Document généré le 25 avr. 2024 03:21

Mesure et évaluation en éducation

Soixante ans de bons et mauvais usages du alpha de Cronbach

Dany Laveault

Volume 35, numéro 2, 2012

Date de réception : 18 juin 2012
Date de réception de la version finale :14 août 2012
Date d'acceptation :14 août 2012

URI : https://id.erudit.org/iderudit/1024716ar
DOI : https://doi.org/10.7202/1024716ar

Aller au sommaire du numéro

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)
2368-2000 (numérique)

Découvrir la revue

Citer cet article

Laveault, D. (2012). Soixante ans de bons et mauvais usages du alpha de Cronbach. *Mesure et évaluation en éducation*, *35*(2), 1–7. https://doi.org/10.7202/1024716ar

Résumé de l'article

Le coefficient alpha de Cronbach est sans doute la mesure de fidélité la plus utilisée, mais aussi la plus abusée de toutes. Quoiqu'elle puisse rendre de précieux services aux concepteurs d'outils d'évaluation, un certain nombre de précautions sont nécessaires pour tenir compte de ses limites et en assurer le bon usage.

Tous droits réservés © ADMEE-Canada - Université Laval, 2012

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Soixante ans de bons et mauvais usages du alpha de Cronbach

Dany Laveault
Université d'Ottawa

Mots clés : alpha de Cronbach, fidélité, théorie classique des scores

Le coefficient alpha de Cronbach est sans doute la mesure de fidélité la plus utilisée, mais aussi la plus abusée de toutes. Quoiqu'elle puisse rendre de précieux services aux concepteurs d'outils d'évaluation, un certain nombre de précautions sont nécessaires pour tenir compte de ses limites et en assurer le bon usage.

Key words: Cronbach's alpha, reliability, classical test score theory

Cronbach's coefficient alpha is probably the most frequently reported measure of reliability and most certainly the most abused one. Though it can still be of tremendous help to test developers, its proper use must take into account a certain number of necessary precautions.

Palavras-chave: alfa de Cronbach, fidelidade, teoria clássica dos resultados

O coeficiente alfa de Cronbach é sem dúvida a medida de fidelidade mais utilizada, mas também a mais sujeita a abusos. Ainda que seja muito útil aos que concebem instrumentos de avaliação, um certo número de precauções são necessárias para ter em conta os seus limites e assegurar o seu bom uso.

Note de l'auteur – Toute correspondance peut être envoyée à Dany Laveault, Faculté d'éducation, Université d'Ottawa, Pavillon Lamoureux, bureau 474, 145, rue Jean-Jacques-Lussier, Ottawa, Ontario K1N 6N5, téléphone (613) 562-5800, poste 4118, télécopieur (613) 562-5146, ou par courriel à l'adresse suivante : [dlaveault@videotron.ca].

Problématique du coefficient alpha

Le coefficient alpha est sans doute l'une des mesures les plus répandues de la fidélité. C'est aussi l'une de celles qui – du fait de sa popularité et de sa facilité d'utilisation – est l'une des moins bien utilisée. C'est à cause des abus, tant dans l'utilisation que dans l'interprétation du coefficient alpha, qu'il s'avère nécessaire de célébrer avec prudence ses soixante ans d'existence (Cronbach, 1951).

Élaboré en 1951 par Lee J. Cronbach, ce coefficient « alpha » a d'abord été conçu, comme son nom l'indique, comme la première d'une série de mesures de calculs des différentes propriétés des scores (Cronbach & Shavelson, 2004). Cette série d'indicateurs numériques ne devait avoir aucun lendemain, mais la popularité du coefficient alpha pour mesurer la fidélité ne s'est pas démentie depuis. En raison de sa popularité, de sa longévité et aussi du nombre élevé de références secondaires à l'article de Cronbach (1951), le nombre d'utilisations inappropriées du coefficient s'est vite accru, au point où plusieurs auteurs (Cortina, 1993 ; Green, Lissitz, & Mulaik, 1977 ; Schmitt, 1996) et Cronbach lui-même (2004) ont cru nécessaire de rappeler et de préciser les limites et les conditions de son utilisation appropriée.

Avec le recul des années, plus de soixante ans après la parution de l'article de Cronbach (1951), le coefficient alpha doit être perçu pour ce qu'il est vraiment, soit un coefficient de cohérence interne des items qui composent le score total à un test. Ce coefficient trouve maintenant sa place parmi un arsenal complexe d'analyses de la fidélité, dont l'étude de la généralisabilité constitue le modèle sans doute le plus abouti en théorie classique des scores (Cardinet, Johnson, & Pini, 2010). Bref, il est important de retenir que le coefficient alpha ne couvre qu'une faible proportion des besoins de calculs de la fidélité et que son utilisation doit être limitée aux cas bien précis où il convient.

Conditions d'utilisation

Le coefficient alpha est une mesure de la cohérence interne, c'est-à-dire du degré d'intercorrélation entre les items. La cohérence interne est une condition nécessaire, mais non suffisante de l'homogénéité des items d'une échelle, cette dernière impliquant l'unidimensionalité de l'échelle en question (Green et al., 1977). Il est en effet possible d'obtenir une valeur élevée du

coefficient alpha avec un ensemble d'items multidimensionnels lorsque le nombre d'items en lui-même est suffisamment grand ou encore si plusieurs des dimensions sont intercorrélées.

Le coefficient alpha est fonction de la corrélation moyenne entre les items (Cortina, 1993) et peut être élevé même s'il existe de grands écarts entre les corrélations. L'erreur type du alpha doit donc être prise en considération afin de déterminer jusqu'à quel point la valeur du alpha peut avoir été affectée par la variance des corrélations entre items. Une erreur type du alpha élevée peut être interprétée de deux manières possibles. Elle peut indiquer soit une erreur élevée de mesure des items, soit la présence de plusieurs dimensions au sein des items (Schmitt, 1996). Lorsque les items ne peuvent être considérés comme tau-équivalents, la fidélité de cohérence interne ne peut être estimée avec exactitude et représente plutôt la borne inférieure de la valeur réelle de fidélité. Pour que le coefficient alpha tende vers sa valeur maximale possible, les items doivent être tau-équivalents, c'est-à-dire être en fortes corrélations et ne différer entre eux que par une constante (Cortina, 1993).

Cortina (1993) a démontré que le coefficient alpha est fortement influencé par le nombre d'items. Lorsque le nombre d'items est suffisamment élevé (de l'ordre de 40 items ou plus), il est relativement facile d'obtenir des valeurs acceptables du alpha (0,70 et plus) en dépit d'une faible moyenne des corrélations entre items ou en présence de multidimentisionalité. À la limite, un constructeur de tests peut atteindre des valeurs du alpha respectables en combinant plusieurs ensembles d'items hétérogènes pourvu que ceux-ci soient corrélés positivement, ne serait-ce que modérément. En combinant sans autres raisons valables des ensembles plus ou moins disparates d'items, il est donc possible, lorsque le nombre d'items est assez élevé, d'obtenir un alpha indiquant une cohérence interne suffisante. Cependant, la signification de cet alpha soulève de nombreuses difficultés à cause de l'absence d'homogénéité entre les items. Un tel alpha indiquerait simplement que le test mesure quelque chose de manière cohérente, mais ce « quelque chose » serait indéfini ou mal défini. Une cohérence élevée obtenue avec un grand nombre d'items ne dispense donc aucunement d'effectuer une validation théorique et empirique de la dimensionnalité du concept mesuré. C'est là l'une des raisons, selon Cortina (1993), de la mauvaise utilisation du alpha, car au-delà d'un certain nombre d'items, la fidélité du test devient une fonction trop importante du nombre d'items. Or « le nombre d'items n'est pas une mesure adéquate de la qualité d'une échelle ou d'un test » (p. 101).

Une autre mauvaise utilisation du coefficient alpha provient d'une interprétation erronée de la limite précédente (Schmitt, 1996). C'est ainsi qu'un faux argument veut qu'il soit possible de se satisfaire de valeurs moindres du alpha avec un petit nombre d'items. La pratique d'indiquer le nombre d'items sur lequel est basé le calcul du alpha est certes une bonne pratique, mais elle ne peut être employée à contre-sens, c'est-à-dire pour se contenter de valeurs de fidélité faibles avec un petit nombre d'items, valeurs qui seraient jugées inacceptables avec un nombre plus élevé. Il demeure que le coefficient alpha est une estimation de l'erreur d'échantillonnage des items et que, peu importe le nombre d'items ayant servi au calcul du score total, une faible valeur du alpha est toujours indicative d'une mesure imprécise du score total.

Une autre interprétation fréquente du coefficient alpha veut que celui-ci représente la moyenne de tous les coefficients de bissection possibles. Or, cette affirmation ne tient que si les items sont tau-équivalents. Le coefficient alpha est la moyenne de tous les coefficients de bissection, si et seulement si les écarts types des items sont égaux. Le coefficient alpha calculé sur les scores standardisés permet de calculer la valeur standardisée du alpha. Ce coefficient standardisé n'est cependant pas approprié pour décrire la fidélité et l'erreur de mesure lorsque les scores bruts sont employés. Il ne devrait pas être calculé avec les valeurs brutes car il pourrait surestimer la cohérence interne de l'ensemble d'items en question.

Limites du coefficient alpha

Selon Cortina (1993), le coefficient alpha est utile pour estimer la fidélité dans un cas bien particulier, soit lorsque nous nous soucions de la variance unique des items dans un test unidimensionnel. Lorsque le coefficient alpha est élevé et que sa valeur n'est pas gonflée par un nombre élevé d'items, nous pouvons en conclure que peu de variance entre les scores résulte d'items particuliers, mais dépend d'un facteur général ou de groupe. Ceci étant dit, le coefficient alpha ne devrait jamais être utilisé comme un indicateur d'unidimensionalité des résultats. Le coefficient alpha ne dispense pas d'une étude appropriée de la dimensionalité des items au moyen d'analyses factorielles exploratoire et confirmatoire. Ceci étant dit, « une fois que l'existence d'un facteur unique a pu être démontrée, le coefficient alpha peut être considéré comme une mesure de la force d'une dimension unique » (Cortina, 1993, p. 103).

Exemples de bonnes pratiques

Si les études et analyses récentes sur le coefficient alpha ont permis de mieux en saisir la portée et de dénoncer les abus quant à son utilisation, elles ont également permis de mieux en encadrer l'utilisation et de formuler des recommandations quant à la manière la plus adéquate d'en rapporter les résultats :

1. La valeur du alpha devrait toujours être accompagnée du nombre d'items sur lequel il a été calculé, de son erreur d'estimation ainsi que de la corrélation moyenne entre les items (Cortina, 1993; Green et al., 1977 ; Schmitt, 1996). Ceci permet au lecteur d'apprécier à quel point la cohérence interne résulte de corrélations élevées entre les items et non d'une inflation de la valeur du alpha occasionnée par un nombre élevé d'items peu corrélés entre eux.
2. Cronbach et Shavelson (2004) suggèrent de rapporter l'erreur de mesure du score en plus du coefficient lui-même. Ceci permet au lecteur de mieux se représenter toute l'étendue des scores possibles avec un échantillon d'items tirés de la même population. Une telle pratique est également conforme avec le standard 2.1 des *Standards for educational and psychological testing* (AERA, APA, & NCME, 2000).
3. Cronbach et Shavelson (2004) proposent d'indiquer le nombre d'items non complétés, surtout si ceux-ci ont été notés 0, introduisant ainsi un biais artificiel dans la corrélation entre les items à la fin d'un test. Dans un tel cas, il peut être approprié de ne pas tenir compte de ces items dans le calcul de la valeur du alpha.
4. L'hétérogénéité dans le contenu des items peut remettre en question la nature aléatoire de l'échantillonnage du contenu. Cronbach et Shavelson (2004) préviennent qu'une analyse de cohérence interne qui ne tient pas compte de l'existence de plusieurs catégories d'items peut donner lieu à des erreurs de mesure plus grandes que ne l'aurait fait une analyse plus spécifique, par exemple, en calculant la fidélité de chaque catégorie.
5. Le coefficient alpha constitue une estimation appropriée de la cohérence interne dans l'éventualité où le score total est utilisé de façon relative, c'est-à-dire pour différencier les individus entre eux. Dans le cas où le score total au test est utilisé de façon absolue pour déterminer, par exemple, si un individu a atteint un seuil de réussite ou tout autre score particulier, alors l'erreur de mesure est sous-estimée. Selon l'usage que

l'on compte faire du score total, Cronbach et Shavelson (2004) recommandent de fournir les composantes de sa variance de manière à permettre le calcul des valeurs d'erreur relative et d'erreur absolue.

6. Il est également possible de mieux informer l'utilisateur potentiel d'un test en lui indiquant le nombre de conditions (par ex. : le nombre d'items) requis pour atteindre une erreur de mesure acceptable. Cronbach et Shavelson (2004) recommandent de présenter un tableau indiquant l'erreur de mesure en fonction d'un certain nombre de conditions. L'utilisateur est alors en meilleure position de faire un choix et d'arbitrer entre coût du test et précision de la mesure.

Conclusion

Pour conclure, nous pouvons affirmer que le coefficient alpha demeure une mesure de fidélité pratique qui mérite d'être employée. Avec le recul des années, des méthodes plus puissantes, mais aussi plus complexes, se sont ajoutées au coefficient alpha et, lorsque les conditions le permettent, celles-ci devraient être employées. C'est le cas notamment de la théorie de la généralisabilité (Cardinet et al., 2010) en théorie classique des scores. Il n'en demeure pas moins que lorsque nous avons besoin de connaître rapidement s'il est justifié de créer un score total à partir de la somme des items d'un score composite, le coefficient alpha permet d'estimer rapidement les chances qu'un autre échantillon d'items tirés du même univers d'items nous permette d'amener des conclusions similaires quant à la valeur relative des résultats, par exemple, quel score est le plus élevé ou le plus bas, sans que l'on puisse se prononcer toutefois quant à sa valeur absolue. Ce besoin est courant et, dans de telles circonstances, le coefficient alpha continue à rendre des services inestimables, tant aux étudiants en psychométrie et en évaluation scolaire qu'aux chercheurs les plus chevronnés.

RÉFÉRENCES

AERA, APA, & NCME (2000). (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying Generalizability Theory using EduG*. New York: Taylor & Francis.

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, 12*, 1-16.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391-418.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37*, 827-838.

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

Date de réception : 18 juin 2012

Date de réception de la version finale : 14 août 2012

Date d'acceptation : 14 août 2012