Revue des sciences de l'éducation

L'alpha de Cronbach est l'un des pires estimateurs de la consistance interne : une étude de simulation

Cronbach's alpha is one of the worst internal consistency estimators: a simulation study

El alfa de Cronbach es uno de los peores estimadores de la consistencia interna: un estudio de simulación

Jimmy Bourque, Danielle Doucet, Josée LeBlanc, Jérémie Dupuis and Josée Nadeau

Volume 45, Number 2, 2019

URI: https://id.erudit.org/iderudit/1067534ar
DOI: https://doi.org/10.7202/1067534ar

See table of contents

Publisher(s)

Revue des sciences de l'éducation

ISSN

1705-0065 (digital)

Explore this journal

Cite this article

Bourque, J., Doucet, D., LeBlanc, J., Dupuis, J. & 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/10.7202/1067534ar

Article abstract

Cronbach's alpha is the most common internal consistency index in educational sciences. The goal of this article is to assess the performance of six internal consistency estimators by conducting a simulation study. The simulation focuses on Cronbach's alpha, lambda-2, Guttman's lambda-4 and lambda-6, the greatest lower bound (GLB) and the omega. Forty-five scenarios were defined by the sample size, the number of items and the value of factorial saturation coefficients. The results suggest that if the instrument has five items, the preferred estimator would be omega. In other cases, it would be the GLB. Alpha and lambda-2 are systematically the two estimators that most underestimate the value of internal consistency and should be avoided. Lambda-6 would be the best estimator offered by SPSS. Overall, this study offers an empirical rationale for a change of practice in educational research.

L'alpha de Cronbach est l'un des pires estimateurs de la consistance interne : une étude de simulation

Jimmy Bourque
Professeur
Université de Moncton

Danielle Doucet
Coordonnatrice de recherche
Université de Moncton

Josée LeBlanc
Psychologue
Réseau de santé Vitalité

Jérémie Dupuis
Étudiant au doctorat
Université de Moncton

Josée Nadeau
Professeure
Université de Moncton

RÉSUMÉ–L'alpha de Cronbach est l'indice de consistance interne le plus répandu en sciences de l'éducation. Le but de cet article est d'évaluer la performance de six estimateurs de consistance interne à partir d'une étude de simulation. La simulation porte sur l'alpha de Cronbach, le lambda-2, le lambda-4 et le lambda-6 de Guttman, la plus grande limite inférieure et l'oméga. Quarante-cinq scénarios ont été définis par la taille de l'échantillon, le nombre d'items et la valeur des coefficients de saturation factorielle. Les résultats suggèrent que, dans le cas où l'instrument compte cinq items, l'estimateur à privilégier serait l'oméga. Dans les autres cas, ce serait la grande limite inférieure. L'alpha et le lambda-2 sont systématiquement les deux estimateurs qui sous-estiment le plus la valeur de la consistance interne et devraient être évités. Le lambda-6 serait le meilleur estimateur offert par SPSS. Dans l'ensemble, cette étude offre un rationnel empirique pour un changement de pratique dans les recherches en éducation.

MOTS-CLÉS—fidélité, consistance interne, mesure, simulation, alpha de Cronbach.

1. Introduction et problématique

Selon la théorie classique des tests, le score total à un test (X) ne représente jamais le score vrai (V, la quantité exacte que l'on tente de mesurer) et est toujours influencé par une part d'erreur de mesure (ε, Streiner, 2003) :

$$X = V + \varepsilon$$

L'erreur se divise elle-même en deux composantes : l'erreur aléatoire, distribuée normalement et dont la moyenne est 0, et l'erreur systématique, dont la distribution est

asymétrique et dont la moyenne diffère de 0. Si l'erreur aléatoire n'introduit pas de biais systématique dans la mesure, l'erreur systématique, lorsqu'elle diffère de 0, fera en sorte que le score observé surestimera ou sous-estimera systématiquement le score vrai.

Les estimateurs de fidélité permettent d'estimer à quel point le score observé se rapproche du score vrai. Par exemple, les indices de consistance interne comme l'alpha de Cronbach sont des mesures de fidélité qui font référence au degré auquel les items mesurent le même concept (Cronbach, 1951). Dans les recherches en éducation, le coefficient alpha est en ce moment le plus couramment utilisé pour évaluer la consistance interne des scores à un test (Cho, 2016 ; Sijtsma, 2009a). Son omniprésence dans les écrits scientifiques, en particulier en éducation, est quelque peu étonnante puisque cet indice ne semble pas constituer le meilleur estimateur de la fidélité réelle, une situation qui serait connue depuis au moins la fin des années soixante-dix (Callender et Osburn, 1979 ; Sijtsma, 2009a).

Paradoxalement, l'alpha est sans contredit l'une des statistiques les plus critiquées et les plus mal comprises en sciences sociales (Cho, 2016 ; Cortina, 1993 ; Green et Yang, 2009 ; Sijtsma, 2009a). Cronbach évoque d'ailleurs lui-même le fait que les nombreux·ses chercheur·se·s ayant utilisé l'alpha n'ont pas nécessairement lu et compris son texte original (Cronbach et Shavelson, 2004). D'ailleurs, plusieurs articles portant sur le coefficient alpha et ses limites se retrouvent dans la revue *Psychometrika*, une publication spécialisée s'adressant à des spécialistes de la psychométrie, et non à l'ensemble des chercheur·se·s des sciences sociales ou des sciences de l'éducation. De nombreux expert·e·s se méfient de l'usage du coefficient alpha, non seulement en tant que mesure de la consistance interne, mais aussi comme mesure de la fidélité (Bentler, 2009 ; Cho et Kim, 2015 ; Davenport, Davison, Liou et Love, 2015 ; Green et Yang, 2009 ; Revelle et Zinbarg, 2009 ; Sijtsma, 2009a). On affirme d'abord que les conditions qui sous-tendent son utilisation (tau-équivalence ou items mesurant un seul facteur et ayant la même variance, unidimensionnalité et non-corrélation des erreurs, entre autres) sont très rarement satisfaites en pratique (Green et Yang, 2009). De plus, en tant que borne inférieure de la fidélité réelle du score à un test, le coefficient alpha sous-estime largement et systématiquement cette valeur (Callender et Osburn, 1979 ; Schmitt, 1996 ; Sijtsma, 2009a). Ceci peut avoir d'importantes conséquences empiriques et cliniques (Green et Yang, 2009 ; Henson, 2001). Au plan empirique, les chercheur·se·s en recherche appliquée pourraient rejeter un instrument psychométrique en pensant qu'il ne mène pas à des scores fidèles. De même, lorsque les erreurs sont corrélées (ce qui est souvent le cas), le coefficient alpha serait surestimé (Green et Yang, 2009). Un·e chercheur·se pourrait donc

conserver un instrument en pensant qu'elle ou il produit des scores fidèles, alors que ce n'est pas le cas. Au plan clinique, la première situation peut faire en sorte de limiter les choix d'instruments disponibles pour les clinicien·ne·s ; la deuxième situation peut entrainer l'utilisation d'instruments qui produisent en réalité des scores non fidèles, ce qui peut mener à un dépistage ou à un diagnostic erroné, à la constatation de faux progrès, à l'admission ou au rejet d'un candidat ou d'un patient, etc.

D'autres coefficients de fidélité moins biaisés existent et permettent de contourner les difficultés rencontrées avec l'alpha (Green et Yang, 2009 ; Osburn, 2000 ; Revelle et Zinbarg, 2009 ; Sijtsma, 2009a ; Trizano-Hermosilla et Alvarado, 2016). Par exemple, Guttman (1945) a présenté un modèle comprenant une série de six coefficients de fidélité, de lambda-1 à lambda-6. Le lambda-3 s'avère mathématiquement équivalent à l'alpha (Callender et Osburn, 1979 ; Cho et Kim, 2015). Bien que le modèle de Guttman et, donc, le lambda-3 aient été avancés avant l'alpha, proposé en 1951, la communauté scientifique a crédité Cronbach, bien malgré lui, en associant son nom à l'alpha (Cho et Kim, 2015 ; Cronbach et Shavelson, 2004). En effet, quoique mathématiquement équivalent, le coefficient alpha était plus facile à calculer que le lambda-3, à une époque où les calculs à la main dominaient (Cho, 2016 ; Cho et Kim, 2015). Le lambda-2 de Guttman était quant à lui supérieur à l'alpha, mais requérait aussi une opération mathématique élaborée (Cho et Kim, 2015). D'autres coefficients ont par la suite été avancés, dont la plus grande limite inférieure (en anglais *greatest lower bound* ; Woodhouse et Jackson, 1977), le coefficient oméga (McDonald, 1978), le coefficient bêta (Revelle, 1979) et la série de coefficients mu (0 à 6), le coefficient mu 1 étant équivalent au lambda-2 (Ten Berge et Zegers, 1978). Enfin, plus récemment, la modélisation par équations structurelles a aussi été mise à contribution pour estimer la fidélité (Bentler, 2009 ; Cho et Kim, 2015 ; Green et Yang, 2009).

L'existence de plusieurs coefficients dont la performance est supérieure à celle de l'alpha et le fait que Cronbach lui-même ait reconnu que l'alpha n'était qu'une mesure de fidélité parmi d'autres (Cronbach et Shavelson, 2004) rendent pertinent le questionnement sur ce qui entraine cette persistante popularité de l'alpha. Quelques explications sont mises de l'avant, comme l'habitude et l'accessibilité (Cho, 2016 ; Revelle et Zinbarg, 2009 ; Sijtsma, 2009a, 2009b ; Trizano-Hermosilla et Alvarado, 2016). En effet, l'alpha est très connu et peut être calculé aisément à partir d'IBM SPSS, le logiciel le plus répandu dans les facultés d'éducation. Cho (2016) compare quant à lui l'alpha à une marque de commerce qui s'est taillé une place et qui conserve une part importante du marché.

Cet article cherche à évaluer, à partir de simulations, la performance du coefficient alpha et de cinq autres estimateurs de consistance interne (lambda-2, lambda-4_{max}, lambda-6, plus grande limite inférieure et oméga) en fonction d'une variété de scénarios quant à la taille de l'échantillon, au nombre d'items et à la valeur des coefficients de saturation factorielle. Plus précisément, nous porterons attention à deux caractéristiques de ces indices : leur valeur moyenne et leur variance, la stabilité constituant un autre trait désirable d'un bon indice de fidélité.

2. Contexte théorique

2.1 Fidélité et consistance interne

Comme nous l'avons mentionné au début de ce texte, selon la théorie classique des tests, le score total à un test (*X*) est constitué d'une part du score réel (*V*) et, d'autre part, de l'erreur de mesure (ε) :

$$X = V + \varepsilon$$

Les mesures de fidélité cherchent à évaluer à quel point une mesure est entachée d'erreur ou, autrement dit, à quel point le score total se rapproche du score réel. Streiner, Norman et Cairney (2015, p. 9) proposent la définition générale suivante de la fidélité : « le degré auquel l'évaluation d'individus à différentes occasions, ou par des observateurs différents, ou par des tests similaires ou parallèles, produit les mêmes résultats ou des résultats similaires » (traduction libre). La fidélité (r_X) peut donc être exprimée soit comme la corrélation entre le score observé et le score vrai (r_{XV}),

$$r_X = r_{XV}$$

soit comme le rapport entre la variance du score vrai (σ_V^2) et la variance du score au test (σ_X^2) :

$$r_X = \sigma_V^2 / \sigma_X^2$$

Il existe plusieurs indices pour estimer la fidélité. Les experts en psychométrie considèrent l'alpha de Cronbach comme une estimation de la consistance interne des items d'un instrument. Selon Cronbach (1951), la consistance interne fait référence à l'homogénéité des items, c'est-à-dire à quel point les items du test sont similaires ou, autrement dit, à quel point ils mesurent la même dimension d'un construit (unidimensionnalité). Or, plusieurs chercheur·se·s ont démontré de façon robuste qu'une valeur élevée de l'alpha ne se traduit pas nécessairement en une homogénéité ou en une unidimensionnalité des items. En revanche, ces chercheur·se·s acceptent plutôt que la consistance interne représente à quel point les items

d'une échelle sont étroitement reliés ou corrélés (Cho et Kim, 2015 ; Cortina, 1993 ; Schmitt, 1996).

2.2 Alpha de Cronbach

On fait souvent référence à l'alpha comme à une estimation de la fidélité telle que Cronbach (1951) l'avait décrite :

$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\sum_{i=1}^{k} s^2(Y_i)}{s_Y^2}\right)$ (Où k est le nombre d'items).

Toutefois, six ans plus tôt, Guttman (1945) avait plutôt conçu l'alpha (qui correspond à son lambda-3) comme une limite inférieure à la fidélité réelle étant donné sa valeur normalement plus petite que lambda-2 lorsque certaines conditions, dont la tau-équivalence, n'étaient pas réunies. Pour sa part, Cortina (1993) avance que la valeur de l'alpha est plus exacte lorsque trois conditions sont satisfaites : 1) l'unidimensionnalité (les items mesurent un seul facteur), 2) la tau-équivalence et 3) les erreurs d'items non corrélées. Toutefois, en pratique, ces conditions (surtout celle de tau-équivalence) sont rarement remplies, ce qui mène à une sous-estimation systématique de la fidélité réelle. De plus, la valeur de l'alpha est fortement influencée par le nombre d'items, le nombre de dimensions orthogonales et la moyenne des corrélations entre items (Cortina, 1993). Par conséquent, d'autres indices de fidélité ont été proposés.

2.3 Lambda de Guttman

Guttman (1945) a dérivé six coefficients ($L_1 – L_6$) qu'il décrit comme des limites inférieures à la fidélité. L'efficacité de chacun de ces indices à mesurer la fidélité réelle est influencée par différentes conditions. Le lambda-1 est le premier indice calculé et sous-estime largement la fidélité réelle. Il n'est pas utilisé comme estimateur de fidélité, mais constitue plutôt une étape intermédiaire vers le calcul des autres indices de Guttman.

$$L_1 = 1 - \frac{\sum_{k=1}^{n} s_k^2}{s_Y^2}$$

(Où s_k^2 est la variance de l'item k, et s_Y^2 est la variance du score dérivé des items.)

Le lambda-2 est toujours plus élevé que le lambda-1 et est supérieur ou égal au lambda-3 (donc à l'alpha de Cronbach) en cas d'indépendance entre les erreurs des items.

$$L_2 = L_1 + \frac{\sqrt{\frac{n}{n-1} C_2}}{s_Y^2}$$

(Où C_2 est la somme des carrés de la covariance entre les items.)

Comme déjà mentionné, le lambda-3 est l'équivalent de l'alpha de Cronbach et est toujours plus élevé que le lambda-1. Le lambda-4 représente quant à lui un coefficient de bissection. Comme il y a plusieurs façons de diviser un test en deux, la valeur de lambda-4 représente la moyenne de toutes les bissections possibles.

$$L_4 = 2\left[1 - \frac{s_a^2 + s_b^2}{s_Y^2}\right]$$

(Où s_a^2 et s_b^2 sont les variances des deux ensembles d'items créés par la bissection.)

Le lambda-5, quant à lui, est efficace lorsqu'il existe une covariance élevée entre un item et les autres, lesquels, en retour, n'ont pas une covariance élevée entre eux, ce qui n'est pas désirable dans le cas d'une échelle psychométrique ou édumétrique.

$$L_5 = L_1 + \frac{2\sqrt{\bar{C}_2}}{s_Y^2}$$

(Où $\bar{C}_2$ est la somme des carrés des covariances d'un item en particulier avec les k - 1 autres.)

Finalement, le lambda-6 reflète la proportion de variance moyenne de chaque item expliquée en régressant cet item sur tous les autres items, c'est-à-dire la corrélation multiple au carré.

$$L_6 = 1 - \frac{\sum_{i=1}^{k} e_i^2}{s_Y^2}$$

(Où e_i^2 est la variance de l'estimateur de l'item i provenant de la régression linéaire sur les autres items.)

Les deux lambdas de Guttman qui disposent du plus grand appui empirique quant à leur capacité à estimer la fidélité réelle sont le lambda-2 (Tang et Cui, 2012) et le lambda-4 (Callender et Osburn, 1979 ; Hunt et Bentler, 2015 ; Osburn, 2000), mais il faut noter que certains des indices, le lambda-6 par exemple, n'ont fait l'objet que de peu d'études.

2.4 Plus grande limite inférieure (*greatest lower bound*)

En 1977, Jackson et Agunwamba se sont intéressés aux indices de Guttman et ont démontré mathématiquement que la valeur maximale du coefficient de bissection lambda-4 (au lieu de la valeur moyenne suggérée par Guttman) est d'habitude la valeur la plus élevée de ces indices et, donc, le meilleur estimateur de la fidélité réelle puisque sa valeur ne peut pas dépasser la valeur de cette dernière (il s'agit d'une limite inférieure). Woodhouse et Jackson (1977) décrivent ensuite comment dériver la grande limite inférieure :

$$GLB = 1 - \frac{tr(C_e)}{s_Y^2}$$

(Où $tr(C_e)$ est la trace de la matrice de covariance des erreurs interitems.)

Dans la majorité des cas, la plus grande limite inférieure est en fait le lambda-4 et l'article démontre la supériorité de ce coefficient sur le lambda-2 et le lambda-3 (l'alpha de Cronbach). Plus récemment, Sijtsma (2009a) a également confirmé que la plus grande limite inférieure constitue un meilleur estimateur de la fidélité que l'alpha. Par contre, Shapiro et Ten Berge (2000) ainsi que Ten Berge et Sočan (2004) ont établi que la plus grande limite inférieure peut surestimer la valeur de la fidélité réelle lorsque calculée à partir de petits échantillons (moins de quelques milliers d'observations) si le nombre de variables est élevé.

2.5 Oméga

De son côté, McDonald (1978, 1985), partant d'une approche d'analyse factorielle, propose le coefficient de fidélité oméga, qui a l'avantage de tenir compte de la force de l'association entre les items et un construit, d'une part, ainsi qu'entre les items et l'erreur de mesure, d'autre part. Par conséquent, selon cet auteur, l'oméga permet une estimation plus exacte de la vraie fidélité de l'échelle. Ce coefficient fonctionne bien même lorsque la condition de tau-équivalence n'est pas respectée et McDonald (1978) démontre mathématiquement que l'oméga sera toujours supérieur ou égal à l'alpha. Pour l'essentiel, l'oméga est le carré de la corrélation entre le score au test et le score au domaine :

$$\omega = \lim_{q \to \infty} R^2 = \frac{\sum_j \sum_k \sum_l f_{jl} f_{kl}}{\sum_j \sum_k \sum_l f_{jl} f_{kl} + \sum_j u_j^2}$$

(Où q est le nombre d'items du domaine, j est le nombre d'items d'une première moitié de l'échelle, k est le nombre d'items de l'autre moitié, l est le nombre de facteurs, f le coefficient de saturation de l'item et u, son unicité.)

Plusieurs études de simulation justifient l'utilisation de l'oméga de McDonald comme un indice de fidélité alternatif à l'alpha. En effet, dans l'étude de Revelle et Zinbarg (2009), l'oméga de McDonald a été meilleur que 12 autres indices. De plus, deux études en 2016 (Kelley et Pornprasertmanit, 2015 ; Trizano-Hermosilla et Alvarado, 2016) ont aussi démontré la supériorité de ce coefficient. D'autres indices de fidélité ont été proposés, le Beta de Revelle (1979), les indices de Ten Berge et Zegers (μ_0, μ_1, μ_2… ; 1978) ainsi que, plus récemment, des méthodes qui relèvent de la modélisation par équations structurelles (Cho et Kim, 2015 ; Green et Yang, 2009)**.** Toutefois, l'explication de chacune de ces méthodes dépasse les buts de cet article et elles ne seront pas testées dans notre simulation.

Enfin, cette étude a pour objectifs d'enrichir les résultats de certaines études antérieures (Callender et Osburn, 1979 ; Revelle et Zinbarg, 2009 ; Woodhouse et Jackson, 1977) à l'intention des chercheurs francophones en éducation tout en émettant des considérations supplémentaires sur la variance, absente des études antérieures. Outre cette attention portée à la variance, l'apport original de cet article découle de la variété des scénarios explorés, notamment en ce qui a trait à la variation de la valeur des coefficients de saturation factorielle.

3. Méthodologie

La recherche rapportée ici porte sur une simulation qui vise à comparer la performance de six indices de consistance interne, soit l'alpha de Cronbach (Cronbach, 1951), les lambda-2, lambda-4_{max} et lambda-6 de Guttman (Guttman, 1945), la plus grande limite inférieure (Ten Berge et Sočan, 2004) et l'oméga de McDonald (McDonald, 1978, 1985). Ces indices sont comparés selon 45 conditions définies par trois paramètres : le nombre d'observations (50, 100, 200, 500 ou 1 000), le nombre d'items (5, 10 ou 20) et la valeur moyenne des coefficients de saturation factorielle (0,35, 0,50 ou 0,80). Notons que pour cette dernière condition, les coefficients de saturation sont extraits d'une distribution normale tronquée à 0 et 1 et dont la moyenne correspond à chacune des trois conditions expérimentales. Nous avons fixé le nombre de réplications à 2 000. La simulation a été effectuée avec le logiciel *R* (la syntaxe est fournie en annexe).

Les *packages* requis pour la simulation sont les suivants : *mvtnorm*, *Matrix*, *Rcsdp*, *psych*, *GPArotation*, *e107*, *truncnorm* et *plyr*. Le moteur de simulation est fonction du nombre de réplications, du nombre d'observations, du nombre d'items et de la valeur des coefficients de saturation factorielle. Dans un premier temps, une matrice de la variance de l'erreur de mesure est créée. Il s'agit d'une matrice de dimension *k* x *k* (*k* étant le nombre d'items), où la

diagonale contient la valeur 1 - l_i^2 (l_i étant la valeur du coefficient de saturation factorielle extrait pour l'item *i* à l'aide de la commande *rtruncnorm*). Dans un deuxième temps, la matrice de la variance attribuable au facteur latent est produite. Il s'agit d'une autre matrice *k* x *k*, sur la diagonale de laquelle on retrouve le carré des coefficients de saturation (l_i^2). La matrice des variances observées est obtenue en additionnant ces deux matrices et est donc fixée à 1 pour tous les termes (1 - l_i^2 + l_i^2).

La matrice de variance-covariance des items est obtenue par la simulation d'une distribution de Wishart à *k* + 1 degrés de liberté (*k* étant le nombre d'items) qui est fonction du nombre d'observations et de la matrice de la variance de l'erreur. Le moteur de simulation est ensuite appelé à calculer l'alpha de Cronbach, les lambda-2, lambda-4_{max} et lambda-6 de Guttman, la plus grande limite inférieure et l'oméga de McDonald. Ces calculs sont répétés 2 000 fois, les données simulées étant obtenues sous forme d'une matrice de variance-covariance fonction du nombre d'observations et de la matrice des variances observées.

Les indicateurs de performance des estimateurs de fidélité rapportés ici sont leur valeur moyenne sur l'ensemble des réplications, la variance et l'intervalle de confiance de 95 % associé à ces deux statistiques. Le moteur se termine avec l'identification des 45 (5 x 3 x 3) conditions expérimentales à partir du nombre d'observations, du nombre d'items et de la valeur des coefficients de saturation factorielle. Le calcul d'intervalles de confiance de 95 % pour les moyennes et les variances permet de déterminer si les indices diffèrent de façon statistiquement significative quant à ces paramètres.

4. Résultats

4.1 Quels sont les estimateurs dont la valeur est la plus élevée ?

Selon les résultats de l'étude de simulation, la plus grande limite inférieure produit l'estimation la plus élevée de la fidélité dans 30 cas sur 45 (67 %). Les seuls cas où la plus grande limite inférieure est supplantée (15/45, 33 %) sont ceux où l'échelle évaluée compte peu d'items (5 dans notre étude) et dont les saturations sont faibles (0,35) ou modérées (0,50). Dans cette situation, c'est plutôt l'oméga qui s'approche le plus de la fidélité réelle. Dans les cas (5/45, 11 %) où il y a peu d'items et où la valeur des coefficients de saturation est élevée (0,80), ces deux indices sont équivalents. À l'opposé, l'alpha de Cronbach et le lambda-2 de Guttman sous-estiment systématiquement et significativement la fidélité réelle, et ce, dans tous les scénarios simulés. L'alpha de Cronbach est le second pire estimateur dans 100 % des cas.

Les facteurs qui influencent le plus la valeur des indices de fidélité sont le nombre d'items et la valeur des coefficients de saturation. La valeur de la fidélité estimée augmente considérablement lorsque le nombre d'items passe de cinq à dix, alors que les faibles coefficients de saturation affectent négativement la valeur estimée, surtout pour les estimateurs les moins efficaces comme l'alpha et le lambda-2. Parmi les indices disponibles sur SPSS, c'est le lambda-6 qui obtient les meilleurs résultats. Dans les tableaux qui suivent, les différentes couleurs sur une même ligne identifient les valeurs significativement différentes (les intervalles de confiance ne se chevauchent pas) alors que des valeurs surlignées de la même couleur sur une même ligne ne diffèrent pas significativement au seuil de 0,05. Le vert représente les valeurs les plus élevées alors que l'orange désigne les indices dont les valeurs sont les plus basses et inférieures au seuil recommandé de 0,80 dans le cas des valeurs moyennes.

Tableau 1
Valeur moyenne des six indices de fidélité

L	K	N	Alpha	Lambda-2	Lambda-4	Lambda-6	GLB	Omega
0,35	5	50	0,70	0,45	0,82	0,83	0,82	0,89
		100	0,70	0,43	0,82	0,82	0,82	0,89
		200	0,70	0,41	0,82	0,82	0,81	0,89
		500	0,70	0,47	0,82	0,82	0,82	0,89
		1000	0,70	0,41	0,82	0,82	0,81	0,89
	10	50	0,77	0,59	0,91	0,94	0,95	0,80
		100	0,77	0,57	0,91	0,94	0,95	0,80
		200	0,77	0,58	0,91	0,94	0,95	0,80
		500	0,77	0,60	0,92	0,94	0,95	0,80
		1000	0,77	0,58	0,91	0,94	0,95	0,80
	20	50	0,83	0,73	0,94	1,00	0,99	0,78
		100	0,83	0,74	0,94	1,00	0,99	0,78
		200	0,83	0,73	0,94	1,00	0,99	0,78
		500	0,83	0,72	0,94	1,00	0,99	0,77
		1000	0,83	0,73	0,94	1,00	0,99	0,78
0,50	5	50	0,73	0,58	0,83	0,85	0,86	0,90
		100	0,73	0,57	0,83	0,84	0,86	0,90
		200	0,73	0,58	0,83	0,84	0,86	0,90
		500	0,73	0,56	0,83	0,84	0,86	0,90
		1000	0,73	0,58	0,83	0,84	0,86	0,90
	10	50	0,81	0,74	0,93	0,96	0,97	0,85
		100	0,81	0,74	0,93	0,96	0,97	0,85
		200	0,81	0,74	0,93	0,96	0,97	0,85
		500	0,81	0,74	0,93	0,96	0,97	0,85
		1000	0,81	0,73	0,93	0,96	0,97	0,84
	20	50	0,88	0,84	0,96	1,00	0,99	0,87
		100	0,88	0,85	0,96	1,00	0,99	0,87
		200	0,88	0,84	0,96	1,00	0,99	0,87
		500	0,88	0,84	0,96	1,00	0,99	0,87
		1000	0,88	0,84	0,96	1,00	0,99	0,87
0,80	5	50	0,86	0,83	0,89	0,94	0,95	0,95
		100	0,87	0,82	0,89	0,94	0,94	0,95
		200	0,86	0,83	0,89	0,94	0,95	0,95
		500	0,87	0,82	0,89	0,94	0,95	0,95
		1000	0,87	0,83	0,90	0,94	0,95	0,95
	10	50	0,94	0,91	0,97	0,99	0,99	0,95
		100	0,94	0,91	0,97	0,99	0,99	0,95
		200	0,94	0,91	0,97	0,99	0,99	0,95
		500	0,94	0,91	0,97	0,99	0,99	0,95
		1000	0,93	0,91	0,97	0,99	0,99	0,95
	20	50	0,97	0,95	0,99	1,00	1,00	0,96
		100	0,97	0,95	0,99	1,00	1,00	0,97
		200	0,97	0,95	0,99	1,00	1,00	0,96
		500	0,97	0,95	0,99	1,00	1,00	0,97
		1000	0,97	0,95	0,99	1,00	1,00	0,97

Tableau 2
Variance des six estimateurs de fidélité sur 2 000 itérations

L	K	N	Alpha	Lambda-2	Lambda-4	Lambda-6	GLB	Omega
0,35	5	50	0,005	0,190	0,009	0,011	0,026	0,004
		100	0,005	0,203	0,010	0,011	0,026	0,005
		200	0,005	0,247	0,008	0,011	0,030	0,005
		500	0,005	0,172	0,009	0,011	0,024	0,005
		1000	0,005	0,255	0,009	0,011	0,029	0,005
	10	50	0,001	0,039	0,001	0,004	0,002	0,004
		100	0,001	0,046	0,001	0,004	0,002	0,003
		200	0,001	0,037	0,001	0,004	0,002	0,004
		500	0,001	0,033	0,001	0,004	0,001	0,004
		1000	0,001	0,043	0,001	0,004	0,002	0,004
	20	50	< 0,001	0,007	< 0,001	< 0,001	< 0,001	0,004
		100	< 0,001	0,007	< 0,001	< 0,001	< 0,001	0,004
		200	< 0,001	0,007	< 0,001	< 0,001	< 0,001	0,004
		500	< 0,001	0,008	< 0,001	< 0,001	< 0,001	0,004
		1000	< 0,001	0,008	< 0,001	< 0,001	< 0,001	0,004
0,50	5	50	0,006	0,175	0,008	0,010	0,019	0,004
		100	0,006	0,121	0,008	0,009	0,020	0,004
		200	0,006	0,089	0,008	0,010	0,019	0,004
		500	0,006	0,331	0,008	0,010	0,020	0,004
		1000	0,007	0,101	0,008	0,010	0,020	0,005
	10	50	0,002	0,017	0,001	0,003	0,001	0,004
		100	0,002	0,016	0,001	0,002	0,001	0,004
		200	0,002	0,016	0,001	0,003	0,001	0,004
		500	0,002	0,015	0,001	0,002	0,001	0,004
		1000	0,002	0,018	0,001	0,002	0,001	0,004
	20	50	0,001	0,003	< 0,001	< 0,001	< 0,001	0,002
		100	0,001	0,002	< 0,001	< 0,001	< 0,001	0,002
		200	0,001	0,003	< 0,001	< 0,001	< 0,001	0,003
		500	0,001	0,003	< 0,001	< 0,001	< 0,001	0,003
		1000	0,001	0,003	< 0,001	< 0,001	< 0,001	0,003
0,80	5	50	0,007	0,016	0,005	0,004	0,003	0,002
		100	0,006	0,025	0,005	0,004	0,005	0,002
		200	0,006	0,018	0,005	0,004	0,003	0,002
		500	0,006	0,022	0,004	0,004	0,004	0,002
		1000	0,006	0,019	0,004	0,003	0,003	0,002
	10	50	0,001	0,002	< 0,001	< 0,001	< 0,001	0,001
		100	0,001	0,002	< 0,001	< 0,001	< 0,001	0,001
		200	0,001	0,002	< 0,001	< 0,001	< 0,001	0,001
		500	0,001	0,002	< 0,001	< 0,001	< 0,001	0,001
		1000	0,001	0,002	< 0,001	< 0,001	< 0,001	0,001
	20	50	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
		100	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
		200	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
		500	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
		1000	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

4.2 Quels sont les estimateurs les plus stables ?

La stabilité (faible variance) d'une statistique est une caractéristique importante en ce qu'elle suggère que sa valeur reflète surtout du signal, par opposition au bruit (*noise*). Dans le cas de cette étude de simulation, les estimateurs de fidélité démontrant la plus faible variance sur les 2 000 échantillons simulés varient selon les conditions expérimentales. Dans le scénario le plus favorable (20 items et coefficients de saturation élevés, 0,80 : 5/45, 11 %), les variances des six indices sont très faibles et statistiquement équivalentes. Dans les cas où l'échelle analysée ne compte que cinq items (15/45, 33 %), c'est l'oméga qui affiche systématiquement la plus faible variance. L'alpha s'avère l'estimateur le plus stable lorsque les coefficients de saturation factorielle sont faibles (0,350). Dans les autres cas, les résultats sont partagés. À l'opposé, c'est le lambda-2 qui démontre systématiquement la variance la plus élevée.

5. Discussion

Les résultats obtenus sont cohérents avec d'autres études. Par exemple, la supériorité de l'oméga pour estimer la fidélité de scores tirés d'échelles constituées de peu d'items correspond aux observations de Revelle et Zinbarg (2009). Selon ces auteur·e·s, dans une étude portant sur 13 indices de fidélité calculés sur neuf bases de données provenant de l'administration de tests comptant de quatre à huit items, l'oméga a surpassé les autres indices dans sept cas sur neuf (échelles de quatre ou de six items). De façon similaire, nos résultats établissent la supériorité de l'oméga pour les scénarios où l'instrument compterait cinq items. Dans les deux cas restants de l'étude de Revelle et Zinbarg (six et huit items), c'est le lambda-4 qui donnait la meilleure valeur. Une variante du lambda-4 (MSPLIT) avait également surpassé l'alpha et le lambda-2 dans l'étude de Callender et Osburn (1979), portant sur une échelle de 10 items cette fois. Dans ce cas, nos résultats corroborent la supériorité du lambda-4 sur les deux autres indices, mais pas celle du lambda-2 sur l'alpha. De plus, comme dans l'étude de Woodhouse et Jackson (1977), la plus grande limite inférieure constitue ici un estimateur systématiquement supérieur ou égal au lambda-4, et toujours plus élevé que le lambda-2 et l'alpha. Par contre, il est possible qu'il s'agisse d'un artéfact du biais d'échantillonnage documenté par Shapiro et Ten Berge (2000) ainsi que Ten Berge et Sočan (2004).

Un ajout important de notre étude est la considération de la variance de ces différents indices. Nous observons que selon les scénarios, il peut y avoir des différences importantes de variance. Par exemple, si l'oméga présente la variance la plus faible pour les scénarios où le

score est dérivé de cinq items seulement, il obtient la variance la plus élevée dans les cas où il y a 20 items et où la valeur des coefficients de saturation est modérée (0,50).

Sur la base de ces résultats, en tenant compte à la fois de la valeur moyenne des estimateurs de fidélité et de leur variance, nous formulons, dans le tableau 3, des recommandations quant à l'utilisation des différents indices selon les scénarios. Ainsi, dans le cas où le nombre d'items composant le test est faible (de l'ordre de cinq items), nos résultats suggèrent que l'utilisation de l'oméga serait indiquée. Dans les autres scénarios, le recours à la plus grande limite inférieure serait préférable, mais vu ses limites (Shapiro et Ten Berge, 2000 ; Ten Berge et Sočan, 2004), il serait probablement plus prudent de recourir à l'oméga ou au lambda-6. Si SPSS constitue la seule option en matière de logiciel d'analyse, le lambda-6 de Guttman constitue une solution plus avantageuse que l'alpha de Cronbach.

Tableau 3
Indice de fidélité recommandé selon les conditions expérimentales

Coefficient de saturation	Nombre d'items	Indice recommandé	Indice recommandé : SPSS
0,35	5	Oméga	Lambda-4 ou Lambda-6
	10	GLB	Lambda-6
	20	Lambda-6	Lambda-6
0,50	5	Oméga	Lambda-6
	10	GLB	Lambda-6
	20	GLB ou Lambda-6	Lambda-6
0,80	5	Oméga ou GLB	Lambda-6
	10	GLB	Lambda-6
	20	GLB	Lambda-6

6. Conclusion

Avec cet article, nous voulions comparer la performance de six estimateurs de la consistance interne pour évaluer la fidélité des scores à un instrument psychométrique. À la suite d'écrits plus ou moins récents retrouvés dans des périodiques anglo-saxons, nous soupçonnions que l'indice le plus utilisé en sciences de l'éducation, l'alpha de Cronbach, serait en fait l'un des moins efficaces. C'est ce que nos résultats ont corroboré : l'alpha et le lambda-2 sous-estiment systématiquement et significativement la fidélité dans tous les scénarios testés. Les meilleurs indices seraient en fait l'oméga s'il y a peu d'items, et la plus grande limite inférieure ou le lambda-6 dans les autres cas.

Notre étude se démarque des écrits antérieurs de deux façons. D'abord, en plus de la valeur moyenne de la fidélité de consistance interne estimée par chaque indice, nous avons

porté attention à sa variance. Comme la stabilité de l'estimateur constitue une caractéristique désirable, nous disposions d'un autre critère d'évaluation. Enfin, nous avons fait varier l'ampleur des coefficients de saturation factorielle, ce qui a permis de constater deux choses. Premièrement, la variance de tous les estimateurs est plus ou moins équivalente et négligeable lorsque les coefficients de saturation et le nombre d'items sont simultanément élevés. Deuxièmement, l'impact des coefficients de saturation sur la valeur des estimateurs de fidélité semble moindre que celui du nombre d'items.

Notre étude compte évidemment un certain nombre de limites, comme le choix restreint d'estimateurs de fidélité comparés. Les récents développements en psychométrie ont suggéré plusieurs autres méthodes d'estimation de la fidélité réelle, dont des approches basées sur la modélisation par équations structurelles qui n'ont pas été abordées ici. Il serait intéressant de voir comment ces nouvelles méthodes se comparent aux approches plus classiques. De plus, le fait que les coefficients de saturation aient été simulés pour être d'un même ordre de grandeur correspond à une situation théorique fréquemment observée en pratique, mais pas toujours : par exemple, le développement et la validation de nouveaux tests peuvent mener à de grandes disparités entre les coefficients de saturation factorielle des items adéquats et de ceux dont la performance est inadéquate. Des recherches futures devraient tester divers scénarios quant aux valeurs relatives des coefficients de saturation, la présence d'un ou quelques items avec des coefficients très faibles parmi un ensemble d'items aux coefficients élevés, par exemple.

ENGLISH TITLE–Cronbach's alpha is one of the worst internal consistency estimators: a simulation study

SUMMARY–Cronbach's alpha is the most common internal consistency index in educational sciences. The goal of this article is to assess the performance of six internal consistency estimators by conducting a simulation study. The simulation focuses on Cronbach's alpha, lambda-2, Guttman's lambda-4 and lambda-6, the greatest lower bound (GLB) and the omega. Forty-five scenarios were defined by the sample size, the number of items and the value of factorial saturation coefficients. The results suggest that if the instrument has five items, the preferred estimator would be omega. In other cases, it would be the GLB. Alpha and lambda-2 are systematically the two estimators that most underestimate the value of internal consistency and should be avoided. Lambda-6 would be the best estimator offered by SPSS. Overall, this study offers an empirical rationale for a change of practice in educational research.

KEYWORDS–fidelity, internal consistency, measure, simulation, Cronbach's alpha.

TÍTULO–El alfa de Cronbach es uno de los peores estimadores de la consistencia interna: un estudio de simulación

RESUMEN–El alfa de Cronbach es el índice de consistencia interna más extendido en ciencias de la educación. El objetivo de este artículo es evaluar el rendimiento de seis estimadores de consistencia interna a partir de un estudio de simulación. La simulación trata del alfa de Cronbach, el lambda-2, el lambda-4 y el lambda-6 de Guttman, el mayor límite inferior y el omega. Se definieron cuarenta y cinco situaciones según el tamaño de la muestra, el número de ítems y el valor de los coeficientes de saturación factorial. Los resultados sugieren que, en el caso en que el instrumento cuente con cinco ítems, el estimador a privilegiar sería el omega. En los otros casos, sería el mayor límite inferior. El alfa y el lambda-2 son sistemáticamente los dos estimadores que más subestiman el valor de la consistencia interna y deberían ser evitados. El lambda-6 sería el mejor estimador ofrecido por el programa informático SPSS. De manera general, nuestro estudio ofrece una justificación empírica para un cambio de prácticas en las investigaciones en educación.

PALABRAS CLAVE–fidelidad, consistencia interna, medida, simulación, alfa de Cronbach.

Références

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137-143.

Callender, J. C. et Osburn, H. G. (1979). An empirical comparison of coefficient alpha, Guttman's lambda-2, and MSPLIT maximized split-half reliability estimates. *Journal of educational measurement*, *16*(2), 89-99.

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational research methods*, *19*(4), 651-682.

Cho, E. et Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational research methods*, *18(*2), 207-230.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98-104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cronbach, L. J. et Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and psychological measurement*, *64(*3), 391-418.

Davenport, E. C., Davison, M. L., Liou, P.-Y. et Love, Q. U. (2015). Reliability, dimensionality, and internal consistency as defined by Cronbach: Distinct albeit related concepts. *Educational measurement: Issues and practice*, *34*(4), 1-6.

Green, S. B. et Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121-135.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177-189.

Hunt, T. D. et Bentler, P. M. (2015). Quantile lower bounds to reliability based on locally optimal splits. *Psychometrika*, *80*(1), 182-195.

Jackson, P. H. et Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567-578.

Kelley, K. et Pornprasertmanit, S. (2015). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological methods*, *21*(1), 69-92.

McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability". *Educational and psychological measurement*, *38*(1), 75-79.

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

Osburn, H. G. (2000). Coefficient Alpha and related internal consistency reliability coefficients. *Psychological methods*, *5*(3), 343-355.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate behavioral research*, *14*(1), 57-74.

Revelle, W. et Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the GLB: Comments on Sijtsma. *Psychometrika*, *74*(1), 145-154.

Schmitt, N. (1996). Uses and abuses of coefficient Alpha. *Psychological assessment*, *8*(4), 350-353.

Shapiro, A. et Ten Berge, J. M. F. (2000). The asymptotic bias of minimum trace factor analysis, with applications to the greatest lower bound to reliability. *Psychometrika*, *65*(3), 413-425.

Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.

Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, *74*(1), 169-173.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.

Streiner, D. L., Norman, G. R. et Cairney, J. (2015). *Health measurement scales. A practical guide to their development and use* (5th édition). Oxford University Press.

Tang, W. et Cui, Y. (2012, 17 avril). *A simulation study for comparing three lower bounds to reliability.* Communication présentée à l'AERA Division D: Measurement and research methodology, section 1: Educational measurement, psychometrics, and assessment (1-25).

Ten Berge, J. M. F. et Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613-625.

Ten Berge, J. M. F. et Zegers, F. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, *43*(4), 575-579.

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), 1-8.

Woodhouse, B. et Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, *42*(4), 579-591.

Correspondance

jimmy.bourque@umoncton.ca
danielle.doucet@umoncton.ca
leblanc_josee@hotmail.com
jeremie.dupuis@umoncton.ca
josee.nadeau@umoncton.ca

Contribution des auteur·e·s

Jimmy Bourque : 40 %
Danielle Doucet : 20 %
Josée LeBlanc : 20 %
Jérémie Dupuis : 10 %
Josée Nadeau : 10 %

Ce texte a été révisé par : Mélissa Singcaster

Texte reçu le : 4 septembre 2018
Version finale reçue le : 24 septembre 2019
Accepté le : 22 novembre 2019

ANNEXE : syntaxe *R*

```
rm(list = ls())

#-----------------------------------------------------
# Packages
#-----------------------------------------------------
library(mvtnorm)
library(Matrix)
library(Rcsdp)
library(psych)
library(GPArotation)
library(e1071)
library(plyr)
library(truncnorm)
#-----------------------------------------------------
# Simulation Driver
#-----------------------------------------------------

runSim <- function(replications, m, k, l, seed = NULL) {
  if (!is.null(seed)) set.seed(seed)

  Lambda.Matrix <- function(k, l) {
    Lambda <- matrix(rtruncnorm(k^2, 0, 1, l^2, 0.025), nrow=k, ncol=k)
return(Lambda)
  }

  Var.Load <- Lambda.Matrix(k, l)

  Sigma.Matrix <- function(k) {
    Sigma <- (Diagonal(k, 1) - Diagonal(k, diag(Var.Load)))
return(Sigma)
  }

  Var.Error <- Sigma.Matrix(k)

  Var.Obs <- Var.Load + Var.Error

  VCOV.MAT <- function(m, Sigma) {
    SIM.MAT <- abs(rWishart(n = 1, df = k + 1, Sigma = as.matrix(Sigma))[,,1] / (m - 1))
    return(SIM.MAT)
  }

  Reliability <- function(S) {
    Alpha <- (alpha(S)[[1]])[[1]]
    Lambda2 <- guttman(S)[[2]]
```

```
  Lambda4 <- splitHalf(S)[[1]]
  Lambda6 <- guttman(S)[[7]]
  GLB <- glb.algebraic(S)[[1]]
  Omega <- omega(S)[[4]]

  OUT <- c(Alpha, Lambda2, Lambda4, Lambda6, GLB, Omega)
names(OUT) <- c('Alpha', 'Lambda-2', 'Lambda-4',"Lambda-6', 'GLB', 'Omega')

  return(OUT)
 }

 results <- replicate(replications, {
            dat <- VCOV.MAT(m, Var.Obs)
            Reliability(dat)
           })

 PERF <- function(results, replications, l) {

  Mean.A <- mean(results[1,])
  Mean.L2 <- mean(results[2,])
  Mean.L4 <- mean(results[3,])
  Mean.L6 <- mean(results[4,])
  Mean.GLB <- mean(results[5,])
  Mean.Omega <- mean(results[6,])

  Var.A <- var(results[1,])
  Var.L2 <- var(results[2,])
  Var.L4 <- var(results[3,])
  Var.L6 <- var(results[4,])
  Var.GLB <- var(results[5,])
  Var.Omega <- var(results[6,])

  RMSE.A <- sqrt(Mean.A^2 + Var.A)
  RMSE.L2 <- sqrt(Mean.L2^2 + Var.L2)
  RMSE.L4 <- sqrt(Mean.L4^2 + Var.L4)
  RMSE.L6 <- sqrt(Mean.L6^2 + Var.L6)
  RMSE.GLB <- sqrt(Mean.GLB^2 + Var.GLB)
  RMSE.Omega <- sqrt(Mean.Omega^ 2 + Var.Omega)

  MCSE.Mean.A <- sqrt(Var.A/replications)
  MCSE.Mean.L2 <- sqrt(Var.L2/replications)
  MCSE.Mean.L4 <- sqrt(Var.L4/replications)
  MCSE.Mean.L6 <- sqrt(Var.L6/replications)
  MCSE.Mean.GLB <- sqrt(Var.GLB/replications)
  MCSE.Mean.Omega <- sqrt(Var.Omega/replications)
```

```
K.A <- kurtosis(results[1,], type = 3)
K.L2 <- kurtosis(results[2,], type = 3)
K.L4 <- kurtosis(results[3,], type = 3)
K.L6 <- kurtosis(results[4,], type = 3)
K.GLB <- kurtosis(results[5,], type = 3)
K.Omega <- kurtosis(results[6,], type = 3)

MCSE.Var.A <- Var.A * sqrt((2 + K.A)/replications)
MCSE.Var.L2 <- Var.L2 * sqrt((2 + K.L2)/replications)
MCSE.Var.L4 <- Var.L4 * sqrt((2 + K.L4)/replications)
MCSE.Var.L6 <- Var.L6 * sqrt((2 + K.L6)/replications)
MCSE.Var.GLB <- Var.GLB * sqrt((2 + K.GLB)/replications)
MCSE.Var.Omega <- Var.Omega * sqrt((2 + K.Omega)/replications)

SK.A <- skewness(results[1,], type = 3)
SK.L2 <- skewness(results[2,], type = 3)
SK.L4 <- skewness(results[3,], type = 3)
SK.L6 <- skewness(results[4,], type = 3)
SK.GLB <- skewness(results[5,], type = 3)
SK.Omega <- skewness(results[6,], type = 3)

MCSE.RMSE.A <- sqrt((1/replications) * (((Var.A^2) * (K.A - 1)) +
                              4 * Var.A^(3/2) * SK.A * Mean.A +
                              4 * Var.A * Mean.A^2))
MCSE.RMSE.L2 <- sqrt((1/replications) * (((Var.L2^2) * (K.L2 - 1)) +
                               4 * Var.L2^(3/2) * SK.L2 * Mean.L2 +
                               4 * Var.L2 * Mean.L2^2))
MCSE.RMSE.L4 <- sqrt((1/replications) * (((Var.L4^2) * (K.L4 - 1)) +
                               4 * Var.L4^(3/2) * SK.L4 * Mean.L4 +
                               4 * Var.L4 * Mean.L4^2))
MCSE.RMSE.L6 <- sqrt((1/replications) * (((Var.L6^2) * (K.L6 - 1)) +
                               4 * Var.L6^(3/2) * SK.L6 * Mean.L6 +
                               4 * Var.L6 * Mean.L6^2))
MCSE.RMSE.GLB <- sqrt((1/replications) * (((Var.GLB^2) * (K.GLB - 1)) +
                                4 * Var.GLB^(3/2) * SK.GLB * Mean.GLB +
                                4 * Var.GLB * Mean.GLB^2))
MCSE.RMSE.Omega <- sqrt((1/replications) * (((Var.Omega^2) * (K.Omega - 1)) +
                                4 * Var.Omega^(3/2) * SK.Omega * Mean.Omega +
                                4 * Var.Omega * Mean.Omega^2))

MeanR <- c(Mean.A, Mean.L2, Mean.L4, Mean.L6, Mean.GLB, Mean.Omega)
VarR <- c(Var.A, Var.L2, Var.L4, Var.L6, Var.GLB, Var.Omega)
RMSER <- c(RMSE.A, RMSE.L2, RMSE.L4, RMSE.L6, RMSE.GLB, RMSE.Omega)
MCSE.MeanR <- c(MCSE.Mean.A, MCSE.Mean.L2, MCSE.Mean.L4, MCSE.Mean.L6,
      MCSE.Mean.GLB, MCSE.Mean.Omega)
```

```
   MCSE.VarR <- c(MCSE.Var.A, MCSE.Var.L2, MCSE.Var.L4, MCSE.Var.L6,
        MCSE.Var.GLB, MCSE.Var.Omega)
   MCSE.RMSER <- c(MCSE.RMSE.A, MCSE.RMSE.L2, MCSE.RMSE.L4,
        MCSE.RMSE.L6, MCSE.RMSE.GLB, MCSE.RMSE.Omega)

   Ind.Perf <- rbind(MeanR, VarR, RMSER, MCSE.MeanR, MCSE.VarR, MCSE.RMSER)
   rownames(Ind.Perf) <- c('Mean',"Variance","RMSE","SE Mean","SE Var.", "SE RMSE")
   colnames(Ind.Perf) <- c('Alpha', 'Lambda-2', 'Lambda-4', 'Lambda-6', 'GLB', 'Omega')
return(Ind.Perf)
  }

  PERF(results, replications, l)
}

#------------------------------------
# Experimental Design
#------------------------------------
source_obj <- ls()

m <- c(50, 100, 200, 500, 1000)

k <- c(5, 10, 20)

l <- c(0.350, 0.500, 0.800)

design_factors <- list(m = m, k = k, l = l) # combine into a design set
params <- expand.grid(design_factors)
params$replications <- 2000
params$seed <- round(runif(nrow(params)) * 2^30)

#-----------------------------------------------------
# run simulations in serial
#-----------------------------------------------------

FINAL <- mdply(params, .fun = runSim)

FINAL$Statistic <- rep(c('Mean',"Variance","RMSE","SE Mean","SE Var.', 'SE RMSE'),
        times=45)

FINAL

save(FINAL, file = 'SimulationResults.Rdata')
```