The T scores were subject to ANCOVA and correlations to compare the differences between subgroups. **Results**: Whatever the subsample, we found the same five-factor structure (similarities were especially high for C). The cumulative explained variances are comparable for all groups. The functioning of eight items varied between men and women; and six between activity groups, however, all but one DIF values were low. No item in domain A varied in its measure. ANCOVA results showed that women have higher A and N scores than men and that professionals have higher C and lower N scores than students. There is no interaction effect between gender and activity. When

we removed the non-invariant items, the effect of gender became significant in favor of women on dimension C even if the difference remained minimal. The difference by activity on N became non-significant. There was a very strong link between activity and age and between gender and activity: in the sample, women were more often students and vice versa. *Conclusions*: Overall and once again, the structure and internal consistency of the Big Five as measured by the BFI-Fr was replicated in new samples. There was a differential functioning of 12 items out of 45 (about a quarter). These differences were small since only one was of moderate magnitude. The deletion of these items did not strongly change the mean comparisons between groups. Therefore, our results do not call into question the usefulness of the BFI-Fr but invite to control the extent of this differential functioning. In all cases (with or without the non-invariant items) the analyses showed significant gender differences for A and N, differences which are typically found in many studies. Students and professionals differ in their levels of C and E. Certain inconsistencies are probably due to the translation from American to French, since the BFI-Fr is based on the lexical approach (in this context, item content was somewhat dependent of languages peculiarities). Thus, the BFI-Fr allows a robust evaluation of the five dimensions but requires a prior control of the invariance of item functioning in order to ensure that the intergroup differences in E-A-C-N-O mean scores do not reflect a psychometric artifact but genuine personality differences.

Keywords (3 à 5): Personality, Five Factor Model, Psychometrics, Invariance, Differential Item Functioning

Introduction

Le modèle en cinq facteurs propose une conception hiérarchique de la personnalité en grands niveaux d'abstraction (les « *Big Five* ») intitulés par les initiales des trois descripteurs prédominants : E pour Extraversion, Énergie, Enthousiasme ; A pour Agréabilité, Altruisme, Affection ; C pour Conscience, Contrôle, Contrainte ; N pour émotions Négatives, Névrosisme, Nervosité, et O pour Ouverture, Originalité, Ouverture d'esprit [1, 2]. Ce modèle est issu, en partie, de l'hypothèse lexicale qui postule que les qualificatifs de la personnalité existent dans le langage courant pour décrire les autres ou se décrire soi-même [3]. Le *Big Five Inventory* (BFI) [4] en permet une évaluation rapide (45 items). La version française, le BFI-Fr, montre de très bonnes qualités psychométriques : les coefficients de cohérence interne vont de 0,77 à 0,83 et les corrélations avec le NEO PI-R [5] vont de 0,69 à 0,79. Les analyses mettent en évidence cinq facteurs expliquant 42 % de la variance totale avec des saturations en général très satisfaisante. Les corrélations avec les dimensions équivalentes du NEO PI-R sont satisfaisantes et les corrélations avec les autres dimensions sont peu élevées [6, 7]. Il s'agit donc d'un outil solide, fiable et efficace. Ces qualités sont également retrouvées pour la forme hétéro-évaluative [8], la structure en cinq composantes expliquant 52 % de la variance totale, avec des saturations factorielles, cohérences internes et des corrélations test-retest très satisfaisantes.

Malgré cela, certains items présentent quelques incohérences, parfois liées à la traduction, comme l'item 29 « *Can be cold and distant* », traduit initialement par « Peut être froid et distant » puis par « Peut être dédaigneux, méprisant » pour correspondre à la culture française ; ou parfois liées à des saturations plus faibles sur le facteur théorique ou des doubles saturations [6]. Une des raisons possibles de ces incohérences pourrait être le manque d'équivalence (ou d'invariance) psychométrique entre sous-groupes de participants (par exemple entre hommes et femmes). Comme le précisent Teresi et Fleishman [9], « le manque d'équivalence de la mesure rend les comparaisons de groupes ambiguës : est-ce que la différence est « réelle », c'est-à-dire une fonction du construit mesuré, ou bien est-elle un artefact du processus de mesure, tel que des interprétations différentes des questions par différents groupes ? » [9, p. 33]. En effet, l'invariance du nombre de facteurs, des saturations, des corrélations et de l'homogénéité interne, n'a pas été à ce jour explicitement évaluée avec le BFI-Fr.

La recherche de l'invariance des dimensions du modèle en cinq facteurs de la personnalité est importante car dans chaque étude de validation française, les auteurs observent une différence liée au genre sur les dimensions A, C et N [6] ; sur E, C et N en hétéro-évaluation [8] ; sur toutes les dimensions avec le BFI-10 mais avec une taille d'effet négligeable [10]. Les femmes présentent des scores plus élevés en N, E, A et C mais ces différences sont moins fortes dans les catégories socio-professionnelles moins favorisées [11]. Ces résultats sont répliqués de façon très régulière dans les études mais il peut y avoir des différences en fonction des cultures et des pays [12]. Par exemple, dans l'étude de Chiorri, Marsch, Ubbiali et Donati (2016), les femmes obtiennent des scores significativement plus hauts que les hommes uniquement en N, A.

De même, pour l'âge, les adultes d'âge moyen présentent des scores plus bas en N et plus hauts en A et C que les adultes plus jeunes [13, 14] et ce dans plusieurs pays [15-17]. Les différences sont moins robustes pour E et O avec des études aux résultats contradictoires. Les évolutions longitudinales ne sont pas linéaires si bien que les résultats diffèrent d'une étude à une autre. Il y a beaucoup plus de finesse au niveau des facettes ou même au niveau des items. Par exemple, Soto et John rapportent une diminution avec l'âge pour les facettes de rumination et de dépression, dans N, mais pas pour les facettes anxiété et irritabilité de la même dimension [18].

Pour mieux comprendre les différences liées à l'âge dans la personnalité, il est nécessaire d'évaluer les traits au niveau des cinq grandes dimensions et des facettes. Cependant, la recherche à ce niveau ne peut fournir qu'un aperçu approximatif incomplet de ces différences [14]. En effet, certains auteurs supposent l'existence d'un niveau hiérarchique supplémentaire, fait de « nuances », s'exprimant au niveau des items considérés isolément, reflète des variations interpersonnelles et intergroupes, liées aux différences de rôles sociaux, de style de vie ou à d'autres circonstances [19]. Dans ce cadre, les items sont conçus pour être des indicateurs directs de la personnalité, conformément à l'hypothèse lexicale qui propose d'étudier le langage naturel pour évaluer les différences interindividuelles [2, 19]. Dans cette perspective, il s'avère nécessaire d'employer des techniques spécifiquement centrées sur le niveau item, telles que celles examinant l'invariance des relations item-trait dans la lignée de la théorie de la réponse à l'item, comme le test du fonctionnement

différentiel de l'item de Mantel-Hanszel [9, 20]. L'avantage de ces techniques par comparaison aux techniques factorielles est qu'elles examinent la relation entre le trait mesuré et la réponse obtenue à un item sans imposer que cette relation soit linéaire [21]. En outre, la relation item-trait est plus finement examinée, car ces techniques considèrent de façons distinctes et comparatives la probabilité avec laquelle chaque modalité de réponse est choisie aux différents niveaux du trait latent mesuré [22]. Dans le cas d'une mesure invariante, la probabilité du choix d'une réponse à un item par deux personnes identiques sur le trait mesuré ne doit pas être significativement différente.

Objectif

L'objectif de cette étude est d'examiner l'invariance des cinq grandes dimensions, en particulier entre un échantillon d'étudiants et un échantillon de professionnels, ainsi qu'en comparant les réponses des hommes et des femmes. Cette évaluation s'avère d'autant plus nécessaire que, dans chaque étude de validation française, on observe une différence liée au genre sur les dimensions A, C et N en auto-évaluation [6] et sur E, C, et N en hétéro-évaluation [8]. Il est donc nécessaire de systématiquement vérifier les différences selon les genres. En outre, les participants inclus dans les études d'adaptation et de validation du BFI-Fr sont issus en majorité de la population étudiante [6, 7, 8] ou, plus rarement, de populations cliniques [23, 24]. De ce fait, une étude examinant les propriétés psychométriques de l'outil sur une population non-clinique et en activité professionnelle reste à effectuer, en particulier dans l'optique d'effectuer des comparaisons entre groupes. Par ailleurs, afin de s'approcher d'une conceptualisation exhaustive de la personnalité incluant la notion de nuances, nous avons simultanément étudié l'invariance au niveau des dimensions mais aussi au niveau des items, par le recours à des analyses factorielles exploratoires et de fonctionnements différentiels des items. À termes, ce type d'études permettra le déploiement de modèles développementaux longitudinaux de la personnalité. Il s'agit non seulement d'un objectif psychométrique, mais aussi théorique.

Méthode

Participants

L'échantillon est composé de 1904 participants dont 1 159 étudiants et 745 non-étudiants (professionnels de tous horizons) ; 672 participants sont des hommes (35,29 %). L'âge moyen des étudiants est de 20,29 ans et l'âge moyen des actifs est de 39,16 ans. Cet échantillon a été progressivement constitué entre 2011 et 2015, période au cours de laquelle le BFI-Fr a été inclus dans douze enquêtes d'initiation à la recherche en psychologie, sous la supervision d'un enseignant-chercheur. L'essentiel des participants ont répondu sous format électronique ($n = 1693$), la prise de contact s'étant effectuée via des réseaux de convenance (personnels et sociaux). Ils ont été invités à proposer le questionnaire à des connaissances afin d'accroître la participation et diversifier le profil des répondants (recrutement par « boule de neige »). Le même principe a été employé pour les questionnaires papier-crayon.

Outils

Tous les participants ont rempli le BFI-Fr [6], composé de 45 items accompagnés d'une échelle de Likert en cinq niveaux : de 1 (désapprouve fortement) à 5 (approuve fortement). Les scores des cinq grandes dimensions sont calculés grâce à la moyenne des items appartenant à l'échelle après inversion des items concernés.

Analyse des données

Les scores aux 45 items ont fait l'objet d'une correction du biais d'acquiescement [25]. Des analyses en composantes principales ont été effectuées à partir des groupes et ce, avec rotation oblimin direct, car les facteurs peuvent être corrélés entre eux. Ensuite, la méthodologie de la théorie de la réponse à l'item (*Item Response Theory*, IRT) a été utilisée à l'aide du programme jMetrik [20] (voir également : <https://itemanalysis.com/jmetrik-download/>). Elle a permis de tester l'invariance et de mettre en évidence les items présentant un fonctionnement différentiel par genres ou par groupes d'activité avec le test du *Mantel-Haenszel Differential Item Functioning* (DIF) [9, 20]. Ce test, dans sa version généralisée, est adapté lorsque deux groupes sont comparés à partir d'une mesure composée d'items polytomiques et pouvant potentiellement présenter des formes variées de fonctionnement différentiel [26]. Par ailleurs, il peut aussi s'appliquer à des données conformes à un modèle de réponses graduées [27] et présente un risque d'erreur de Type 1 modéré et une puissance adéquate [28].

Le test DIF examine « dans quelle mesure, après contrôle du construit sous-jacent, la réponse à un item est liée à l'appartenance groupale » [20, p. 34]. Il estime la force du lien entre la réponse à l'item et l'appartenance de groupe en contrôlant l'effet de la variable latente, au moyen d'une valeur DIF. Lorsque cette valeur est égale à zéro, la réponse à l'item ne varie pas d'un groupe à l'autre mais seulement en fonction de la variable latente mesurée. Inversement, plus elle est non-nulle, plus l'appartenance de groupe a un impact sur la réponse à l'item, indépendamment de la variable mesurée. Dans ce cas, une valeur positive indique que l'item présente des scores plus élevés pour l'un des groupes (groupe focal), alors qu'une valeur négative indique des valeurs plus élevées pour l'autre groupe (groupe de comparaison). Une classification qualitative des items (fonctionnement différentiel faible, modéré ou fort) sur la base de valeurs DIF seuil est alors couramment appliquée (classification ETS). Avant d'effectuer cette classification, la valeur DIF est divisée par l'étendue des scores de l'échelle de mesure (ici 5) puis considérée en valeur absolue. Dès lors, si elle est inférieure à 0,05 le fonctionnement différentiel est considéré faible, si entre 0,05 et 0,10 il est considéré modéré et s'il est au-dessus de 0,10 il est considéré fort [20]. Conformément à la méthodologie recommandée [9, 20] nous avons procédé de manière itérative. Tout d'abord, la valeur DIF pour chaque item a été estimée en prenant en référence comparative la distribution du score total de tous les items mesurant le construit théorique. Cette analyse a été répétée une seconde fois en excluant du score total les items présentant les plus hautes valeurs DIF (classées comme modérées ou fortes) au cours de la première étape (procédure de purification). Cela est de nouveau répété si un item supplémentaire occasionne à nouveau de fortes variations entre groupes (nouvelle purification). Les valeurs DIF ainsi que leurs classifications qualitatives rapportées dans le Tableau 1 correspondent à la dernière itération.

L'examen du DIF nécessite d'avoir au préalable mis en évidence la dimensionnalité latente aux différents items employés. A cette fin, l'application d'une analyse factorielle s'avère particulièrement adaptée, en ce qu'elle permet de simultanément examiner l'association importante de l'item avec son facteur théorique, tout en vérifiant la faible relation de l'item avec les autres facteurs théoriques [21, 29]. Le test du Mantel-Haenszel [9, 20] a donc été réalisé après avoir examiné les résultats des analyses en composantes principales.

Les moyennes et écarts-types rapportés dans le Tableau 2 ont été transformés en scores T, sur la base des moyennes marginales estimées lors d'une ANCOVA contrôlant l'effet du genre, de l'âge et de l'activité, conformément à Soto, John, Gosling et Potter [14]. Cela a permis d'établir comme scores de référence une moyenne de 50 et un écart-type de 10 sans que ceux-ci soient le reflet d'un groupe de genre, d'un groupe d'âge ou d'activité plus qu'un autre. Selon Cohen [30] une différence de 2 points représente une taille d'effet faible, une différence de 5 points représente une taille d'effet moyenne et une différence de 8 points représente une taille d'effet importante. Dans le même esprit, les corrélations du Tableau 3 sont partielles, c'est-à-dire qu'elles reflètent les relations entre facteurs de personnalité en contrôlant l'effet des variables tiers. Cela a permis de comparer les différences de corrélations observées entre sous-groupes (ex. hommes vs. femmes) tout en contrôlant l'effet des autres variables (ex. âge et activité). En outre, les relations entre le genre, l'âge et le groupe d'activité ont été examinées à l'aide d'ANOVA à un facteur ou d'un test du khi-deux d'indépendance (χ^2). Enfin, les analyses statistiques ont été menées à l'aide du logiciel SPSS v22 et la valeur de significativité retenue est 0,01 étant donné la taille de l'échantillon.

Résultats

[Invariance des dimensions du BFI-Fr selon les sous-groupes](#)

Le Tableau 1 présente les saturations factorielles observées par sous-échantillons ainsi que les comparaisons inférentielles. Quel que soit le découpage de l'échantillon, on retrouve bien la structure des cinq dimensions. C'est surtout le cas pour C pour laquelle les saturations sont comprises entre 0,46 et 0,77 selon le genre, et entre 0,47 et 0,75 selon l'activité, sans aucune saturation secondaire. Bien que certaines saturations principales apparaissent peu élevées (cinq sur 36 sont inférieures à 0,50 ; neuf sont inférieures à 0,60), les saturations secondaires sont toutes faibles, n'excédant pas 0,21 ; mettant alors en évidence un modèle clair. C'est un peu moins évident pour les autres dimensions. En effet, on observe pour E une saturation secondaire importante en C pour l'item 11 chez les hommes (0,36) et chez les non-étudiants (0,36) alors que la saturation principale est seulement légèrement supérieure (0,39 pour les hommes ; 0,38 pour les non-étudiants). Une autre saturation secondaire apparaît à 0,32 pour l'item 26, en O uniquement chez les non-étudiants alors que la saturation principale est à -0,42. Pour le reste, les saturations principales oscillent entre 0,38 et 0,78 en valeur absolue et les saturations secondaires entre 0 et 0,29 si on fait exception des plus élevées mentionnées ci-dessus. Pour A, trois saturations secondaires pour l'item 42 chez les hommes (0,35) et les femmes (0,30) ainsi que chez les étudiants (0,35), posent problème car les saturations principales sont

respectivement de -0,38¹ ; 0,45 et 0,43. Les autres saturations varient entre 0,32 et 0,71 pour les principales et entre 0 et 0,24 pour les secondaires. Pour N, l'item 29 sature plus (0,38) dans A que dans N (0,26), pour les hommes. On retrouve le même problème selon l'activité car il sature plus sur A pour les non-étudiants (-0,35) que dans N (0,27). Pourtant les autres saturations principales sont comprises entre 0,30 et 0,80 alors que les saturations secondaires sont comprises entre 0 et 0,27. Pour O, trois items posent problème. L'item 15 présente une saturation secondaire plus élevée en C chez les étudiants (0,32) qu'en O (0,30). L'item 41 a une saturation principale faible chez les hommes (-0,26) et chez les non-étudiants (-0,22). L'item 35 présente lui aussi une saturation principale faible chez les hommes (-0,24) et les étudiants (-0,24). Les autres saturations principales varient entre 0,31 et 0,80 alors que les saturations secondaires varient entre 0 et 0,29. La variance expliquée cumulée est de 39,93 % pour les hommes, de 40,89 % pour les femmes, de 40,90 % pour les étudiants et de 38,65 % pour les non-étudiants. Elles sont donc comparables pour tous les groupes et proches de celles des études de validation française.

Invariance des items selon le test de Mantel-Haenszel

En ce qui concerne les résultats de la méthodologie de la théorie de la réponse à l'item, huit items varient modérément selon le genre (trois items de O, deux items de E, deux items de N et un item de C) ; cinq items varient modérément et un item fortement selon l'activité (un item de O, deux items de E, un item de N et deux items de C). Parmi ces items, deux varient dans toutes les comparaisons (N29 et E1). Plus précisément, les items N4 (DIF = 0,22), N29 (DIF = 0,21), C28 (DIF = 0,22), O15 (DIF = 0,30), E26 (DIF = 0,20) ont obtenu des scores plus élevés pour les hommes, alors que pour les femmes ce sont les items O30 (DIF = -0,39), O41R (DIF = -0,33) et E1 (DIF = -0,32) qui ont occasionné des scores plus élevés. En ce qui concerne l'activité, les items E31R (DIF = 0,27) et C3 (DIF = 0,25) ont obtenu des scores plus élevés chez les participants en activité professionnelle, alors que les items N29 (DIF = -0,34), C8R (DIF = -0,21), O44 (DIF = -0,31) et E1 (DIF = -0,49) ont fait l'objet de scores plus élevés pour les étudiants. Cela fait donc au total 12 items sur 45 ayant au moins une variation entre groupes. Toutefois, selon la classification ETS, seul un item présente une valeur DIF modérée, les onze autres ayant des valeurs DIF faibles. Aucun item du domaine A n'a varié dans sa mesure.

[Différences selon le genre](#)

Le Tableau 2 présente les statistiques descriptives pour chaque facteur, pour chaque groupe, ainsi qu'en fonction de la conservation ou non dans l'inventaire des items avec un fonctionnement différentiel. Sans suppression des items non-invariants entre groupes (BFI-Fr complet), les résultats de l'ANCOVA montrent des effets du genre sur A ($F = 17,72$; $p < 0,01$; $\eta^2_p = 0,01$) et sur N ($F = 83,84$; $p < 0,01$; $\eta^2_p = 0,04$), les femmes ayant des scores A (56,89) et N (47,76) supérieurs aux hommes (respectivement 55,47 et 42,37).

¹ Toutes les saturations observées pour A chez les hommes sont inversées par comparaison avec les autres sous-groupes. Cela n'est cependant pas problématique car le signe attribué à un item par l'analyse est indépendant du fait que l'item mesure positivement (item standard) ou négativement (item à inverser) la dimension. C'est l'item qui sature le plus qui détermine le signe du facteur. Or, pour les hommes c'est un item standard qui sature le plus, alors que pour les trois autres sous-groupes, c'est un item inversé. Une fois qu'on inverse le codage des items le nécessitant les saturations sont toutes positives.

Lorsque l'on supprime les items non-invariants, l'effet du genre devient significatif en faveur des femmes sur la dimension C même si la différence reste minime.

Le Tableau 3 présente les corrélations partielles entre les cinq facteurs ainsi que les coefficients d'homogénéité interne au niveau de l'échantillon total et par sous-échantillons. Pour l'ensemble de l'échantillon, on observe de nombreuses corrélations significatives, cependant, elles sont comprises entre 0,05 et 0,29. La corrélation la plus élevée concerne N et E (-0,29) puisque l'une concerne la sensibilité aux émotions négatives et l'autre aux émotions positives. Il y a moins de corrélations significatives chez les hommes que chez les femmes, même si les tendances sont les mêmes que dans l'échantillon total. Les alphas de Cronbach, compris entre 0,69 et 0,83, sont très satisfaisants. Ils sont plus bas pour les hommes en A et O mais très proches du seuil traditionnel de 0,70.

Différences selon l'activité

Les résultats montrent également des effets de l'activité sur C ($F = 78,30 ; p < 0,01 ; \eta^2_p = 0,04$) et sur N ($F = 7,54 ; p < 0,01 ; \eta^2_p = 0,004$), les professionnels ayant des scores C (55,29) plus élevés que les étudiants (47,73), des scores E (47,85) plus élevés que les étudiants (46,61) et des scores N plus faibles (43,50) que les étudiants (47,38). En revanche, il n'y a pas d'effet d'interaction entre le genre et l'activité. L'âge (covariable) est significativement mais faiblement associé positivement à C ($F = 23,41 ; p < 0,01 ; \eta^2_p = 0,01 ; r_p = 0,11$) et à A ($F = 13,57 ; p < 0,01 ; \eta^2_p = 0,01 ; r_p = 0,08$), ainsi que négativement à N ($F = 7,09 ; p < 0,01 ; \eta^2_p = 0,01 ; r_p = -0,06$). Pour N, la différence selon l'activité devient non significative pour les deux versions réduites (bien qu'il subsiste des différences notables). Les alphas de Cronbach, également compris entre 0,69 et 0,83, sont très satisfaisants mais sont plus bas pour les non-étudiants en O.

Très logiquement, il existe un lien très fort entre l'activité et l'âge puisque l'âge moyen des non-étudiants est de 39,16 ans alors qu'il est de 20,29 ans pour les étudiants ($F = 1 198,90 ; p < 0,0001$ en ANOVA à un facteur ; valeur F corrigée par la méthode de Welch pour cause de variances inégales au sein des deux groupes selon le test de Levene ; $2 123,67 ; p < 0,0001$). Il y a aussi un lien entre le genre et l'activité : dans l'échantillon, les femmes sont plus souvent étudiantes que les hommes et inversement ($\chi^2(1) = 25,44 ; p < 0,0001$). Ces éléments confirment donc la nécessité de contrôler les effets des variables (ce qui est fait dans les Tableaux 2 et 3).

Discussion

L'objectif de cette étude est d'examiner l'invariance du BFI-Fr à plusieurs niveaux : au niveau des dimensions, mais aussi au niveau des items, cela en fonction du genre et de l'activité professionnelle (étudiant-e-s/professionnel-le-s).

Invariance des items

De façon générale, encore une fois, la structure et la consistance interne des *Big Five* est retrouvée, dans de nouveaux échantillons. Toutefois, on observe un fonctionnement différentiel de 12 items (près d'un quart).

Nos résultats révèlent que ces différences sont de faible ampleur car une seule est modérée, les autres étant faibles. Par ailleurs, en vérifiant *a posteriori* les résultats de la suppression de ces items, cela modifie peu les comparaisons entre les groupes (la nouvelle différence significative est de très faible amplitude). Dès lors, nos résultats ne remettent pas en question l'utilité du BFI-Fr, mais invitent à contrôler l'ampleur de ce fonctionnement différentiel. Il est donc tout à fait possible d'utiliser le BFI-Fr afin de procéder à des études sur le comportement des items afin d'évaluer les différences interindividuelles et les nuances.

Indépendamment des comparaisons de groupes, seuls quelques items saturent dans d'autres dimensions, phénomène déjà présent dans les études de validation du BFI-Fr [6, 8]. Par exemple, l'item 29 présente probablement un problème de traduction/transposition culturelle de la notion d'humeur changeante, pouvant être interprétée dans notre culture comme relevant du registre des relations à autrui et donc de l'Agréabilité. Même chose pour l'item 42 « *aime coopérer avec les autres* » de A qui sature aussi sur E puisque les deux dimensions sont les deux facteurs prosociaux. L'item 15 « *est ingénieux, une grosse tête* » a tendance également à saturer en C puisque les termes utilisés pour décrire ce trait symbolisent plus en France la possession d'un potentiel intellectuel qu'une ouverture d'esprit face à un problème. C'est aussi le cas pour l'item 35 (« *préfère un travail simple et routinier* ») qui pourrait correspondre aux facettes ordre et autodiscipline de C. En revanche, les items de la dimension A ne présentent pas de telles limites. Ce type de raisonnement, au niveau des items, est d'autant plus important que le BFI-Fr est issu d'un travail d'analyse du langage courant. C'est pourquoi de plus en plus d'études s'intéressent aux nuances. Il s'agit d'un niveau intermédiaire, sous les facettes, au niveau de la réponse à l'item, qui permet de capter un concept supplémentaire indépendant, qui pourrait participer aux modèles théoriques sous-jacents [19].

Invariance des dimensions du BFI-Fr

Les analyses intergroupes font apparaître des différences significatives faibles en A et moyennes en N que l'on conserve ou supprime les items non-invariants. Ce sont des différences que l'on retrouve classiquement dans de nombreuses études de comparaison selon le genre. Il est même surprenant que nous n'observions pas de différence sur les autres dimensions sur lesquelles les femmes obtiennent habituellement des scores supérieurs. En effet, les femmes ont tendance à être plus extraverties, plus agréables, plus consciencieuses et plus sensibles aux émotions négatives que les hommes, ces résultats s'observant à l'échelle internationale avec le BFI [11] mais aussi avec le NEO PI-R [31]. Tout comme dans notre étude, la différence la plus importante correspond à N, significative dans 49 pays (sur 55) alors qu'elle est significative dans 34 pays pour A, moins fréquente pour E (25 pays) et C (23 pays). Cette différence, toujours difficile à interpréter, peut être liée aux variations socio-économiques, l'espérance de vie, l'éducation, ou encore des prédispositions génétiques.

Les étudiants et les non-étudiants se distinguent systématiquement sur C et sur N lorsque l'échelle complète est utilisée. L'évaluation sur cette dimension est plus aisée dans un cadre professionnel car cela nécessite de se projeter dans des tâches et dans des capacités adaptatives liées à la performance. C'est également une dimension qui représente un caractère fortement désirable et utile, en particulier sur le marché du travail

[32]. En ce qui concerne les différences sur N, il est possible que les jeunes soient plus sujets à des troubles psychiatriques mineurs car plus proches de l'adolescence, une période développementale d'émancipation vis-à-vis des repères parentaux et d'incertitudes en termes d'avenir (insertion professionnelle, reconnaissance sociale, aspects financiers...) [33, 34].

Bien sûr, l'analyse des différences au niveau des items est indissociable de l'analyse des différences au niveau des dimensions. Ainsi, les différences liées au genre et à l'activité au niveau des dimensions peuvent s'expliquer par les différences obtenues au niveau des items. Par exemple, l'item qui obtient la valeur DIF la plus élevée (0,49) est l'item E1 « est bavard ». Cet item, à lui seul, capte probablement, la nuance qui permet de mieux comprendre la différence entre les étudiants et les non-étudiants, qui ne se joue pas uniquement au niveau de l'Extraversion mais bien dans le fait d'être plus ou moins bavard (comportements habituels). Les variations observées sur les autres items de E (E26, E31R) sont plus faibles. C'est le contraire pour la dimension A pour laquelle aucun item ne varie. La différence significative observée entre les hommes et les femmes s'explique exclusivement au niveau de la dimension, c'est-à-dire dans la conceptualisation globale de l'agréabilité. Autre cas de figure, pour O, nous observons uniquement des variations au niveau des items mais aucune différence significative au niveau de la dimension.

Limites

Cette étude présente plusieurs limites. La première concerne l'échantillonnage. En effet, la méthode par « boule de neige » a pu entraîner des biais de recrutement, notamment pour la répartition hommes/femmes. Ainsi, il pourrait être intéressant de vérifier l'invariance dans une étude systématisée, en analysant également les variations culturelles. La seconde limite est d'ordre méthodologique. Ici, l'invariance est étudiée à travers les différences au niveau des items et au niveau des dimensions dans plusieurs sous-groupes. Il serait intéressant de l'étudier en fonction de la stabilité dans le temps, la fidélité selon plusieurs évaluations, ou encore la validité prédictive. Enfin, il faudrait mettre en regard ces résultats avec des méthodes longitudinales et des mesures répétées afin d'affiner les connaissances théoriques et les modèles développementaux. De plus, il existe une limite conceptuelle car chaque outil propose des concepts et des dimensions opérationnelles différentes. Par exemple, les dimensions du BFI-Fr ne recourent pas nécessairement celles du NEO PI-3 et encore moins au niveau des facettes [1]. Qu'en est-il au niveau des nuances et de l'invariance ? Le BFI-Fr a été créé pour évaluer les cinq grands domaines et, plus récemment, des facettes mais pas les nuances (idiosyncrasies). Les items du BFI-Fr ne sont donc pas censés refléter ces nuances. Nos résultats, et notamment les résultats DIF nous permettent seulement d'envisager l'existence de ce niveau supplémentaire. C'est une question à explorer dans les prochaines recherches et ce, probablement, plus avec le BFI-2 qui possède plus d'items, ou encore mieux, avec les grands inventaires comme le NEO PI-3.

Conclusion

Les résultats nous poussent à envisager de reproduire ce type d'étude avec tous les outils et, en premier lieu, avec le BFI-2, dont l'étude de validation est en cours. La méthode de la théorie de réponse à l'item (IRT), facile d'accès, pourrait permettre d'étudier systématiquement l'invariance des outils de mesure afin d'en améliorer les évaluations, toutes disciplines et objectifs confondus. Ces éléments, jusqu'alors non considérés dans les études utilisant le BFI, confortent dans l'ensemble sa robustesse en termes d'invariance intergroupes, les écarts à cette invariance étant peu importants. Ainsi, cet outil permet une évaluation robuste des cinq grandes dimensions, avec cependant un contrôle préalable de l'invariance du fonctionnement des items. Cela afin de s'assurer que les comparaisons de moyennes aux scores reflètent bien l'influence des groupes et non un artefact psychométrique.

Les auteurs déclarent ne pas avoir de liens d'intérêts.

Références

- [1] John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues. In: John OP, Robins RW, Pervin LA, Ed. Handbook of personality: theory and research, 3rd ed. New York, NY, US: Guilford Press; 2008: 114–58.
- [2] Plaisant O, Guertault J, Courtois R et al. Histoire des « Big Five » : OCEAN des cinq grands facteurs de la personnalité. Introduction du Big Five Inventory français ou BFI-Fr. *Anna Med Psychol* 2010;168: 481-486.
- [3] Goldberg LR. An alternative “description of personality”: The Big-Five factor structure. *J Pers Soc Psychol* 1990;59: 1216-29.
- [4] John OP, Donahue EM, Kentle RL. The Big Five Inventory—versions 4a and 54. Berkeley, CA, US: University of California, Institute of Personality and Social Research ; 1991.
- [5] Rolland JP. Manuel du NEO PI-R. Inventaire de personnalité révisé. Paris : ECPA ; 1998.
- Costa PT, McCrae RR. NEO PI-R professional manual. Odessa, FL, US: Psychological Assessment Resources, Inc; 1992.
- [6] Plaisant O, Courtois R, Réveillère C, et al. Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Ann Med Psychol* 2010;168: 97–106.
- [7] Courtois R, Petot JM, Lignier B, et al. Le Big Five Inventory français permet-il d'évaluer des facettes en plus des cinq grands facteurs ? *L'Encéphale* 2018; 44: 208-214.
- [8] Lignier B, Petot JM, Plaisant O, Zebdi R. Validation du Big Five Inventory français (BFI-Fr) dans un contexte d'hétéro-évaluation. *Ann Med Psychol* 2016;174(6): 436-441.
- [9] Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Quality of Life Res* 2007;16(1): 33-42.
- [10] Courtois R, Petot, JM, Plaisant O, et al. Validation française du Big Five Inventory à 10 items (BFI-10). *L'Encéphale* 2020,46 :455-462.
- [11] Schmitt DP, Realo A, Voracek et al. Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *J Pers Soc Psychol* 2008;94(1): 168–182.

- [12] Chiorri C, Marsh HW, Ubbiali A, Donati D. Testing the Factor Structure and Measurement Invariance Across Gender of the Big Five Inventory Through Exploratory Structural Equation Modeling. *J Pers Assess* 2016;98(1): 88-99.
- [13] McCrae RR, Costa PT, de Lima MP, et al. Age differences in personality across the adult life span: Parallels in five cultures. *Dev Psychol* 1999;35: 466-477
- [14] Soto C, John OP, Gosling SD, Potter J. Age Differences in Personality Traits From 10 to 65: Big Five Domains and Facets in a Large Cross-Sectional Sample. *J Pers Soc Psychol* 2011;100(2): 330-348.
- [15] McCrae RR, Terracciano A, Personality Profiles of Cultures Project. Personality profiles of cultures: Aggregate personality traits. *J Pers Soc Psychol* 2005;89(3): 407-425.
- [16] Donnellan MB, Lucas RE. Age differences in the big five across the life span: Evidence from two national samples. *Psychol Aging* 2008;23: 558-566.
- [17] Lucas RE, Connellan MB. Age differences in personality: Evidence from a nationally representative Australian sample. *Dev Psychol* 2009;45(5) :1353-1363.
- [18] Soto CJ, John OP. Development of Big Five Domains and Facets in Adulthood: Mean-Level Age Trends and Broadly Versus Narrowly Acting Mechanisms. *J Pers* 2012;80(4) :881-914.
- [19] Möttus R, Rozgonjuk D. Development is in the details: Age differences in the Big Five Domains, Facets and Nuances. *J Pers Soc Psychol* 2020 (Advance online publication).
- [20] Meyer J. *Applied Measurement with iMetrik*. New York, NY, US: Routledge; 2014.
- [21] De Ayala RJ, Hertzog MA. The Assessment of Dimensionality for Use in Item Response Theory. *Multivariate Behav Res* 1991;26(4):765-792.
- [22] Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Qual Quant* 47(4), 2341-2360.
- [23] Bucourd E, Martailé V, Mulleman D et al. Comparison of the Big Five personality traits in fibromyalgia and other rheumatic diseases. *Joint Bone Spine* 2017;84(2): 203-207.
- [24] Plaisant O, Srivastava S, Mendelsohn GA et al. Relations entre le Big Five Inventory français et le manuel diagnostique des troubles mentaux dans un échantillon clinique français. *Ann Med Psychol* 2005;163: 161-7.
- [25] Soto C, John OP. The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *J Pers Soc Psychol* 2017;113(1): 117-143.
- [26] Fidalgo ÁM. A New Approach for Differential Item Functioning Detection Using Mantel-Haenszel Methods. The GMHDIF Program. *Span J Psychol* 2011;14(2):1018-1022.
- [27] Fidalgo ÁM, Bartram D. A Comparison Between Some Generalized Mantel-Haenszel Statistics for Detecting DIF in Data Simulated Under the Graded Response Model. *Appl Psychol Meas* 2010;34(8):600-606.
- [28] Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: A comparison of three Mantel-Haenszel procedures. *Appl Meas Educ* 14(3), 235-260.
- [29] Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Stud Educ Eval* 68, 100963.
- [30] Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
- [31] Costa PT, Terracciano A, McCrae RR. Gender differences in personality traits across cultures: robust and surprising findings. *J Pers Soc Psychol* 2001;81(2): 322-331.
- [32] Crisanta-Alina M, Gangloff B, Auzoult L. Désirabilité et utilité sociale des traits du Big Five. *Psicol Soc* 2011;28(11): 33-51.
- [33] Jones PB. Adult mental health disorders and their age at onset. *Br J Psychiatry* 2013;54: s5-10.

[34] OVE (Observatoire de la vie étudiante). Enquête nationale Conditions de vie des étudiants 2016.
www.ove-national.education.fr

Tableau 1. Saturations factorielles observées par sous-échantillons (matrices des types) et fonctionnement différentiel de chaque item par rapport à son facteur théorique (valeurs DIF).

Items	Comparaisons par genre										Comparaisons par activité										DIF	
	N		C		A		O		E		N		C		A		O		E			
	H	F	H	F	H	F	H	F	H	F	E	NE	E	NE	E	NE	E	NE	E	NE		
N39	.78	.80	.10	.11	-.06	.04	-.06	-.03	-.12	-.11	0,03	-.80	.78	.12	.10	.05	.02	-.02	-.09	-.08	.15	0,07
N14	.79	.76	.10	.12	-.14	.06	-.03	.00	-.05	-.03	-0,10	-.77	.78	.14	.12	.07	.12	-.02	-.03	-.02	.04	0,06
N9R	-.76	-.75	-.07	-.04	-.03	.01	-.02	-.01	-.01	.02	-0,05	.77	-.72	-.09	-.06	-.01	.09	-.01	.01	.01	-.01	0,02
N19	.73	.68	.02	-.01	-.07	.00	-.02	.05	-.09	-.11	0,03	-.68	.76	.00	.07	.00	.06	.05	-.04	-.12	.07	-0,02
N34R	-.60	-.68	.04	.11	-.11	.02	.02	.03	-.11	-.19	-0,01	.69	-.61	.04	.07	.05	.10	.04	.03	-.18	.13	-0,14
N24R	-.58	-.53	.06	.00	-.18	.02	-.01	.06	-.20	-.16	-0,12	.58	-.54	.00	.07	.07	.07	.04	.05	-.15	.24	0,08
N4	.51	.44	-.03	-.07	.10	-.11	.13	.13	-.21	-.17	0,22*	-.45	.50	-.08	-.01	-.11	-.16	.13	.14	-.17	.24	-0,06
N29	.26	.30	-.21	-.16	.38	-.24	-.06	.02	.00	-.09	0,21*	-.30	.27	-.16	-.15	-.27	-.35	.00	-.02	-.05	.06	-0,34*
C3	.07	.08	.74	.77	-.05	.02	-.07	-.06	-.05	-.05	-0,05	-.14	.11	.75	.67	.02	.12	-.04	-.11	-.04	-.02	0,25*
C33	-.03	-.01	.75	.72	-.02	.01	.07	-.03	-.02	.06	0,04	.02	.05	.72	.63	-.03	.12	.02	.00	.05	-.09	0,11
C13	.03	.03	.71	.70	-.07	.05	-.06	-.04	-.05	-.01	-0,03	-.09	.03	.70	.58	.06	.13	-.07	.01	.00	-.01	0,08
C23R	-.12	-.04	-.67	-.70	.09	-.07	.07	.06	-.02	-.03	0,06	.11	-.01	-.68	-.58	-.03	-.15	.03	.15	-.01	.08	0,13
C18R	-.07	-.04	-.63	-.62	-.04	.07	-.03	.03	.06	.07	0,02	.04	-.08	-.62	-.59	.06	.11	.04	-.04	.06	-.10	0,01
C43R	.09	.12	-.58	-.64	-.09	.14	-.04	-.02	.02	.15	-0,03	-.08	.12	-.62	-.56	.17	.02	-.01	-.05	.12	-.08	-0,19
C28	-.13	-.01	.61	.63	.01	.05	-.05	.08	-.08	.02	0,22*	.04	-.14	.59	.54	.06	-.07	.06	.04	-.02	.02	0,04
C8R	-.12	.00	-.48	-.51	.02	-.02	.00	.06	.02	-.01	-0,02	.02	-.08	-.49	-.60	-.05	.09	.07	.00	.03	-.03	-0,21*
C38	-.09	-.04	.46	.50	-.04	.09	.13	.15	.11	.15	0,01	.06	-.03	.48	.47	.07	.04	.13	.21	.16	-.11	-0,16
A27R	-.02	-.03	-.07	-.04	.61	-.71	.02	.01	.05	.09	-0,15	.03	-.07	-.05	-.06	-.69	-.64	.01	.04	.07	-.06	0,03
A32	.06	-.01	.04	.01	-.63	.61	-.04	-.04	.07	.04	0,03	.01	.01	.05	.03	.61	.60	-.04	-.04	.04	-.05	-0,12
A37R	-.10	-.08	-.05	-.08	.57	-.59	-.09	-.12	.11	.13	-0,13	.09	-.07	-.11	-.05	-.60	-.54	-.09	-.14	.10	-.18	-0,03
A12R	.11	.15	.07	.06	.58	-.56	-.06	-.15	.24	.23	0,01	-.16	.07	.04	-.01	-.57	-.57	-.13	-.09	.23	-.22	-0,07
A17	-.02	-.13	-.02	-.10	-.53	.45	-.07	-.02	.02	-.02	0,02	.10	-.07	-.07	-.03	.49	.48	-.03	-.07	-.04	-.05	-0,11
A2R	.02	.10	-.08	-.04	.43	-.43	-.11	-.15	.10	.19	0,09	-.06	.12	.01	-.10	-.44	-.45	-.20	.00	.14	-.20	0,16
A7	-.02	-.03	.05	.17	-.47	.42	-.04	-.09	.08	.10	0,02	.05	-.01	.13	.12	.41	.47	-.12	.01	.10	-.08	0,02
A42	.01	-.06	.08	.12	-.38	.45	-.01	-.03	.35	.30	0,08	.05	-.05	.05	.16	.43	.38	-.08	.08	.35	-.21	0,14
A45R	.17	-.13	.11	.03	.32	-.49	-.06	-.11	.08	.07	-0,13	.05	-.04	.10	.05	-.40	-.49	-.10	-.10	.03	-.10	0,16
A22	-.05	-.02	-.08	-.08	-.42	.37	.00	-.09	.21	.22	0,17	.01	-.17	-.06	-.10	.37	.39	-.09	-.01	.18	-.22	-0,18
O25	-.08	-.07	.14	-.02	.17	-.02	.76	.78	.05	.02	0,10	.07	-.12	.00	.09	-.05	-.14	.80	.72	.05	.04	0,07
O5	-.05	-.06	.13	.07	.08	.03	.72	.70	.03	.06	0,01	.04	-.10	.08	.04	-.01	.00	.74	.64	.07	-.02	0,11
O20	-.03	-.01	-.09	-.14	.16	-.06	.68	.70	.04	-.01	-0,02	.00	-.06	-.14	-.06	-.09	-.08	.71	.66	-.01	-.04	-0,14
O30	.14	.08	-.14	-.07	-.16	.07	.59	.64	-.07	-.09	-0,39*	-.16	.13	-.07	-.10	.10	.21	.64	.55	-.01	.13	-0,13
O44	.01	-.01	-.10	-.06	-.06	-.01	.54	.64	-.06	.01	-0,15	.02	.05	-.04	.00	.00	.11	.63	.51	.00	-.01	-0,31*
O40	-.01	.01	.00	.07	-.04	-.01	.54	.63	-.08	-.01	0,08	.02	-.03	.10	-.03	-.01	.04	.60	.60	-.02	.08	-0,04
O10	.01	-.05	.08	.07	-.04	.00	.38	.36	.15	.12	0,11	.09	.01	.10	.08	.00	-.01	.39	.31	.09	-.20	0,04
O15	-.18	.00	.29	.27	.23	-.13	.37	.32	.00	-.01	0,30*	.10	-.15	.32	.21	-.15	-.27	.30	.42	-.01	.05	0,02
O41R	-.07	-.05	.21	.06	.15	-.09	-.26	-.38	.08	.02	-0,33*	.07	-.16	.09	.16	-.11	-.22	-.39	-.22	.03	.05	0,01
O35R	-.01	-.05	-.01	.02	.14	-.10	-.24	-.31	-.12	-.24	-0,12	.07	-.02	-.01	.04	-.13	-.08	-.24	-.41	-.25	.12	0,06
E21R	-.11	-.13	.09	.08	-.08	.07	.00	-.02	-.76	-.75	-0,13	.13	-.14	.08	.12	.10	.05	.00	-.02	-.75	.78	-0,07
E36	-.08	-.07	-.06	.04	-.14	.08	-.01	.03	.69	.71	0,01	.08	-.08	-.06	.14	.09	.07	.03	-.02	.74	-.62	-0,03
E1	.10	.14	-.19	-.12	.04	-.03	-.03	-.05	.65	.72	-0,32*	-.17	.11	-.10	-.08	-.06	.00	-.05	-.05	.68	-.74	-0,49**
E6R	.09	.12	.10	.09	-.09	.20	-.06	-.09	-.68	-.69	0,07	-.12	.11	.10	.09	.17	.17	-.07	-.09	-.70	.66	-0,08
E16	-.10	-.11	.15	.08	-.24	.18	.08	.05	.60	.58	-0,14	.07	-.11	.08	.23	.25	.12	.02	.12	.60	-.54	-0,07
E31R	.28	.18	-.13	-.08	-.19	.27	-.02	-.08	-.55	-.56	0,17	-.20	.22	-.03	-.12	.29	.18	-.05	-.08	-.59	.50	0,27*
E26	-.15	-.16	.19	.14	.29	-.24	.29	.28	.48	.51	0,20*	.18	-.16	.11	.18	-.28	-.26	.26	.32	.54	-.42	0,13
E11	-.10	-.14	.36	.27	-.14	.12	-.03	.01	.39	.50	0,02	.11	-.18	.28	.36	.11	.13	.00	-.04	.51	-.38	0,13

Note. N = névrosisme, A = agréabilité, C = conscienciosité, O = ouverture, E = extraversion. H = hommes, F = femmes, E = étudiants, NE = non-étudiants. Valeurs supérieures à .30 en valeur absolue en caractère gras. * valeurs DIF classées modérées, ** valeurs DIF classées fortes. Pour les comparaisons de genre, lorsque la valeur DIF est positive cela signifie que l'item a occasionné des scores plus élevés chez les hommes, lorsqu'elle est négative cela signifie que les scores ont été plus élevés chez les femmes indépendamment du trait de personnalité mesuré. Pour les comparaisons entre groupe d'activité, les valeurs positives signifient que les scores ont été plus élevés pour les non-étudiants (personnes en activité professionnelle), les valeurs négatives signifient des scores plus élevés pour les étudiants.

Tableau 2. Statistiques descriptives (moyennes et écart-types) pour chaque grand facteur, pour chaque groupe et en fonction de la version de l'inventaire (complet ou sans les items avec fonctionnements différentiels par genre ou activité)

	Inventaire standard		Suppression items avec DIF entre groupes de genres		Suppression items avec DIF entre groupes d'activité		
	<i>M</i>	<i>ET</i>	<i>M</i>	<i>ET</i>	<i>M</i>	<i>ET</i>	
E	Échantillon total	47,09	11,19	46,92	10,94	47,20	11,17
	Hommes	47,26	10,98	47,24	10,90	47,56	11,03
	Femmes	47,00	11,30	46,74	10,97	47,01	11,25
	Étudiants	46,61*	11,41	46,25*	11,18	46,55*	11,44
	Non-étudiants	47,85*	10,79	47,96*	10,48	48,22*	10,67
A	Échantillon total	56,39	8,24	55,38	7,89	55,89	8,11
	Hommes	55,47*	8,46	54,50*	8,11	54,99*	8,33
	Femmes	56,89*	8,07	55,86*	7,73	56,39*	7,95
	Étudiants	56,17	8,36	55,17	8,02	55,68	8,24
	Non-étudiants	56,73	8,03	55,71	7,69	56,23	7,91
C	Échantillon total	50,69	10,04	49,72	9,75	50,45	10,04
	Hommes	50,65	10,07	49,44*	9,83	50,57	9,99
	Femmes	50,72	10,02	49,88*	9,70	50,39	10,07
	Étudiants	47,73*	9,89	46,94*	9,63	47,47*	9,86
	Non-étudiants	55,29*	8,40	54,05*	8,23	55,08*	8,44
N	Échantillon total	45,86	11,56	47,95	12,10	46,07	11,98
	Hommes	42,37*	11,70	43,82*	12,20	42,24*	12,19
	Femmes	47,76*	11,03	50,20*	11,43	48,16*	11,33
	Étudiants	47,38*	11,62	49,41	12,22	47,50	12,07
	Non-étudiants	43,50*	11,05	45,68	11,56	43,86	11,49
O	Échantillon total	49,96	8,98	50,03	9,32	50,38	8,70
	Hommes	50,09	8,67	50,60	8,93	50,62	8,46
	Femmes	49,89	9,15	49,72	9,52	50,24	8,82
	Étudiants	50,08	9,41	50,15	9,72	50,35	9,10
	Non-étudiants	49,77	8,28	49,84	8,66	50,41	8,03
5F	Échantillon total	50	10	50	10	50	10

Note. N = névrosisme, A = agréabilité, C = conscienciosité, O = ouverture, E = extraversion. Les moyennes et écart-types ont été transformés en scores T sur la base des moyennes estimées après contrôle de l'effet du genre, de l'âge et de l'activité des participants lors d'une analyse de covariances [14]. La valeur de 50 correspond à la moyenne et 10 à l'écart-type de l'échantillon total sur les cinq grands facteurs confondus. Selon Cohen [30] une différence de 2 points représente une taille d'effet faible, une différence de 5 points représente une taille d'effet moyenne et une différence de 8 points représente une taille d'effet importante. Les astérisques (*) correspondent à des différences significatives en ANCOVA ($p < .01$).

Tableau 3. Corrélations partielles entre les cinq facteurs et alphas de Cronbach (en diagonales) au niveau de l'échantillon total et par sous-échantillons de genre et d'activité.

		E	A	C	N	O
Échantillon total (<i>N</i> = 1896)	E	.82				
	A	-.05	.70			
	C	.13*	.10*	.82		
	N	-.29*	-.20*	-.14*	.81	
	O	.18*	.12*	.08*	-.12*	.73
Par genre (hommes diagonale haute, <i>n</i> = 667 ; femmes diagonale basse, <i>n</i> = 1226)	E	.83/.80	-.03	.12*	-.32*	.15*
	A	-.07	.71/.69	.08	-.20*	.11*
	C	.13*	.11*	.83/.81	-.09	.07
	N	-.27*	-.21*	-.17*	.79/.81	-.18*
	O	.19*	.12*	.08*	-.09*	.75/.69
Par activité (étudiants diagonale haute, <i>n</i> = 1152 ; non-étudiants diagonale basse, <i>n</i> = 741)	E	.80/.83	-.06	.07	-.24*	.17*
	A	-.04	.71/.70	.07	-.20*	.09*
	C	.25*	.07	.76/.81	-.09*	.05
	N	-.37*	-.19*	-.08*	.80/.80	-.08*
	O	.20*	.17*	.13	-.21**	.69/.75

Note. * $p < .01$. N = névrosisme, A = agréabilité, C = conscienciosité, O = ouverture, E = extraversion. Pour l'échantillon total, les effets du genre, de l'âge et de l'activité ont été contrôlés. Pour la comparaison des groupes de genre les effets de l'âge et de l'activité ont été contrôlés, alors que la comparaison des groupes d'activités différentes le genre et l'âge ont été contrôlés. Les écarts d'effectifs par rapport au nombre de participants sont dus à des valeurs manquantes au niveau de l'âge. Les corrélations en gras sont significativement différentes selon le test de transformation en valeurs *z* de Fisher.