Mesure et évaluation en éducation

Une note sur le coefficient oméga (ω) et ses déclinaisons pour estimer la fidélité des scores

Sébastien Béland and Florent Michelot

Volume 43, Number 3, 2020

URI: https://id.erudit.org/iderudit/1084526ar
DOI: https://doi.org/10.7202/1084526ar

See table of contents

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)
2368-2000 (digital)

Explore this journal

Cite this article

Béland, S. & Michelot, F. (2020). Une note sur le coefficient oméga (ω) et ses déclinaisons pour estimer la fidélité des scores. *Mesure et évaluation en éducation*, *43*(3), 103–122. https://doi.org/10.7202/1084526ar

Article abstract

In recent decades, some authors have proposed rejecting Cronbach's alpha (α) coefficient (1951) to adopt McDonald's omega (ω) coefficient (1985, 1999) which is based on a factor analysis model. After presenting some inherent limits of the α, this article follows two aims : (i) understanding the theoretical logic behind omega coefficients and (ii) presenting ω coefficients' specificities regarding their computation method. To this end, we identify conditions for the use of the different forms of ω (total or hierarchical, with an EFA or CFA). An example of analysis and recommendations are suggested to better argue the score reliability.

Une note sur le coefficient oméga (ω) et ses déclinaisons pour estimer la fidélité des scores

Sébastien Béland
Université de Montréal

Florent Michelot
Université de Moncton

Mots clés : oméga, fidélité des scores, modèle d'analyse factorielle, adéquation des données

Au cours des dernières décennies, certains auteurs ont suggéré de rejeter le coefficient alpha (α) de Cronbach (1951) pour adopter le coefficient oméga (ω) de McDonald (1985, 1999) basé sur un modèle d'analyse factorielle. Après avoir présenté certaines limites inhérentes à l'α, nous présentons le coefficient ω et ses déclinaisons en poursuivant deux objectifs : comprendre la logique théorique derrière les coefficients de fidélité oméga et exposer les spécificités des coefficients ω sur le plan de leur méthode de calcul. À cet effet, nous distinguons les conditions d'usage des différentes formes d'ω (total ou hiérarchique, dans le cadre d'une AFE ou d'une AFC). Un exemple d'analyse et des recommandations sont proposés pour mieux argumenter la fidélité des scores.

Key words: omega, reliability of scores, factor analysis model, model fit

In recent decades, some authors have proposed rejecting Cronbach's alpha (α) coefficient (1951) to adopt McDonald's omega (ω) coefficient (1985, 1999) which is based on a factor analysis model. After presenting some inherent limits of the α, this article follows two aims : (i) understanding the theoretical logic behind omega coefficients and (ii) presenting ω coefficients' specificities regarding their computation method. To this end, we identify conditions for the use of the different forms of ω (total or hierarchical, with an EFA or CFA). An example of analysis and recommendations are suggested to better argue the score reliability.

Palavras-chave: ómega, fidelidade das pontuações, modelo de análise fatorial, adequação dos dados

Nas últimas décadas, alguns autores sugeriram a rejeição do coeficiente alfa (α) de Cronbach (1951) para se adotar o coeficiente ómega (ω) de McDonald (1985, 1999) baseados num modelo de análise fatorial. Após a apresentação de certas limitações inerentes ao α, apresentamos o coeficiente ω e as suas variações, prosseguindo dois objetivos: compreender a lógica teórica subjacente aos coeficientes de fidelidade ómega e expor as especificidades dos coeficientes ω em termos do seu método de cálculo. Para este fim, distinguimos as condições de uso das diferentes formas de ω (total ou hierárquico, no quadro de um AFE ou AFC). Um exemplo das análises e das recomendações são propostas para melhor argumentar a fidelidade das pontuações.

Note des auteurs : La correspondance liée à cet article peut être adressée à sebastien.beland@umontreal.ca et à florent.michelot@umoncton.ca. Nous souhaitons remercier les évaluateurs anonymes de ce manuscrit pour la qualité de leurs commentaires.

Introduction

Bien qu'il soit toujours pertinent pour analyser des items qui mesurent approximativement un facteur (Raykov et Marcoulides, 2019 ; Sijtsma et Pfadt, 2021), le coefficient alpha (noté α dans ce qui suit) de Cronbach (1951) est critiqué depuis plusieurs décennies par de nombreux auteurs. Par exemple, pour Sijtsma (2009), « la seule et unique raison de rapporter l'α est que les meilleures revues ont tendance à accepter les articles qui utilisent des méthodes statistiques qui existent depuis longtemps, telles que l'α » (p. 118, trad. libre), alors que McNeish (2018) considère que recourir à « toutes les options sont préférables au fait de continuer à utiliser l'α » (p. 423, trad. libre). Cela explique probablement pourquoi Bourque et al. (2019) titraient leur article *L'alpha de Cronbach est l'un des pires estimateurs de la consistance interne*.

Parmi les options proposées, Dunn et al. (2013) mentionnent l'oméga (noté ω dans ce qui suit) de McDonald (1985, 1999)[1]. Selon Maydeu-Olivares et al. (2007) :

> Si un modèle avec un bon ajustement peut être trouvé, l'utilisation d'une estimation de la fidélité basée sur ledit modèle est sans conteste la meilleure option. Par exemple, s'il s'avère qu'un modèle à un facteur s'adapte bien aux données, alors la fidélité des scores du test est donnée par le coefficient ω et le chercheur consciencieux devra employer ce coefficient. (p. 172, trad. libre)

En effet, Trizano-Hermosilla et Alvarado (2016) ont utilisé la simulation informatique auprès de trois échantillons ($n = 250$, 500 et 1000) pour montrer qu'ω serait plus efficace qu'α pour estimer la fidélité des scores de tests comportant 6 et 12 items. Un résultat similaire a été observé par Bourque et al. (2019). Cho (2021) croit toutefois qu'il faut nuancer de tels résultats, car « il existe peu de preuves empiriques que la fidélité de l'AF [analyse factorielle] est plus précise que la fidélité qui ne repose pas sur un modèle d'AF » (p. 4, trad. libre). En effet, peu d'études ont considéré l'erreur d'échantillonnage dans leurs comparaisons de l'ω avec d'autres coefficients. Et on en sait très peu sur le comportement de cet indice avec des données dichotomiques et polytomiques. En fait, « la majorité des travaux sur cette question ont porté sur la fidélité d'instruments à multiples items comportant des [items] continus » (Raykov et al., 2010, p. 265, trad. libre).

Malheureusement, les coefficients de fidélité basés sur un modèle d'analyse factorielle comme les coefficients ω sont souvent utilisés de façon mécanique, alors qu'ils nécessiteraient certaines précautions pour s'assurer qu'ils évaluent bien la fidélité des scores découlant d'une seule passation d'un test.

Cet article poursuit deux objectifs : comprendre la logique théorique derrière les coefficients de fidélité ω basés sur un modèle d'analyse factorielle et exposer les spécificités des coefficients ω sur le plan de leur méthode de calcul.

La fidélité des scores basée sur la théorie classique des tests

Comparativement au concept de validité, qui a été amplement discuté au fil des années (Newton et Shaw, 2014), la conception générale de la fidélité n'a guère évolué. Par exemple, Kuder et Richardson (1937) définissaient le coefficient de fidélité comme une estimation du pourcentage de la variance totale qui peut être expliquée par la variance vraie et, par conséquent, qui n'est pas causée par l'erreur de mesure. Cette définition est encore utilisée aujourd'hui.

Pour comprendre les fondements théoriques des différents coefficients ω, il est nécessaire de retourner à l'équation fondamentale de la théorie classique des tests (TCT), qui est $Y = V + E$, où Y est le score observé d'un individu, V le score vrai et E l'erreur de mesure aléatoire. Pour passer de la théorie à la pratique, cette équation constitue un défi de taille puisqu'on ne peut pas avoir accès à la valeur réelle de V.

En présence des réponses de plusieurs individus, donc de plusieurs scores vrais, la variance du score observé Y se décompose de la façon suivante :

$$\sigma_Y^2 = \sigma_V^2 + \sigma_E^2 \tag{1}$$

Puisque la fidélité des scores concerne la capacité d'un outil de mesure à générer peu d'erreurs, c'est théoriquement la corrélation (notée ρ) entre Y et V qui nous intéresse, soit :

$$\rho(Y,V) = \frac{\sigma_{Y,V}}{\sigma_Y \sigma_V} = \frac{\sigma_V^2}{\sigma_Y \sigma_V} \tag{2}$$

Il est important de savoir que la fidélité est traditionnellement représentée par le carré de ρ, soit ρ^2. Cette forme est similaire au coefficient de détermination, qui quantifie, dans ce cas, la force de la relation linéaire entre *V* et *Y*. Ainsi, la fidélité des scores devient :

$$\rho(Y,V)^2 = \rho^2_{TCT} = \left(\frac{\sigma_{Y,V}}{\sigma_Y\sigma_V}\right)^2 = \left(\frac{\sigma^2_V}{\sigma_Y\sigma_V}\right)^2 = \left(\frac{\sigma_V}{\sigma_Y}\right)^2 = \frac{\sigma^2_V}{\sigma^2_Y} = \frac{\sigma^2_V}{\sigma^2_V + \sigma^2_E} = 1 - \frac{\sigma^2_E}{\sigma^2_Y} \quad (3)$$

et le populaire coefficient de fidélité des scores α tente d'estimer ρ^2_{TCT}, car, rappelons-le, il n'est pas possible d'avoir directement accès à *V*. Ainsi, la valeur de ρ^2_{TCT} sera nécessairement bornée entre 0 et 1, et plus la valeur approchera de 1, plus la fidélité des scores sera élevée. Mentionnons aussi que l'erreur type de mesure (*standard error of measurement*) s'obtient après quelques manipulations algébriques :

$$\sigma_E = \sigma_Y\sqrt{1 - \rho^2_{TCT}} \quad (4)$$

Elle est dépendante de la variance totale au test et de la valeur de ρ^2_{TCT}.

Cette définition de la fidélité des scores a été retenue par un grand nombre d'auteurs (p. ex., Lord et al., 1968), mais d'autres options sont suggérées, dont celle de Sijtsma et Pfadt (2021), qui définissent la fidélité comme étant la corrélation entre deux tests parallèles. En revanche, Cho (2021) conteste cette perspective, car les tests parallèles sont plutôt rares dans la réalité.

Le coefficient α pour estimer ρ^2_{TCT}

Le coefficient α, présenté par Cronbach (1951), existait aussi sous des formes équivalentes chez Kuder et Richardson (1937) et chez Guttman (1945). Mathématiquement, ce coefficient est une estimation de ρ^2_{TCT} et s'écrit :

$$\alpha = \frac{J}{J-1}\left(1 - \frac{\sum_{j=1}^{J}\sigma^2_{item_j}}{\sigma^2_{Total}}\right) \quad (5)$$

où $\sum_{j=1}^{J}\sigma^2_{item_j}$ est la somme des variances à chacun des *J* items et estime l'élément σ^2_E de ρ^2_{TCT}. L'élément σ^2_{Total} est la somme de tous les éléments contenus dans une matrice de variance-covariance/corrélations.

À l'instar des coefficients présentés par Guttman (1945) et du *greatest lower bound* (*glb*) présenté par Jackson et Agunwamba (1977), le coefficient repose sur l'hypothèse de tau-équivalence de la variance, ce qui veut dire que les items doivent être fortement corrélés et ne doivent différer les uns des autres que par une constante. Les scores tau-équivalents supposent que σ^2_V a des valeurs similaires pour tous les items, ce qui est peu réaliste dans la recherche en sciences sociales, singulièrement en éducation. Considérer que σ_V^2 peut différer d'un item à l'autre, comme cela est proposé dans la conception congénérique de la fidélité des scores, est beaucoup plus réaliste. On estime donc que c'est cette conception qu'on devrait davantage retrouver en sciences de l'éducation.

Le Tableau 1 qui suit, adapté de Cho (2016), montre deux hypothèses[2] quant à la variance telle qu'elle est présentée à l'équation 1 : dans l'exemple présentant une hypothèse de tau-équivalence, on constate que la variance vraie σ_V^2 est identique, quel que soit l'item (en l'occurrence $\sigma_V^2 = 5$), tandis qu'elle varie dans le cas de l'hypothèse congénérique (où σ_V^2 varie, selon l'item, de 4 à 16).

Tableau 1
Deux hypothèses sur la variance des items

		σ_Y^2 **Variance des scores observés au test**			σ_V^2 **Variance vraie**			σ_E^2 **Variance d'erreur**		
		Y1	**Y2**	**Y3**	**Y1**	**Y2**	**Y3**	**Y1**	**Y2**	**Y3**
Scores tau-équivalents	Y1	8	5	5	5	5	5	3	0	0
	Y2	5	10	5	5	5	5	0	5	0
	Y3	5	5	12	5	5	5	0	0	7
Scores congénériques	Y1	10	6	8	4	6	8	6	0	0
	Y2	6	16	12	6	9	12	0	7	0
	Y3	8	12	22	8	12	16	0	0	6

Pour Sijtsma et Pfadt (2021), le coefficient α serait utile pour savoir si des scores peuvent répéter une procédure d'administration précise, par exemple si l'on a des doutes sur la qualité globale du test ou de certains items. Cependant, certains auteurs déclarent qu'en présence de variances d'items congénériques, α peut être biaisé (Raykov, 1997, 2001). De leur côté, Revelle et Condon (2019) déclarent qu'α n'est pas plus une estimation de ρ_{TCT}^2 qu'une mesure de la fidélité des scores : il s'agit simplement d'un coefficient dont la valeur varie en fonction du nombre d'items et des corrélations moyennes entre ces items. En fait, α constituerait un indice adéquat si tant est que trois grandes conditions soient remplies : (i) lorsque le test comporte une seule dimension ; (ii) si l'on admet une perspective de fidélité des scores tau-équivalente ; et (iii) lorsque les erreurs de mesure ne sont pas corrélées entre elles (Raykov et Marcoulides, 2011).

Considérant les limites de l'α, il convient maintenant d'explorer davantage le coefficient qui est au cœur de cet article, à savoir l'ω basé sur un facteur général avec la possibilité d'identifier des facteurs secondaires (aussi appelés « facteurs de groupe »).

La fidélité des scores basée sur un modèle d'analyse factorielle

Les modèles d'analyse factorielle exploratoire (AFE) et confirmatoire (AFC) ont comme objectif de produire ou d'évaluer des groupes d'items nommés facteurs. Un exemple connu est le modèle exploratoire de Spearman, qui s'écrit $Y = \lambda_j\theta + u_j^2$, où λ_j est un coefficient de saturation factorielle (*factor loading*) de l'item j, où θ est un trait latent et où u_j^2 est une erreur de mesure aléatoire.

Selon Lord et al. (1968) et McDonald (1999), l'équation $Y = V + E$ peut être réécrite de la façon suivante dans le cadre du modèle d'analyse factorielle de Spearman : $Y = C + U$, où C est la part commune du modèle, et où U est la part unique qui n'est pas expliquée et à laquelle s'ajoute l'erreur de mesure aléatoire. Notez que les facteurs de groupe sont identifiés à partir de U.

Estimer la fidélité d'un ensemble d'items nécessite de pouvoir isoler les constituants de la variance émergeant d'un modèle d'analyse factorielle :

$$\sigma_Y^2 = \sigma_C^2 + \sigma_U^2 \tag{6}$$

où σ_C^2 est la variance commune du modèle tirée de λ_j et où σ_U^2 est la variance unique du modèle. Si l'on s'inspire de la logique de l'équation 3, la fidélité basée sur l'analyse factorielle est théoriquement égale à :

$$\rho_{AF}^2 = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_U^2} \tag{7}$$

où McDonald (1999) supposait que $\sigma_V^2 \approx \sigma_C^2$ et que $\sigma_E^2 \approx \sigma_U^2$. Cependant, cette approche ne repose pas exactement sur les mêmes éléments que ceux contenus dans la théorie classique des tests, ce qui explique pourquoi nous utilisons les notations ρ_{AF}^2 et ρ_{TCT}^2 .

Rappelons que la conception de la fidélité selon la théorie classique des tests diffère de celle basée sur un modèle d'analyse factorielle. En effet, la première est basée sur le rapport entre la variance vraie et la variance totale (ou la capacité du score au test à être répété selon une autre perspective), tandis que la seconde repose sur la proportion des scores que la variance du modèle d'analyse factorielle peut expliquer (Sijtsma et Pfadt, 2021). Il y a donc une différence conceptuelle entre ces deux perspectives.

Les coefficients ω pour estimer ρ^2_{AF}

Dans leur forme la plus simple, les modèles d'analyse factorielle de Spearman et de Thurstone sont constitués de deux grandes catégories de paramètres qui peuvent être soit contraints à des valeurs précises, soit laissés libres de varier : (i) le coefficient de saturation de l'item, qui représente la force de la relation entre un item et un facteur ; et (ii) les paramètres d'erreur de ces modèles. Selon Bonett (2021), le coefficient ω prend la forme générique suivante en présence d'un modèle de mesure avec un facteur général tau-équivalent :

$$\omega = \frac{(J\lambda_j)^2}{(J\lambda_j)^2 + \sum_{j=1}^{J} u_j^2} \tag{8}$$

où λ_j est un coefficient de saturation et u_j^2 est la part inexpliquée du modèle auquel s'ajoute l'erreur de mesure. Ce qui est plus important est la forme suivante du coefficient dans le cadre d'un modèle de mesure avec un facteur général congénérique :

$$\omega = \frac{(\sum_{j=1}^{J} \lambda_j)^2}{(\sum_{j=1}^{J} \lambda_j)^2 + \sum_{j=1}^{J} u_j^2} \tag{9}$$

où l'élément $(\sum_{j=1}^{J} \lambda_j)^2$ est utilisé pour estimer σ_C^2 et où l'élément $\sum_{j=1}^{J} u_j^2$ est utilisé pour estimer σ_U^2.

À ce stade, il est nécessaire de rappeler qu'il existe une confusion terminologique autour du coefficient ω (Cho, 2021). D'ailleurs, McDonald, en 1985, n'effectuait pas de distinction formelle entre les différentes versions d'ω. En réalité, il existe deux grandes déclinaisons de ce coefficient : ω_{Total} et $\omega_{Hiérarchique}$, rapportés dans le Tableau 2. Nous avons fait le choix de présenter la façon dont ils sont calculés dans le logiciel R puisque sa librairie *psych* (Revelle, 2019) fait office de référence en mesure.

Une première chose à observer est que les différentes déclinaisons d'ω peuvent être calculées à partir de différents modèles d'AFE et d'AFC. Brunner et al. (2012) présentent d'ailleurs des exemples de calculs du coefficient ω à partir de différents modèles d'AFC.

Un deuxième élément à constater est que les différentes versions d'ω ne donnent pas nécessairement des valeurs similaires. Pour illustrer cette idée importante, nous utilisons les 12 items de compréhension de phrases à l'écoute du Test de classement en anglais, langue seconde (TCALS-II) obtenus au Cégep de l'Outaouais en 1998 (Raîche, 2002 ; voir la section *Un exemple* ci-dessous pour plus d'informations). Les corrélations entre ces 12 items sont présentées dans la section de gauche du Tableau 3, tandis que les coefficients de saturation d'une AFE hiérarchique sont affichés à droite.

Tableau 2

Distinction des coefficients ω utilisés dans la librairie psych

Modèle	Nom	Équation	Descriptif	Dans le logiciel R
AFE	$\omega_{T:AFE}$	$1 - \frac{\sum(1-h_j^2)}{\sigma^2_{Total}}$	Basé sur la somme des variances uniques d'une AFE hiérarchique après rotation oblique, ce qui permet de trouver la solution de la transformation Schmid-Leiman sur le facteur général. Les facteurs secondaires sont définis à partir de l'erreur résiduelle générée par le modèle. Ici, l'élément h_j^2 est la communauté de l'item et sa variance unique est $u_j^2 = 1 - h_j^2$.	Fonction *omega* dans la librairie *psych*
	$\omega_{H:AFE}$	$\frac{(\sum c_j)^2}{\sigma^2_{Total}}$	Basé sur le facteur général d'une AFE hiérarchique après rotation oblique, ce qui permet de trouver la solution de la transformation Schmid-Leiman sur le facteur général. Les facteurs secondaires sont définis à partir de l'erreur résiduelle générée par le modèle. Ici, l'élément c_j est le coefficient de saturation de l'item sur le facteur général.	Fonction *omega* dans la librairie *psych*
AFC	$\omega_{T:AFC}$	$1 - \frac{\sum(1-h_j^2)}{\sigma^2_{Total}}$	Basé sur la somme des variances uniques du modèle bifactoriel d'une AFC où tous les items sont estimés sur un facteur général. Les facteurs secondaires sont définis à partir de l'erreur résiduelle d'une AFE. Ici, l'élément h_j^2 est la communauté de l'item et sa variance unique est $u_j^2 = 1 - h_j^2$.	Fonctions *omegaSem* et *omegaFromSem* dans la librairie *psych*
	$\omega_{H:AFC}$	$\frac{(\sum c_j)^2}{\sigma^2_{Total}}$	Basé sur le facteur général d'un modèle bifactoriel d'une AFC où tous les items sont estimés sur un facteur général. Les facteurs secondaires sont définis à partir de l'erreur résiduelle d'une AFE. Ici, l'élément c_j est le coefficient de saturation de l'item sur le facteur général.	Fonctions *omegaSem* et *omegaFromSem* dans la librairie *psych*

Notes. L'élément σ^2_{Total} est estimé soit à l'aide de la variance du modèle d'analyse factorielle utilisé, soit en calculant la variance/corrélation dans l'échantillon observé (tout comme pour le coefficient α).

Les coefficients $\omega_{T:AFE}$ et $\omega_{H:AFE}$ sont calculés à partir d'une analyse factorielle hiérarchique (voir la recommandation de Jensen et Weng [1994] concernant l'extraction d'un facteur général).

Cet exemple concerne la librairie *psych* du logiciel R, mais le coefficient ω se retrouve dans plusieurs autres logiciels. Par exemple, Hayes et Coutts (2020) ont produit une macro pour SPSS et SAS. Les logiciels JASP, jamovi et Mplus permettent aussi de calculer l'ω.

Tableau 3

Corrélations entre items et coefficients de saturation d'un modèle exploratoire hiérarchique, après rotation oblique, permettant de trouver la solution Schmid-Leiman pour le facteur général

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	c_j	F1	F2	u^2
Y_1	1	0,26	0,31	0,31	0,21	0,15	0,25	0,29	0,27	0,22	0,17	0,2	0,43	-0,01	0,37	0,68
Y_2	0,26	1	0,22	0,17	0,15	0,16	0,15	0,18	0,19	0,13	0,17	0,14	0,30	0,00	0,25	0,85
Y_3	0,31	0,22	1	0,25	0,25	0,15	0,26	0,22	0,25	0,2	0,15	0,18	0,40	-0,02	0,35	0,72
Y_4	0,31	0,17	0,25	1	0,27	0,26	0,31	0,38	0,36	0,29	0,23	0,27	0,48	0,15	0,25	0,68
Y_5	0,21	0,15	0,25	0,27	1	0,21	0,26	0,28	0,25	0,21	0,23	0,17	0,38	0,09	0,23	0,79
Y_6	0,15	0,16	0,15	0,26	0,21	1	0,24	0,26	0,26	0,23	0,21	0,25	0,35	0,17	0,12	0,83
Y_7	0,25	0,15	0,26	0,31	0,26	0,24	1	0,38	0,36	0,32	0,29	0,31	0,47	0,22	0,17	0,70
Y_8	0,29	0,18	0,22	0,38	0,28	0,26	0,38	1	0,5	0,49	0,39	0,43	0,57	0,40	0,08	0,50
Y_9	0,27	0,19	0,25	0,36	0,25	0,26	0,36	0,5	1	0,48	0,38	0,42	0,56	0,39	0,08	0,53
Y_{10}	0,22	0,13	0,2	0,29	0,21	0,23	0,32	0,49	0,48	1	0,38	0,48	0,51	0,48	-0,05	0,50
Y_{11}	0,17	0,17	0,15	0,23	0,23	0,21	0,29	0,39	0,38	0,38	1	0,37	0,43	0,35	0,00	0,69
Y_{12}	0,2	0,14	0,18	0,27	0,17	0,25	0,31	0,43	0,42	0,48	0,37	1	0,47	0,44	-0,04	0,58

Note. Puisqu'il y a deux facteurs secondaires (F1 et F2) dans cette analyse, les coefficients de saturation du facteur général sont égaux.

Dans les deux cas, il faut analyser les données à l'aide d'une AFE hiérarchique, appliquer une rotation oblique des axes factoriels et, enfin, trouver la solution de la transformation Schmid-Leiman sur le facteur général. Pour obtenir une estimation de σ^2_{Total}, on doit effectuer la somme de tous les éléments contenus dans la matrice de corrélations du Tableau 3, ce qui donne la valeur de 47,24. Ensuite, pour calculer $\omega_{T:AFE}$, il est nécessaire de prendre la somme des variances uniques u^2, donc :

$$0{,}68 + 0{,}85 + 0{,}72 + 0{,}68 + \ldots + 0{,}53 + 0{,}50 + 0{,}69 + 0{,}58 = 8{,}05.$$

Tous les ingrédients sont en main pour calculer la valeur du coefficient :

$$\omega_{T:AFE} = 1 - (8{,}05 / 47{,}24) = 0{,}83$$

Le coefficient $\omega_{T:AFE}$ nécessite plutôt de mettre au carré la somme des saturations de ce facteur général, que nous notons $(\Sigma c_j)^2$:

$$0{,}43 + 0{,}30 + 0{,}40 + 0{,}48 + \ldots + 0{,}56 + 0{,}51 + 0{,}43 + 0{,}47 = 5{,}35$$

et de diviser cette valeur par σ^2_{Total}.

Dans ce cas-ci, $\omega_{H:AFE} = 5{,}35^2 / 47{,}24 = 0{,}61$.

En AFC, la démarche est comparable, mais les coefficients de saturation ne sont pas estimés de la même façon. Reprenons la même matrice de corrélations, mais en estimant les paramètres à l'aide d'une AFC bifactorielle, où les facteurs secondaires sont définis en AFE.

D'abord, le coefficient $\omega_{T:AFC}$ consiste à calculer la somme des variances uniques (soit 0,63 + 0,82 + 0,73 + 0,64 + … + 0,52 + 0,47 + 0,70 + 0,58 = 7,87) et à reprendre la valeur $\sigma^2_{Total} = 47{,}24$, car la matrice de corrélations utilisée est la même que dans l'exemple précédent. Ainsi, $\omega_{T:AFC} = 1 - (7{,}87 / 47{,}24) = 0{,}83$.

De son côté, le coefficient $\omega_{H:AFC}$ utilise l'information qui se trouve dans la colonne c_j du Tableau 4, soit $(\Sigma c_j)^2 = 5{,}71^2 = 32{,}60$, et se calcule de la façon suivante : $\omega_{H:AFC} = 32{,}60 / 47{,}24 = 0{,}69$. À titre indicatif, nous avons obtenu un $\alpha = 0{,}81$ pour ces données.

L'existence de différences entre les AFE et AFC n'est pas surprenante, car les coefficients de saturation ne sont pas estimés de la même façon. Il convient donc de sélectionner le modèle d'analyse factorielle le plus adéquat au moment d'estimer la fidélité des scores avec cette catégorie de coefficients. Ainsi, calculer les coefficients ω à partir d'une AFC implique d'avoir des hypothèses fortes entre les items à l'étude et le facteur général

Tableau 4

Corrélations entre items et coefficients de saturation d'un modèle confirmatoire bifactoriel

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	c_j	F1	F2	u^2
Y_1	1	0,26	0,31	0,31	0,21	0,15	0,25	0,29	0,27	0,22	0,17	0,2	0,45	0,00	0,41	0,63
Y_2	0,26	1	0,22	0,17	0,15	0,16	0,15	0,18	0,19	0,13	0,17	0,14	0,30	0,00	0,31	0,82
Y_3	0,31	0,22	1	0,25	0,25	0,15	0,26	0,22	0,25	0,2	0,15	0,18	0,41	0,00	0,32	0,73
Y_4	0,31	0,17	0,25	1	0,27	0,26	0,31	0,38	0,36	0,29	0,23	0,27	0,60	0,00	0,04	0,64
Y_5	0,21	0,15	0,25	0,27	1	0,21	0,26	0,28	0,25	0,21	0,23	0,17	0,46	0,00	0,05	0,78
Y_6	0,15	0,16	0,15	0,26	0,21	1	0,24	0,26	0,26	0,23	0,21	0,25	0,41	0,08	0,00	0,82
Y_7	0,25	0,15	0,26	0,31	0,26	0,24	1	0,38	0,36	0,32	0,29	0,31	0,55	0,13	0,00	0,68
Y_8	0,29	0,18	0,22	0,38	0,28	0,26	0,38	1	0,5	0,49	0,39	0,43	0,61	0,35	0,00	0,50
Y_9	0,27	0,19	0,25	0,36	0,25	0,26	0,36	0,5	1	0,48	0,38	0,42	0,59	0,35	0,00	0,52
Y_{10}	0,22	0,13	0,2	0,29	0,21	0,23	0,32	0,49	0,48	1	0,38	0,48	0,47	0,56	0,00	0,47
Y_{11}	0,17	0,17	0,15	0,23	0,23	0,21	0,29	0,39	0,38	0,38	1	0,37	0,42	0,35	0,00	0,70
Y_{12}	0,2	0,14	0,18	0,27	0,17	0,25	0,31	0,43	0,42	0,48	0,37	1	0,44	0,48	0,00	0,58

Note. Puisqu'il y a deux facteurs secondaires (F1 et F2) dans cette analyse, les coefficients de saturation du facteur général sont égaux.

(Flora, 2020). Du côté de l'AFE, l'utiliser est intéressant pour mieux comprendre la relation existante entre les items, le facteur général et les facteurs secondaires.

Les coefficients de fidélité des scores peuvent être accompagnés de leur intervalle de confiance pour attester de leur précision. Rappelons que, dans une perspective fréquentiste, un intervalle de confiance à 95 % indique que, si l'étude est répétée de façon similaire un très grand nombre de fois, 95 % des intervalles vont inclure la vraie valeur de la fidélité des scores. Fait intéressant, Kelley et Pornprasertmanit (2016) montrent que ce sont les intervalles de confiance obtenus par *bootstrap* qui sont recommandés.

Enfin, Deng et Chan (2017) reviennent sur le fait que plusieurs chercheurs ont discuté de la différence négligeable qui existerait entre ω et α. Ils ont alors élaboré une stratégie pour tester la différence entre les coefficients ω et α. Leurs analyses montrent qu'il y a une différence statistiquement significative dans 13 des 16 échelles évaluées. Toujours selon Deng et Chan (2017), les deux coefficients ne peuvent donc pas être utilisés de façon interchangeable, comme certains le laisseraient entendre.

Quelques mots sur la relation entre l'adéquation des données à un modèle et la fidélité

Pour rappeler les sages paroles de Westfall et Henning (2013), « un modèle produit des données ; le modèle a des paramètres inconnus ; les données réduisent l'incertitude quant aux paramètres inconnus » (p. 11, trad. libre). Pour cette raison, il est important de vérifier si les données présentent une bonne adéquation au modèle d'analyse factorielle utilisé. Il importe donc d'utiliser le modèle d'analyse le plus adéquat.

Plusieurs coefficients permettant d'évaluer la qualité d'un modèle sont disponibles. Par exemple, dans le contexte de l'AFC, le test du chi-carré, qui est dépendant de la taille de l'échantillon, évalue l'hypothèse nulle selon laquelle le modèle présente une adéquation parfaite aux données, l'hypothèse alternative étant que ce n'est pas le cas. Différents coefficients peuvent aussi être interprétés sur la base des balises reconnues : pour l'indice de qualité d'ajustement (*goodness of fit index* ou GFI), on souhaitera ainsi obtenir une valeur plus grande que 0,95 ; pour l'indice d'ajustement comparatif (*comparative fit index* ou CFI), une valeur plus grande que 0,90 ; pour l'erreur quadratique moyenne de l'approximation (*root mean square error of approximation* ou RMSEA), une valeur plus petite que 0,08. Le lecteur curieux pourra lire Kline (2016) pour en connaître davantage à ce propos.

Stanley et Edwards (2016) synthétisent la relation complexe entre la valeur d'un coefficient de fidélité des scores et l'adéquation du modèle utilisé de la façon suivante :

Tableau 5
Relation entre la fidélité et l'ajustement des données

	Ajustement acceptable	**Ajustement inacceptable**
Fidélité acceptable	Situation idéale	Problème potentiel de dimensionnalité
Fidélité inacceptable	Erreur de mesure	Il faut choisir un autre coefficient de fidélité

Quand la fidélité estimée et l'ajustement au modèle sont acceptables, il n'y a pas de problème. En revanche, les choses se compliquent quand l'adéquation des données au modèle est bonne, mais que la fidélité estimée est plutôt mauvaise.

Dans une telle situation, la valeur de $\omega_{T:AFE}$ sera moins élevée, car les valeurs de la variance unique $u^2 = 1 - h^2$ sont relativement élevées, comparativement aux valeurs de h^2. Cela veut dire que le modèle d'analyse factorielle génère de l'erreur. Comme l'explique Edwards (2013) :

> Ce n'est pas parce qu'il y a adéquation des données au modèle qu'une échelle est bonne. L'adéquation au modèle nous indique dans quelle mesure un ensemble particulier de paramètres estimés peut reproduire une certaine variante des données observées. Il se peut que les données observées nous disent qu'une échelle n'est pas très bonne (p. ex., tous les éléments ont une faible pente), mais que le modèle peut reproduire ces données avec une grande précision (ce qui conduit à une bonne adéquation). (p. 111, trad. libre)

Lorsque l'adéquation est mauvaise, mais que la fidélité des scores est bonne, la situation peut, là aussi, poser problème. Dans ce cas, le modèle définit probablement mal la dimensionnalité des données et l'analyste doit s'assurer d'utiliser le bon modèle d'analyse factorielle avant de prétendre estimer adéquatement la fidélité des scores. Enfin, quand la fidélité estimée et l'ajustement aux données sont médiocres, il faut sélectionner un autre coefficient.

Argumenter plutôt que simplement mentionner la valeur d'un ω

Comme nous l'avons vu, les coefficients de fidélité basés sur un modèle d'analyse factorielle sont complexes : en cela, se contenter de mentionner la valeur d'un coefficient en guise de preuve de fidélité des scores semble lacunaire.

Ainsi, il n'est pas suffisant de se satisfaire d'une valeur étalon (et arbitraire, par ailleurs) égale ou supérieure à disons 0,7 pour montrer qu'un test présente des scores qualifiables de « fidèles ». Il importe de poursuivre les essais en étudiant la relation entre la valeur d'un coefficient et l'adéquation des données au modèle d'analyse factorielle utilisé.

Nous croyons que les articles publiés devraient, autant que faire se peut (et particulièrement lorsque la publication vise à discuter du développement d'un instrument), donner les informations minimales suivantes sur les coefficients de fidélité basés sur un modèle d'analyse factorielle :

1. la justification du modèle utilisé pour calculer le coefficient (p. ex., en se demandant si on veut explorer ou confirmer une théorie sur la relation entre les items et le facteur général et les facteurs secondaires) ;
2. les informations sur l'adéquation des données au modèle utilisé ;
3. la valeur du coefficient de fidélité et ses intervalles de confiance.

De façon plus générale, cette proposition s'inscrit dans le droit fil de Parkes (2007), qui propose différentes étapes d'argumentation relatives à la fidélité. Bien que celles-ci dépassent le strict cadre de l'article, nous croyons que les informations relatives au contexte d'utilisation de l'instrument, à ce que les auteurs considèrent comme une fidélité acceptable et aux liens avec la validité des scores, sont bienvenues afin d'argumenter sur la fidélité ou non des scores.

Un exemple

Nous allons maintenant analyser la fidélité des scores (n = 1372) du Test de classement en anglais, langue seconde (TCALS-II) obtenus au Cégep de l'Outaouais en 1998 (Raîche, 2002). Ce test, qui comporte 85 questions à quatre choix de réponse dichotomisés en bonnes et mauvaises réponses, a comme objectif de classer les étudiants dans des cours adaptés à leur niveau d'anglais durant leurs études au collégial. Les administrateurs utilisent donc le score total de l'étudiant en guise d'estimation d'habileté en anglais langue seconde. Raîche (2002) a déjà décrit ce test, et son analyse en composante principale montre que celui-ci comporte une dimension. Le lecteur curieux pourra consulter cette référence pour en apprendre plus sur le test.

Puisque l'enjeu est important, le seuil minimal acceptable est arbitrairement fixé à 0,85, plutôt que le seuil souvent utilisé de 0,7. L'objectif, ici, est de montrer que la valeur des coefficients de fidélité de type ω doit être accompagnée d'informations sur l'adéquation des données au modèle utilisé. Nous allons, dans ce qui suit, calculer les coefficients $\omega_{T:AFE}$ et $\omega_{H:AFE}$ en plus de l'α, car il est excessivement connu.

Le coefficient ω et ses déclinaisons sont obtenus à partir d'un modèle d'analyse exploratoire basé sur une matrice de corrélations tétrachoriques. Toutes les analyses sont produites à l'aide de la librairie *psych* du logiciel R (Revelle, 2019). Enfin, les intervalles de confiance à 95% sont calculés à partir de 500 échantillons *bootstrap*.

D'abord, $\sigma^2_{Total} = 1768{,}629$. La valeur des coefficients de fidélité des scores est plutôt encourageante. Ici, $\alpha = 0{,}96$, IC 95% [0,97 ; 0,98], ce qui témoigne d'une très bonne fidélité des scores. Nous obtenons aussi des résultats intéressants, quoiqu'assez différents, selon qu'on calcule ω_T ou ω_H avec deux facteurs secondaires, soit $\omega_{T:AFE} = 0{,}97$, *IC* 95% [0,96 ; 0,98] et $\omega_{H:AFE} = 0{,}75$, *IC* 95% [0,74 ; 0,76]. Les indices d'adéquation des données aux modèles utilisés sont tous acceptables. Par exemple, l'indice de Tucker-Lewis (TLI) est égal à 0,91 et présente une valeur satisfaisante. Même chose pour le RMSEA = 0,038, IC95% [0,037 ; 0,038].

La morale de cette histoire est qu'il ne faut pas se laisser aveugler par la valeur des coefficients $\omega_{T:AFE}$ et $\omega_{H:AFE}$. Il est plutôt indispensable d'argumenter la fidélité des scores. Le TCALS-II présente de bonnes preuves de fidélité des scores pour la passation de 1998 et l'adéquation des données au modèle d'analyse factorielle renforce cette hypothèse.

Conclusion

Il existe différentes conceptions de la fidélité et cet article visait à clarifier l'apport du recours à l'ω dans cette discussion, sachant qu'il peut être basé sur différents modèles d'analyse factorielle à un facteur général. Nous avons donc présenté la logique théorique derrière les déclinaisons de ce coefficient et comment leurs équations se distinguent. Cela a permis de formuler des recommandations dans la section *Argumenter plutôt que simplement mentionner la valeur d'un* ω, puis de conclure sur un exemple d'analyse empirique.

Selon Sijtsma et Pfadt (2021),

> « si vous souhaitez corriger les performances d'un test en modélisant les influences ciblées (*target influences*) et les influences non ciblées (*non-target influences*) que vous estimez indésirables, puis déterminer la fidélité sans les influences non ciblées, vous pouvez utiliser le coefficient ω pour le modèle factoriel qui est en adéquation avec les données recueillies » (s. p., trad. libre).

Rappelons que la bonne utilisation du coefficient ω et de ses déclinaisons nécessite des conditions minimales telles qu'un facteur général, une bonne adéquation des données au modèle d'analyse factorielle utilisé ainsi que des données continues ou une matrice de corrélations adéquates si les données sont discrètes.

Limites

Cet article n'est évidemment pas sans limites. La limite la plus importante est probablement liée au fait que nous n'avons pas discuté de l'application des coefficients ω dans un contexte multidimensionnel selon le modèle de facteur commun.

Une autre limite concerne le fait que nous n'avons pas étudié le comportement des différentes versions de l'ω dans le cadre d'études de simulation. Il y aura un travail à faire à ce propos, tout spécialement dans le cadre de données dichotomiques, comme nous l'avons fait dans cet article, et de données polytomiques.

Enfin, il sera important d'approfondir l'impact de l'utilisation de différents modèles d'analyse factorielle sur l'estimation d' $\omega_{Hiérarchique}$ et d' ω_{Total}. De plus, il faudra travailler à améliorer l'estimation des erreurs contenues dans le modèle hiérarchique et dans le modèle bifactoriel. Et à l'instar de Pfadt et al. (2021), il sera pertinent d'analyser le comportement d'ω et ses déclinaisons en contexte bayésien.

Réception : 3 avril 2020

Version finale : 22 octobre 2021

Acceptation : 25 octobre 2021

NOTES

1. Il est à noter qu'on aussi trouve la logique de l'oméga chez Lord et al. (1968) et chez Greene et Carmines (1980).
2. Notez que l'hypothèse de fidélité des scores parallèle n'est pas discutée dans cet article.

RÉFÉRENCES

Bonett, D. (2021). *Statistics for psychologists : An introduction to multivariate statistical models* (vol. 4) [document en préparation].

Bourque, J., Doucet, D., LeBlanc, J., Dupuis, J. et Nadeau, J. (2019). L'alpha de Cronbach est l'un des pires estimateurs de la consistance interne : une étude de simulation. *Revue des sciences de l'éducation*, *45*(2), 78-99. https://doi.org/g4pm

Brunner, M., Nagy, G. et Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796-846. https://doi.org/fccf97

Cho, E. (2016). Making reliability reliable : A systematic approach to reliability coefficients. *Organizational Research Methods*, *19*(4), 651-682. https://doi.org/f838w6

Cho, E. (2021). Neither Cronbach's alpha nor McDonald's omega : A commentary on Sijtsma and Pfadt. *Psychometrika*. https://doi.org/g4pn

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/cc5

Deng, L. et Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, *77*(2), 185-203. https://doi.org/gfj8nh

Dunn, T. J., Baguley, T. et Brunsden, V. (2013). From alpha to omega : A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399-412. https://doi.org/f2dpm2

Edwards, M. C. (2013). Purple unicorns, true models, and other things I've never seen. *Measurement : Interdisciplinary Research and Perspectives*, *11*(3), 107-111. https://doi.org/g4pp

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484-501. https://doi.org/ghj99z

Greene, V. L. et Carmines, E. G. (1980). Assessing the reliability of linear composites. *Sociological Methodology*, *11*, 160-175. https://doi.org/ckt88n

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282. https://doi.org/b45tjm

Hayes, A. F. et Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But... *Communication Methods and Measures*, *14*(1), 1-24. https://doi.org/ggwd6m

Jackson, E. H. et Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items : Algebraic lower bounds. *Psychometrika*, *42*(4), 567-578. https ://doi.org/d98rr5

Jensen, A. R. et Weng, L.-J. (1994). What is a good g? *Intelligence*, *18*(3), 231-258. https :// doi.org/ddb8b3

Kelley, K. et Pornprasertmanit, P. (2016). Confidence intervals for population reliability coefficients : Evaluation of methods, recommendations, and software for homogeneous composite measures. *Psychological Methods*, *21*(1), 69-92. https ://doi.org/f8fq9j

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4e éd.). Guilford Press.

Kuder, G. F. et Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151-160. https ://doi.org/czx5nr

Lord, F. M., Novick, M. R. et Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Maydeu-Olivares, A., Coffman, D. L. et Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*(2), 157-176. https ://doi.org/dv8gpk

McDonald, R. P. (1985). *Factor analysis and related methods*. Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory : A unified treatment*. Lawrence Erlbaum.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412-433. https ://doi.org/gcsk2k

Newton, P. et Shaw, S. (2014). *Validity in educational and psychological assessment*. SAGE.

Parkes, J. (2007), Reliability as argument. *Educational Measurement : Issues and Practice*, *26*(4), 2-10. https ://doi.org/ccwsdf

Pfadt, J. M., van den Bergh, D., Sijtsma, K., Moshagen, M. et Wagenmakers, E. J. (2021). Bayesian estimation of single-test reliability coefficients. *Multivariate Behavioral Research*, 1-30. https ://doi.org/g4pr

Raîche, G. (2002). *Le dépistage du sous-classement aux tests de classement en anglais, langue seconde, au collégial*. Collège de l'Outaouais. https ://eduq.info/xmlui/bitstream/handle/11515/1012/raiche_PAREA_2001_depistage.pdf

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*(4), 329-353. https ://doi.org/d4mfcg

Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*(1), 69-76. https ://doi.org/fk938x

Raykov, T., Dimitrov, D. M. et Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, *17*(2), 265-279. https ://doi.org/10.1080/10705511003659417

Raykov, T. et Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Raykov, T. et Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, *79*(1), 200-210. https ://doi.org/gf4j8w

Revelle, W. (2019) *psych : Procedures for psychological, psychometric and personality research – R Package Version 1.9.12.* Northwestern University, Evanston, IL. https ://CRAN.R-project.org/package=psych

Revelle, W. et Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, *31*(12), 1395-1411. https://doi.org/gf527p

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120. https://doi.org/c76r6c

Sijtsma, K. et Pfadt, J. M. (2021). Invited review part II : On the use, the misuse, and the very limited usefulness of Cronbach's alpha : Discussing lower bounds and correlated errors. *Psychometrika*. https://doi.org/gmwnj5

Stanley, L. M. et Edwards, M. C. (2016). Reliability and model fit. *Educational and Psychological Measurement*, *76*(6), 976-985. https://doi.org/g4ps

Trizano-Hermosilla, I. et Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions : Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, 769. https://doi.org/gfjz27

Westfall, P. et Henning, K. (2013). *Understanding advanced statistical methods*. CRC Press.