



SOBRE EL COEFICIENTE ALPHA DE CRONBACH Y SU INTERPRETACIÓN EN LA EVALUACIÓN EDUCATIVA

Michaël Canu

**Universidad El Bosque
Bogotá, Colombia**

Mauricio Duque

**Universidad de Los Andes
Bogotá, Colombia**

Resumen

El coeficiente alpha de Cronbach, propuesto desde 1951 como medida de la "reliability" o de la coherencia interna de una prueba, fue desarrollado en psicometría a partir del trabajo de Kuder y Richardson sobre las correlaciones inter-items. Este indicador tiene todavía una importancia muy grande en el análisis de los resultados de evaluaciones psicotécnicas y a menudo en educación. Sin embargo, varios estudios han resaltado las restricciones de este indicador. De un lado, como lo han escrito Kuder y Richardson en su artículo original, es una estimación del error aleatorio en la varianza de un resultado de prueba, lo que surge que se consideran solamente los errores de medición aleatorios. De otro lado, este coeficiente tiene una sensibilidad bastante importante con respecto a los "outliers", es decir los valores (respuestas) aberrantes, es decir los sujetos que tienen una puntuación lejos del promedio. De hecho, se puede considerar que existe una diferencia de fondo bastante importante entre el comportamiento esperado del coeficiente en los resultados de pruebas psicotécnicas y en los de pruebas en educación: se considera en el primer caso que la coherencia del sujeto (él que contesta la prueba) es alta y se puede entonces medir la coherencia de la prueba a través de sus respuestas mientras en el segundo caso, la coherencia del sujeto puede hacer parte de lo que se evalúa. Entonces, ¿Cómo interpretar entonces el resultado de este tipo de coeficiente si dos fuentes de error con características desconocidas se combinan en la aplicación de una sola prueba? A partir del estudio de algunas pruebas conceptuales en educación en ingeniería, y de comparaciones con

pruebas psicotécnicas, este artículo propone aportar elementos de comprensión sobre el coeficiente Alpha, su significado, y argumentos acerca del comportamiento de esta medición de coherencia en casos extremos. Nuevos indicadores diseñados a partir del alpha permiten por ejemplo caracterizar la población estudiada en término de coherencia interna relativo al contenido de la prueba.

Palabras clave: evaluación; alpha de cronbach; confiabilidad interna; coherencia; prueba conceptual

Abstract

The Cronbach's alpha coefficient, proposed from 1951 like measure of the "reliability" or internal consistency of a test, was developed in psychometrics from the work of Kuder and Richardson on the correlations inter-items. This indicator has still a very big importance in the analysis of the results of psycho-technical or educational assessment. Nevertheless, several studies have highlighted the restrictions of this indicator. On the one hand, as Kuder and Richardson notes themselves in their original article, it is an estimation of the random error in the variance of a result of test, which will arise that are only the random mistakes of measurement. On the other hand, this coefficient has an important enough sensibility according to the "outliers", the aberrant values, the subjects that have a punctuation far from the average. In fact, it is possible to consider to be that it exists a fundamental difference between the behavior expected from the coefficient in the results of educational tests and psycho-technical tests: it can be said that in the first case the coherence of the subject that answers is pretty high and one can measure the coherence of the test across his answers, while in the second case, the coherence of the subject is part of what is evaluated. So, how interprets the result of this type of coefficient if two sources of mistake with unknown characteristics combine in the application of a unique test? From the study of some conceptual assessments in engineering education, and of comparisons with response tests, this article proposes to give elements of comprehension on the coefficient Alpha, its meaning, and arguments it brings over of the behavior of this measurement of coherence in extreme cases. New indicators designed from alpha allow to characterize for example the population studied in terms of internal coherence relative to the content of the test.

Keywords: assessment; coefficient alpha; reliability; internal consistency; concept inventory

1. Introducción

El coeficiente Alpha de Cronbach, propuesto desde 1951 como medida de la confiabilidad interna (internal consistency or reliability) de una prueba, fue desarrollado en psicometría a partir del trabajo de Kuder y Richardson (1937) sobre las correlaciones inter-items, que fue derivado de las técnicas de partición de Spearman (Half-split technics, 1910).

La fórmula del Alpha de Cronbach se presenta así:

Donde k es el número de ítems, S^2 es la varianza total de la prueba cuadrada y S_i^2 es la varianza cuadrada del ítem i .

De hecho, el coeficiente Alpha de Cronbach es una generalización de la fórmula 20 de Kuder y Richardson (K-R 20, Kuder y Richardson, p.158) para ítems con evaluación numérica no dicotómica (es decir diferente de 0/1).

¿Que mide este coeficiente?

La idea de la evaluación es utilizar una herramienta (la prueba) que permite atribuir un número (la nota) a cada persona para evaluar su desempeño. Intuitivamente, se entiende que dos pruebas diferentes que miden la misma información (conocimiento, opinión, etc.) no van a dar exactamente el mismo resultado. Pero también, aunque el instrumento ideal de medición debería dar siempre el mismo puntaje para la misma persona que presenta esta misma prueba, en la realidad, el resultado va a variar. Esta variación es debida al error de medición. Esta auto-correlación (self-correlation) de la prueba con ella misma es una medición de su consistencia interna (internal consistency). Aunque como lo recuerda Cronbach (2004, p.395), no hubo demostración formal de la equivalencia entre confiabilidad (reliability) y consistencia interna, la confiabilidad es considerada como una estimación de esta consistencia interna de la prueba a través de varias aplicaciones (Webb et al., 2006, Ch. 4, p. 1).

2. Interpretación del Alpha

Interpretación y utilización "clásica":

El Alpha se utiliza entonces principalmente para dar una estimación de la consistencia interna de una prueba. Es como lo decimos, una estimación del porcentaje de la varianza total de la prueba que se puede considerar como no debida al error de medición (es decir la "verdadera" varianza; Kuder y Richardson, 1937).

Como se observa en la fórmula del Alpha, el coeficiente depende de tres parámetros: la varianza total de la prueba, la suma de las varianzas de cada ítem y el número de ítems. Son parámetros no independientes en general, y el coeficiente varía entre 0 y 1.

Si todos los ítems son totalmente independientes cada uno de los otros, el coeficiente vale 0 ya que la suma de las varianzas es igual la varianza de la suma en este caso. Se puede interpretar como una incoherencia máxima de la prueba: cada ítem evalúa un aspecto diferente de los otros.

Al contrario, si hay una dependencia máxima entre todos los ítems, por ejemplo, si cada ítem es el mismo que los otros, la varianza total es entonces igual al número de ítems por la varianza de cada ítem (ya que son iguales) y la suma de las varianzas de cada ítem es igual al producto del número de ítems por la varianza de cada uno. El Alpha vale entonces 1. Sería un caso ideal, límite, para el cual cada ítem de la prueba mide

exactamente el mismo aspecto, lo que no ocurre en la realidad y de hecho no es deseable.

En general, se considera que un Alpha más grande que 0,7 indica una buena confiabilidad de la prueba. Pero en muchos casos, un Alpha entre 0,5 y 0,7 se considera como aceptable en educación (Rodríguez, 2013).

El número de ítems no es tampoco un parámetro independiente de los otros y en general, no se puede considerar que, si se aumenta o disminuye el número de ítems, aumenta o disminuye automáticamente el Alpha. Ello viene del hecho que cada ítem adicional o cancelado tiene un impacto en término de varianza (total y de cada ítem).

La utilización principal del coeficiente Alpha tiene entonces que ver con la confiabilidad de la prueba, es decir con la estructura misma de la prueba: el constructo. Permite saber qué tanto los ítems miden los mismos aspectos del constructo de la prueba. De hecho, se puede calcular un coeficiente dependiendo del alpha para saber si un ítem particular tiene un efecto positivo o negativo sobre la confiabilidad de la prueba global: vamos a llamarlo *CEA*: Cambio En Alpha si se quita el ítem *i*.

Este indicador tiene una alta correlación - en general - con la *discriminación* del mismo ítem, es decir con la capacidad de este ítem de discriminar entre los estudiantes que tienen un alto puntaje y los que tienen un bajo puntaje. Se puede utilizar para saber si un ítem particular induce una respuesta, en promedio, debida al azar. Existen varios métodos para calcular la discriminación, uno de los más utilizados es el índice de discriminación de Kelley (utilizada por ejemplo en el Learning Management Systems Blackboard). Vamos a utilizarlo en su versión "a 27%" (Ebel, 1954) para comparar la discriminación de los ítems de nuestras pruebas con el CAE.

Como lo vimos, la varianza de cada ítem depende del ítem, es decir da informaciones sobre el impacto de un ítem dado al interior de la prueba, y permite entonces dar elementos de decisión con respecto a un ítem (eliminarlo, por ejemplo).

Interpretación y utilización alternativa:

Sin embargo, la aplicación y la interpretación del Alpha requiere que varias condiciones se verifiquen. En su artículo, Kuder y Richardson precisan en efecto que la confiabilidad de la prueba permite evaluar el *error aleatorio* de medición. En psicometría, se puede considerar que existe un error de medición porque lo que conduce el sujeto a contestar de una manera u otra a un ítem particular o un grupo de ítems es básicamente inconsciente: se supone que frente a dos ítems de forma diferente que buscan la misma respuesta, el sujeto va efectivamente y de manera automática contestar de igual manera: el proceso interno que conduce a la respuesta será el mismo en los dos casos. El único error de medición es entonces de tipo aleatorio. En educación, buena parte de las evaluaciones intentan justamente evaluar de alguna manera este proceso interno que conduce a la respuesta dada.

Entonces, al error de medición (aleatorio), se superpone un efecto que no tienen que ver con el azar y que depende de la capacidad del sujeto para hacer el vínculo adecuado entre varias formas de un mismo ítem apuntando a un único aspecto del constructo.

En este caso, el Alpha va a dar una información que no solamente depende de la coherencia de los ítems (confiabilidad de la prueba) sino también de la "coherencia del estudiante". Podríamos decir de la "confiabilidad interna del estudiante", es decir de su capacidad a contestar de la misma manera a ítems diferentes que apuntan al mismo conocimiento.

Es importante anotar que se puede considerar que este efecto podría cancelarse con la aplicación de la misma prueba a un número importante de sujetos, o aplicando esta prueba a un grupo de expertos en el tema para los cuales no hay duda sobre sus capacidades para contestar de manera coherente los diferentes ítems (aunque presentados de forma diferente).

Sin embargo, cuando uno quiere desarrollar una prueba a escala de una facultad, por ejemplo, no se pueden conseguir miles de estudiantes para aplicar la misma prueba. En este caso muy común, se puede considerar que el Alpha va a dar indicaciones tanto de la coherencia interna de la prueba como la de los estudiantes. Eso es lo que vamos a ilustrar en el resto del artículo.

3. Cambio En Alpha Estudiantil

Como el Alpha depende de las respuestas de los estudiantes, se puede examinar el impacto de cada estudiante en el Alpha de la misma manera en que se utiliza para los ítems. Eso nos conduce a calcular para cada estudiante el coeficiente Alpha global de la prueba sin él. El efecto obtenido puede ser positivo o negativo, según si la respuesta del estudiante a todos los ítems va a bajar o subir el Alpha. Este indicador, que vamos a llamar CEAE (Cambio En Alpha Estudiantil) nos permite obtener, por ejemplo, una representación de la repartición de los estudiantes en término de nota con respecto a su impacto sobre el Alpha. Es decir, ¿qué parte de los estudiantes tienen un impacto negativo sobre el Alpha global de la prueba? Esto puede servir a identificar una subpoblación de estudiantes que tiene un comportamiento particular (en término de coherencia) con respecto a la prueba.

4. Metodología

Se utilizó una prueba conceptual sobre los circuitos eléctricos (Electric Circuit Concept Inventory, ver parágrafo siguiente) asociada a un curso de circuitos eléctricos de 5to semestre de una universidad colombiana para conseguir los datos necesarios para este trabajo. Durante dos semestres del mismo curso (primero y segundo semestre de 2016) con dos secciones diferentes y dos profesores colectando los datos.

Una prueba conceptual equivalente se aplicó dos veces cada semestre en un esquema Pre-test/Post-test, al inicio y al final del curso. 66 estudiantes (40 en la sección 1 y 26 en la sección 2) participaron el primero semestre de 2016 y 72 estudiantes en el segundo semestre (36 en la sección 1 y 36 en la sección 2).

La primera prueba conceptual (primero semestre) tenía 22 ítems y la segunda (segundo semestre) tenía 29 ítems.

5. El "Electric Circuit Concept Inventory"

Una característica fundamental del ingeniero experto en una área específica es la capacidad de utilizar conceptos disciplinarios para resolver problemas (Pellegrino, Di Bello, & Brophy, 2013). Sin embargo, a menudo los estudiantes no son capaces de generalizar sus conocimientos a otros problemas y contextos que difieren de los que les permitieron construir dicho conocimiento. Desde hace tres décadas se vienen desarrollando diferentes pruebas de nivel conceptual para evaluar tanto los estudiantes de ingeniería que las metodologías de enseñanza utilizadas. Los *Concept Inventories* (CI) dan un buen ejemplo de la utilización de modelos conceptuales sobre la comprensión de los estudiantes para generar un conjunto de preguntas de manera sistemática (a partir de elementos cognitivos y de la manera de observar el razonamiento utilizado). El CI consiste en preguntas de respuestas múltiples en las cuales los distractores (proposición de respuestas incorrectas) fueron desarrollados primero para preguntas abiertas y analizar las respuestas dadas. Las preguntas se basan en la investigación en educación incluyendo investigaciones sobre las concepciones (erróneas o alternativas) de los estudiantes. De hecho, los distractores son en general representativos de las concepciones erróneas de los estudiantes y no son escogidas al azar como sucede con frecuencia en las evaluaciones de respuestas múltiples clásicas. Por eso, los CI están muy enfocados en un concepto particular y preciso de un área disciplinar como el concepto de fuerza, de leyes del movimiento, de mecánica estática, o dinámica, de electricidad, etc.

Varios autores desarrollaron CI en temas de electricidad como (Rahman & Ogunfunmi, 2010) o (Engelhardt Paula & Beichner, 2004) que concibieron una prueba conceptual sobre el tema de fundamentos de circuitos eléctricos llamada DIRECT (Determining and Interpreting Resistive Electric circuits Concepts Test). Este CI fue utilizado por ejemplo por (Getty, 2009) para evaluar una intervención didáctica basada en indagación con estudiantes de universidad (evaluación formativa y sumativa). Las preguntas del ECI de Engelhardt and Beichner con menos correlación fueron excluidas de la prueba para disminuir el tamaño de esta, pasando de 29 a 22 ítems.

Ninguna de las situaciones de la prueba se utilizó en los ejercicios de clase. Este CI que contiene entonces las 22 preguntas de mejor correlación del ECI original de Engelhardt Paula and Beichner (2004) mide cinco aspectos de fundamentos de electricidad: el aspecto físico de los circuitos (esquemas), los circuitos equivalentes, la energía y la potencia, la corriente y la conservación de la corriente (Ley de Corriente de Kirchhoff, y

el voltaje y la ley de ohm (Ley de Voltaje de Kirchhoff). Para mayor información sobre los análisis de estas pruebas, consultar Canu & Duque (2016).

6. Resultados y discusión

Como anunciado en los párrafos precedentes, el Alpha puede ser utilizado para calcular la discriminación de los diferentes ítems, lo que evita la utilización de otras herramientas. La figura 1 muestra la alta correlación (s de spearman = 0.91 en los dos pre-tests) que existe entre el CAE y el índice de discriminación de Kelley. Eso permite justificar la utilización de este coeficiente para evaluar la discriminación de los ítems.

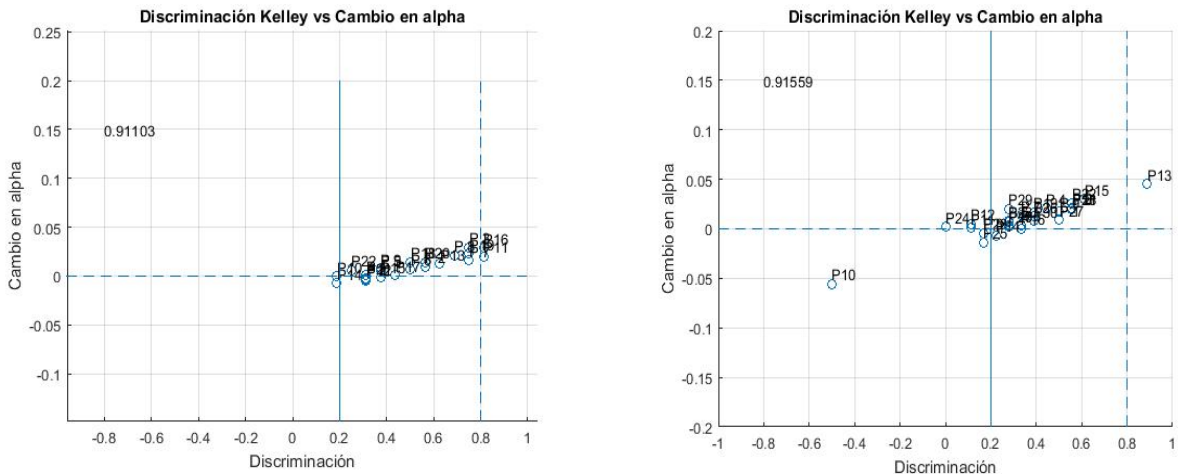


Figura 1: Correlación entre discriminación clásica y utilizando el Alpha (CAE)

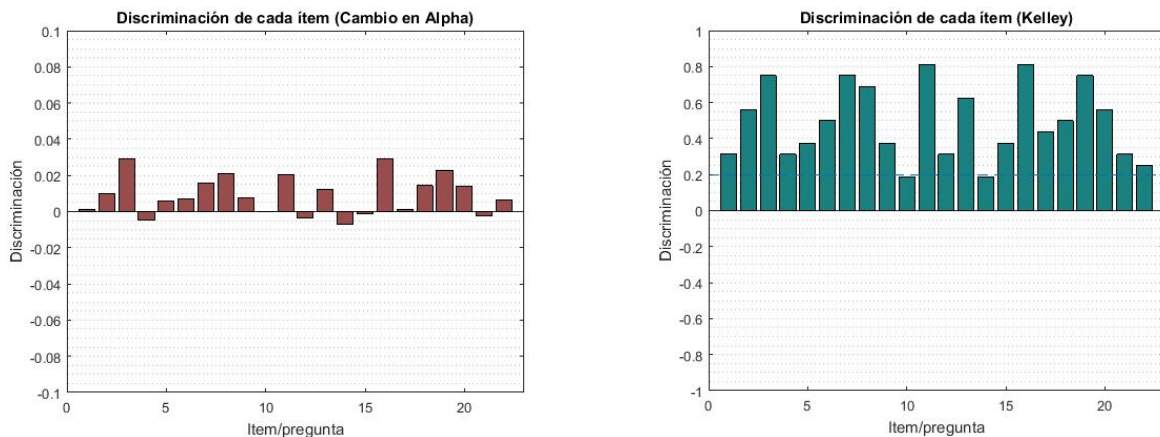


Figura 2: Discriminación de los ítems con el CAE y el de Kelley (pretest 2016-10).

La figura 2 muestra que utilizando el CAE los ítems poco discriminatorios aparecen en la parte negativa, lo que permite identificarlos de manera más directa. Al contrario del índice de Kelley para el cual se puede discutir el valor límite a partir del cual el ítem tiene baja discriminación, un CAE negativo indica una baja discriminación. Eso viene del hecho que los ítems que bajan el Alpha son los que inducen respuestas, en promedio, cerca de respuestas "al azar". Es decir que los estudiantes con alto o bajo puntaje

contestan de una manera similar mientras se esperaría que sólo los estudiantes con alto puntaje contestaran correctamente a este ítem.

El resultado más interesante proviene de la utilización del Alpha para estudiar el comportamiento de los estudiantes. La figura 3 muestra el cambio en el alpha estudiantil correlacionado con la nota de los estudiantes para el pre y el post test de los dos semestres.

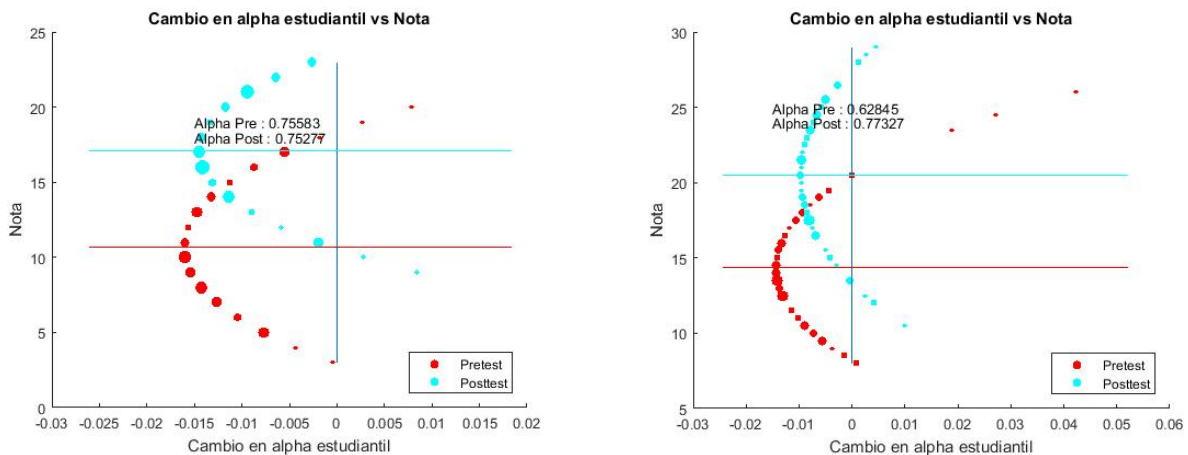


Figura 3: Correlación entre CEAE y Nota para el semestre 2016-10 (iz.) y 2016-20 (der.)

Se puede observar (curvas paraboloides) que el promedio de las notas sube entre el pre y el post test durante los dos semestres, lo que permite decir que el conocimiento conceptual de los estudiantes aumenta entre el inicio y el final del curso (¡lo que esperan todos los profesores!). Sin embargo, mientras que el Alpha de la prueba sube de 0,63 a 0,77 entre el pre y el post test del segundo semestre, no hay cambio significativo durante el primer semestre (0,76 -> 0,75). Es importante anotar que esta misma prueba conceptual (se quitaron 7 preguntas de baja discriminación de la prueba original) tenía un Alpha de 0,70-0,71 en el artículo de Engelhardt and Beichner.

¡Además, durante el segundo semestre el Alpha de 0,63 se puede descomponer entre una parte de los estudiantes (la sección 1) que obtiene un Alpha global de 0,68 y otra con un Alpha global de 0,43! El algoritmo realizado en Matlab con el CEAE, nos permite detectar de manera automática esta subpoblación de la muestra total que baja el Alpha de más de 2%.

Las gráficas precedentes nos muestran también un efecto bastante previsible: son los estudiantes del promedio de la muestra que tienen un CEAE negativo (esto es que bajan el Alpha), es decir los que tienen posiblemente la más baja coherencia con respecto a los conceptos de la prueba. Al contrario, los “buenos” y los “malos” tienen mayor coherencia.

7. Conclusiones y perspectivas

Aunque utilizado de manera amplia, el coeficiente Alpha de Cronbach tiene un comportamiento bastante complejo. El pasaje de mediciones psicométricas a

mediciones en educación cambia una hipótesis fundamental del Alpha: se asume que la diferencia entre el verdadero puntaje del sujeto (true score) y el puntaje medido es el error de medición y que este error es un error aleatorio (Cronbach, 2004). En evaluación de tipo conceptual, se puede observar que interviene un efecto que viene de la coherencia del estudiante con respecto a la prueba (a los conceptos evaluados, el constructo) y que este efecto no se puede descuidar con muestras pequeñas de estudiantes. Además, este efecto se puede utilizar para extraer informaciones sobre la muestra de estudiantes y en particular sobre la presencia de subpoblaciones con un comportamiento coherente o incoherente. Falta extraer interpretaciones precisas del CEAE y en particular un índice integrado que permite tomar decisiones en término de diseño de la prueba o de muestra representativa de estudiantes.

8. Referencias

- Canu, M., & Duque, M. (2016). Lecciones aprendidas de un curso de circuitos eléctricos en modalidad híbrida. In ACOFI (Ed.), *Encuentro Internacional de Educación en Ingeniería: El cambio para la competitividad y la sostenibilidad* (p. 85). Cartagena de Indias.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391-418.
- Ebel, R.L. (1954). Procedures for the Analysis of Classroom Tests. *Educational and Psychological Measurement*, 14, 352-364.
- Engelhardt Paula, & Beichner, R. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics*, 72(1), 98-115.
- Getty, J. C. (2009). *Assessing Inquiry Learning in a Circuits/Electronics Course*. Paper presented at the 39th ASEE/IEEE Frontiers in Education Conference, San Antonio, Texas.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.
- Pellegrino, J., Di Bello, L., & Brophy, S. (2013). The Science and Design of Assessment *Engineering Education in cambridge handbook of engineering education research*, 571-598.
- Rahman, M., & Ogunfunmi, T. (2010). *A concept inventory for an Electric Circuits course : Rationale and fundamental topics*. Paper presented at the Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris.
- Rodríguez Jiménez, O. R. (2013). LA EVALUACIÓN OBJETIVA EN INGENIERÍA APORTES EN PROCESOS DE EVALUACIÓN Y MEJORA CURRICULAR. In ACOFI (Ed.), *World Engineering Education Forum* (p. 1). Cartagena de Indias.
- Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology*, 3, 271-295.
- Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). 4 Reliability Coefficients and Generalizability Theory. *Handbook of Statistics*, 26, 81-124.

Sobre los autores

- **Michaël Canu:** Maestría en Control de sistemas y de producción de la Ecole des Mines de Nantes (France), Maestría y Doctorado en didáctica de las disciplinas de la universidad Paris Diderot – Paris 7 (Francia), Doctor en Ingeniería de la Universidad de Los Andes. Profesor asociado de la Universidad El Bosque. mcanu@unbosque.edu.co
- **Mauricio Duque:** Ingeniero eléctrico, Msc en ingeniería Universidad de los Andes, DEA y Doctor en Ingeniería de INRP, Grenoble, Francia. Profesor asociado de la Universidad de los Andes. maduque@uniandes.edu.co.

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2017 Asociación Colombiana de Facultades de Ingeniería (ACOFI)