

Propiedades psicométricas del Stroop, test de colores y palabras en población colombiana no patológica*

Psychometric Properties of the Stroop color-word Test in non-pathological Colombian Population

LUCÍA CARLOTA RODRÍGUEZ BARRETO **
NINFA DEL CARMEN PULIDO ***
CARLOS ALEJANDRO PINEDA ROA ****

Universidad Pedagógica y Tecnológica de Colombia,
Tunja, Colombia

RESUMEN

El objetivo del presente artículo es revisar las propiedades psicométricas del Stroop, test de colores y palabras de Golden (2007) en población colombiana no clínica. 1332 participantes, 561 varones (42.1%) y 771 mujeres (57.9%) entre 7 y 80 años de edad, ($M = 29.80$ $DE = 17.95$). agrupados en los rangos de edad de 16 a 44 años (48.5%), 7 a 15 (28.1%), 45 a 64 (18.1%) y 65 a 80 (5.2 %). La confiabilidad fue evaluada utilizando test-retest a través del coeficiente de correlación intraclase con IC al 95% con diferencia de diez minutos entre las dos aplicaciones. La validez de constructo fue medida mediante el Análisis de componentes principales ACP. La validez convergente fue calculada a través de r de Pearson entre las subescalas de Stroop y el TMT-B. El Análisis paralelo de horn y el ACP arrojaron evidencia de tres componentes que explican el 86.84% de la varianza. Las correlaciones r de Pearson, el Stroop y TMT-B fueron significativas a 0.01 con valores de -0.41 para P; -0.35 para C; de -0.40 para PC, y de -0.14 para interferencia. La confiabilidad encontrada fue superior a 0.70 en todas las láminas del Stroop. La validez discriminante no halló diferencias entre hombres y mujeres, pero sí para edad, estrato y escolaridad. Las medias y desviaciones estándar en la población colombiana fueron inferiores en todas las escalas a la versión española, lo que puede ser explicado por diferencias culturales importantes. Los resultados permiten afirmar que el Stroop es válido y confiable para diversos grupos etarios de poblaciones no patológicas en Colombia.

Palabras clave

Stroop; validez de constructo y convergente; confiabilidad test-retest; población colombiana

ABSTRACT

The aim of this article is to review the psychometric properties of the Stroop Color and Word Test Golden (2007) in Colombian nonclinical population. 1,332 participants, 561 males (42.1%) and 771 women (57.9%) between 7 and 80 years old, ($M = 29.80$ $SD = 17.95$) grouped into age ranges of 16-44 years (48.5%), 7-15 (28.1%), 45-64 (18.1%) and 65-80 (5.2%). Reliability was assessed using test-retest through intraclass correlation coefficient with IC95% far ten minutes between the two applications. Construct validity was measured by Principal Component Analysis ACP. Convergent validity was calculated by Pearson r between Stroop subscales and TMT-B. Parallel analysis of Horn and gave evidence ACP three components explaining 86.84% of the variance. Pearson r correlations Stroop and TMT-B were significant at 0.01 to -0.41 values for P; -0.35 for C; from -0.40 to -0.14

doi:10.11144/Javeriana.upsy15-2.ppst

Para citar este artículo: Rodríguez Barreto, L. C., Pulido, N. del C., & Pineda Roa, C. A. (2016). Psychometric properties of the Stroop color-word test in non-pathological Colombian population. *Universitas Psychologica*, 15(2), 255-272. <http://dx.doi.org/10.11144/Javeriana.upsy15-2.ppst>

- * Artículo de investigación científica y tecnológica.
** Ph.D. en Neurociencias. Docente investigadora Grupo Psicología Clínica y de la Salud. Correo electrónico: lucia.rodriguez@uptc.edu.co
*** Mg. en Psicología. Docente investigadora Grupo de investigación Salud Trabajo y Calidad de Vida. Correo electrónico: ninfa.pulido@uptc.edu.co
**** Mg. en Salud Sexual y Reproductiva. Docente investigador grupo de investigación Clínica y Salud. Correo electrónico: carlos.pineda01@uptc.edu.co

for PC and Interference. Reliability found was higher than 0.70 in all Stroop scales. Discriminant validity found no difference between men and women, but if for age stratum and schooling. Means and standard deviations were lower Colombian population at all scales to the Spanish version, which can be explained by significant cultural differences. These results support the Stroop is valid and reliable for different age groups of no pathological populations in Colombia.

Keywords

Stroop; construct validity and convergent; test-retest reliability; Colombian population

Introducción

En Colombia, la mayoría de instrumentos psicológicos y neuropsicológicos que se utilizan para evaluar y emitir diagnósticos carecen de criterios de calidad técnica que permitan garantizar la validez y la confiabilidad, con las respectivas consecuencias éticas y prácticas, con el impacto que esto conlleva en la calidad del servicio en los cuales se utilizan indiscriminadamente. Las malas prácticas incluyen el uso de pruebas que en su mayoría son realizadas y normalizadas con las características de lenguaje, nivel socioeconómico y cultural de otros países, lo que constituye una preocupación permanente de su aplicación en nuestra población (León 2013).

La prueba de J. R. Stroop se utiliza para evaluar atención, flexibilidad cognitiva y el “efecto Stroop”, que consiste en la inhibición de respuestas automáticas en favor de otras más inusuales (Stroop, 1935). La prueba ha sido validada y aplicada en personas entre 7 y 80 años para orientar diagnósticos como daños cerebrales, drogadicción, demencia senil, psicopatología o estrés. (Golden, 2007). También ha sido utilizada para la identificación del trastorno de estrés postraumático (Bremner et al., 2004, citados por Golden, 2007) y el reconocimiento de déficit de atención con hiperactividad, pacientes con esquizofrenia, Alzheimer, depresión, distimia o dolor crónico, entre otros (Golden, 2007).

La valoración en población sana o disfuncional requiere la aplicación de instrumentos válidos y confiables diseñados y adaptados específicamente para la población que se pretende evaluar, además de contar con datos normativos adecuados para establecer el diagnóstico (Bauer, Tobias, & Valenstein,

1993; Squire & Shimamura, 1996, citado por Pérez, 2009). La presente validación toma en consideración solamente población sana.

El ejercicio del quehacer responsable y ético exige, entonces, adelantar estudios de validación que permitan contar con una información técnica que derive en el uso adecuado de los instrumentos de evaluación en nuestro contexto. De esta manera, y dadas las limitaciones existentes de instrumentos de evaluación neuropsicológica para la población colombiana, se presenta como alternativa el estudio psicométrico la versión del Stroop, test de colores y palabras de Charles J. Golden (2007), que fue adaptado para la población española por el departamento de I+D de TEA ediciones S.A en el año de 1994 (Golden, 2007).

El Stroop se utiliza ampliamente en el ámbito de evaluación neuropsicológica, inclusive en el contexto colombiano; no obstante, no se cuenta con un estudio amplio que revele las características de validez y confiabilidad en población no patológica en Colombia.

Antecedentes y generalidades del Stroop (efecto Stroop)

El test Stroop inicialmente pretendía conocer los efectos de la interferencia perceptual y conocer qué efectos tenía esta interferencia sobre la conducta. Más tarde se presentó como un test sensible para discriminar personas con daño cerebral, siendo capaz, incluso, de describir la localización de la lesión (Bausela & Santos, 2006).

El test Stroop comenzó su progreso a partir de investigaciones de los primeros psicólogos empíricos quienes observaron que la tipificación de colores era siempre más tardía en adultos que sabían leer que la lectura de los nombres de los colores. Posteriormente, Stroop en 1935, sugirió que la discrepancia entre las tareas “leer las palabras” y “nombrar el color” se debía a que los colores estaban relacionados a una variedad de respuesta conductual mientras que las palabras solo estaban asociadas a la lectura.

Existen varias versiones de la prueba. En 1967, Houston y Jones emplearon la versión tradicional del Stroop colocando un elemento distractor de la

presencia o no de ruido (Bausela & Santos, 2006). En 1972, Daryl Imple presentó combinaciones adecuadas de colores y palabras relacionadas con el color (p. ej., la palabra SANGRE en tinta roja) y encontró que las latencias de denominación fueron apreciablemente más rápidas que las que se obtuvieron con palabras incongruentes relacionadas con el color (p. ej., la palabra HIERBA en tinta roja) o con palabras no relacionadas (p. ej., la palabra CUADRADO en tinta roja) (Bausela & Santos, 2006). Hay otras versiones, como la de Dodrill (1978), que consiste en una única lámina que contiene 176 ítems (11x16) palabras-color (rojo, naranja, verde, azul) impresas al azar en estos colores. En la primera parte la persona lee el nombre de la palabra. En la segunda parte el participante debe denominar el color en que está impresa la palabra. Dodrill evalúa la ejecución de la prueba sobre la base del tiempo total empleado en la primera parte y la diferencia entre la segunda y la primera. La versión Victoria (Regard, 1981), la cual es un instrumento notable para la evaluación de la atención selectiva y control de la inhibición y es muy utilizado en los aspectos de investigación de las funciones ejecutivas del cerebro (Malek, Hekmati, & Amiri, 2013).

La versión informatizada (Douchamps, 1988, citado por Bausela & Santos, 2006) se trata de un software: los estímulos consisten en cuatro colores-palabra (amarillo, rojo, verde o azul) presentados de uno en uno en diferentes colores (amarillo, rojo, verde o azul) y aparecen de manera aleatoria. Sin embargo, la versión normalizada de Golden (1975) es la más manipulada y en la que se ha fundado la versión española. Consta de tres láminas de 100 ítems distribuidos al azar en cinco columnas de 20 elementos. La primera lámina está establecida por las palabras roja, verde y azul, dado que el efecto de interferencia era el mismo independientemente del número de colores empleados impresos en color negro. La segunda lámina consiste en otros 100 elementos iguales (XXXX) impresos en las diferentes tintas de color (rojo, verde y azul). El mismo color no aparece dos veces seguidas en la misma columna y los colores no siguen el mismo orden de las palabras de la primera lámina. La tercera lámina

son las palabras de la primera lámina impresas en los colores de la segunda coincidiendo en el orden de representación de las mismas.

Antecedentes de validación del Stroop en poblaciones de diversas culturas

Antecedentes relacionados a confiabilidad y validez (validez interna)

Maureira, Aravena, Gálvez, & Flores (2014) en Chile, aplicaron el test de Stroop en la versión Española del Stroop realizada por el Departamento I + D de TEA Ediciones (Golden, 1994) con 85 estudiantes de educación física, cuyas edades oscilaban entre 17 y 31 años con promedio de 20.78 (DT=2.81). El espacio entre las dos aplicaciones fue de 15 días. Conforme a la teoría, los autores señalan que el rendimiento, como el Stroop, debe mantenerse similar en estas edades, (Ostrosky, Ardila, & Rosselli, 1991) debido a lo cual se realizó un análisis estadístico a través de una prueba t para verificar si existían diferencias significativas entre el cuartil inferior y superior de las edades del grupo. Los autores encontraron que “el test de Stroop presenta un valor de confiabilidad test-retest de 0.884 y que los valores de las puntuaciones entre los cuartiles de edad más bajos y más altos no presentaron diferencias significativas ($t=1.771$; $gl=72$; $p=0.081$) lo que entrega información sobre la validez de constructo del instrumento” (Maureira et al., 2014, p. 346).

Rognoni et al. (2013) obtuvieron resultados similares a los encontrados por Maureira et al. (2014) en grupos etarios semejantes. En el 2013, el equipo de investigación de Rognoni obtuvo datos normativos para el Stroop (Golden, 2001) y la Tower of London-Drexel University versión (TOL) en una muestra de 179 participantes entre 18 y 49 años de edad, sin deterioro cognitivo y con puntuaciones mayores a 24 en el *Mini-mental State Examination* y el *Memory Impairment Screen*. No hallaron efectos de la edad y el género en las pruebas. La escolaridad influyó en las variables del Stroop (explicando el 4 al 10% de la varianza) y en algunas del TOL en la que el nivel de escolaridad explicó el 9% de la varianza en la variable “movimientos totales y tiempo

de latencia”, indicándose como posibilidad que a mayor escolaridad hay menores movimientos que implican mayor control inhibitorio y planificación de la conducta (Rognoni et al., 2013, p. 76).

Espinosa (2008), en Bogotá, Colombia, calculó la confiabilidad por estabilidad mediante test-retest. Realizó la aplicación en un primer momento y a los tres días resultando valores r de Pearson para $P=0.538^{**}$; $C=0.421^{**}$; $PC=0.311^{**}$, y $INT=0.225^*$ (significancia bilateral 0.01^{**} y 0.05^*). Para la validez de constructo realizó un análisis factorial mediante el método de componentes principales obteniendo tres factores que explicaron el 98.6%.

En la adaptación portuguesa del Stroop realizada por Soares (2009) se hizo la comparación de tres muestras, una con 290 sujetos saludables entre 25 y 80 años, otra con 32 participantes con diagnóstico de enfermedad tipo Alzheimer de inicio tardío y una tercera muestra de comparación con características similares al primer grupo. Adicionalmente, este autor realizó una sola aplicación y encontró consistencia interna global medida a través del alfa de Cronbach $\alpha=0.663$. Mediante análisis factorial exploratorio halló una estructura de tres componentes similar a Golden (2001, citado por Soares, 2009) con una correlación fuerte que explicó el 96.16% de la varianza total. Las puntuaciones medias para la población general en el test de Stroop de la muestra portuguesa fueron $P=90$ ($DT=24$); $C=66$ ($DT=17$); $PC=41$ ($DT=10$); $INT= 3.9$ ($DT=8.1$). Las puntuaciones en P , C y PC fueron menores en la población portuguesa en comparación con las puntuaciones españolas de Golden; sin embargo, los portugueses tuvieron puntuaciones mayores en interferencia.

En el estudio realizado por Henao-Arboleda et al. (2010) en Medellín, Colombia, los autores establecieron datos normativos para población colombiana de una batería de pruebas neuropsicológicas que incluía el test de Stroop, con una investigación de tipo observacional y longitudinal que evaluó la confiabilidad mediante test-retest en 848 participantes de población general, 151 personas con demencia tipo Alzheimer (DTA) y 36 con deterioro cognitivo Leve (DCL), mayores de 50 años o más y residentes en la ciudad de Medellín. En relación a la

velocidad como función cognitiva, encontraron una consistencia interna total (intervalo de confianza del 95%) con valores alfa de Cronbach en $P=0.78$, $C=0.75$ y $PC=0.78$. En cuanto a la población general se encontraron valores de alfa para P de 0.74, C de 0.72 y para PC de 0.74. El mismo estudio de Henao-Arboleda et al. (2010) en la replicabilidad test-retest (un mes entre primera y segunda aplicación) con una muestra de 119 sujetos, obtuvo puntuaciones medias en la primera aplicación de $P=30.4$; $C=41.1$, y $PC=83.9$ y en la segunda de $P=28.6$; $C=39.3$, y $PC=75.7$, con un coeficiente de correlación intraclass (CCI) de $P=0.86$, $C=0.79$ y $PC=0.84$ para cada tarea del Stroop.

Antecedentes relacionados a variables socio demográficas (validez externa o con relación a un criterio)

La validación de un test no viene sólo a través de un análisis de componentes principales o de un análisis factorial común y calculando el alfa de Cronbach de la prueba y sus subpruebas; es preciso demostrar también que el atributo (y subatributos) discrimina entre las personas que componen la muestra que hemos utilizado (Crocker & Algina, 1986).

Diferencias por la variable edad

El Stroop es una prueba de amplio uso en la investigación de problemas de atención y en el diagnóstico de trastornos atencionales, además permite explorar la flexibilidad mental, la atención selectiva, la inhibición cognitiva y la velocidad de procesamiento de la información (MacLeod, 1991; Rognoni et al., 2013).

En 1932, Ligon halló con una prueba similar al Stroop, en una muestra de 635 estudiantes de escuelas públicas, diferencias significativas en las tareas palabra-color, encontrando además, que la velocidad aumentaba con la edad. Por su parte, Comalli, Wapner, & Werner (1962) en una muestra de 235 sujetos con edades entre 7 a 80 años, hallaron resultados similares a Ligon, evidenciando que en niños de 7 años la velocidad era más lenta que en el grupo de sujetos de 17 a 19 años.

En el estudio de Rand, Wapner, Werner, & MacFarland (1963, citado por Golden, 2007) con cuatro grupos de edad entre los 6 a 16 años, hallaron que los errores disminuían con la edad y aumentaban al pasar de P a C y de C a PC; los autores evaluaron otras características, como las inadecuaciones (decir respuestas que no se referían ni a la tinta ni a la palabra escrita), las respuestas mal articuladas, repetición de palabras entre ítems, las omisiones, frases lingüísticas y no lingüísticas insertadas. Se encontró que la ejecución en la mayoría de características decreció con la edad. Por su parte, Cramer (1967, citado por Golden, 2007) adaptó el Stroop para niños en edad escolar y concluyó que el concepto primordial en este grupo etario es la forma.

En Chile, Conca & Ibarra (2004) estandarizaron el test de Stroop. La muestra estuvo conformada por 615 participantes con edades comprendida entre 8 a 12 años, pertenecientes al sector urbano y de niveles socioeconómicos alto, medio y bajo. Utilizando Anova de un factor y test de Scheffé, el estudio mostró diferencias estadísticamente significativas para la edad ($p=0$), lo que llevó a las autoras a concluir que se requieren tres normas diferenciales: una para los 8 años, otra para los 9 y 10 años y una última para los 11 y 12 años para P, C y PC. En INT la media disminuyó a medida que aumentaba la edad.

En Holanda, Van der Elst, Van Boxtel, Van Breukelen, & Jolles (2006) administraron el test de Stroop a 1856 participantes, entre 24 a 81 años, los datos normativos fueron clasificados por edad, sexo y nivel educativo. Los resultados mostraron que todas las láminas P, C, PC e interferencia fueron significativamente afectadas por la edad. La ejecución en todas las láminas disminuyó con la edad. A partir de sus hallazgos, los autores explican el concepto de “reserva cognitiva”, en el que la educación se convierte en un factor protector ante los daños producidos por el envejecimiento.

Un estudio con escolares realizado en Tenerife, España, cuyo objetivo fue hallar datos normativos para el test de Stroop en una muestra de 1032 estudiantes entre 6 y 12 años (edad tomada en años y meses), sin enfermedad neurológica ni psiquiátrica, mediante un diseño transversal con la edad como

variable independiente distribuida en seis grupos (6-3 a 7-4, 7-5 a 8-4, 8-5 a 9-4, 9-5 a 10-4, 10-5 a 11-4 y 11-5 a 12-4), aplicando el test de Stroop (Golden, 2001) además del test de cinco dígitos (Five-Digit Test, FDT) como alternativa no lectora para determinar capacidad de interferencia, mediante un análisis de varianza, encontró que P, C, PC e INT presentaron diferencias significativas en las puntuaciones medias en función de la edad (Martín et al., 2012, p. 45).

Así, para el test de Stroop, Martín et al. (2012) encontraron que P presentó diferencias significativas para la mayoría de los grupos, excepto para el rango de edad 10-5 a 11-4 vs. 11-5 a 12-4. En C, las medias fueron diferentes para los tres primeros grupos, excepto para los grupos de mayor edad que no mostraron diferencias. Para PC no se hallaron diferencias en el grupo 6-3 a 7-4 y 7-5 a 8-4. Para INT, el grupo de menor edad presentó puntuaciones elevadas, similares a los grupos de mayor edad.

En los resultados para FDT se halló menor control de interferencia en niños prelectores, esta prueba sirvió como mejor predictor para evaluar la capacidad de inhibición en estas edades, se encontró que con el crecimiento progresivo de la capacidad de inhibición aumenta entre los nueve y once años (Martín et al., 2012).

En síntesis, edad es una variable que afecta los resultados con las diferentes láminas del Stroop.

Diferencias por la variable sexo

En relación a las diferencias por sexo, Peretti (1969, como se cito en Golden, 2007) estudió una muestra de 50 varones y 50 mujeres encontrando una puntuación media para PC de 48.2 en mujeres y 55.1 varones que lo condujo a concluir la existencia de una diferencia significativa al 0.01 a favor de los varones. En la traducción, adaptación y validación española del Stroop realizada por el Departamento I+D de TEA Ediciones en 1994 (en adelante TEA) para población general, se encontraron puntuaciones medias de P=118 (DT=18), C= 77 (DT=14), PC=49 (DT=10) e INT=2.70 (DT=8.53); los resultados para esta validación fueron superiores a muestras americanas, el estudio encontró diferencias

significativas para INT en varones con 3.63 y mujeres 1.88, pero no hubo diferencias significativas para P, C y PC en la variable sexo. TEA (1994) señaló que en relación a niños y ancianos se hace necesario contar con muestras más representativas para llegar a datos más concluyentes.

Contrario a los resultados de Peretti, en el estudio de Van der Elst et al. (2006) las mujeres tuvieron mejor desempeño en las láminas C, PC e interferencia que los hombres, pero no hubo diferencias en la lámina P.

Por otro lado, Conca & Ibarra (2004) no encontraron diferencias significativas en todo el Stroop según la variable sexo. En efecto, se observó que las mujeres tienen un mejor desempeño en las dos primeras láminas P y C de la prueba, sin embargo, este desempeño se invierte a favor de los hombres en la tercera lámina. Al realizar el análisis estadístico a través de “t de student” para cada una de las láminas del test, se concluyó que no existen diferencias significativas en dicha variable ($p= 0.051$). A la misma conclusión llegaron Espinosa (2008) y en Colombia Soares (2009) en Portugal al señalar que no encontraron diferencias en cuanto a la variable sexo para ninguna de las puntuaciones del Stroop.

En conclusión, la evidencia actual es contradictoria respecto a las diferencias por sexo respecto de cada una de las láminas del Stroop.

Diferencias por la variable estrato

Siguiendo a Conca & Ibarra (2004), se hallaron diferencias significativas en el nivel socioeconómico ($p=0$), lo que permitió concluir que para las láminas P, C y PC se requieren dos normas diferenciales: una para los estratos alto y medio y otra para el estrato bajo. A partir de los resultados obtenidos se puede concluir que, para el índice de interferencia, hubo diferencias estadísticamente significativas ($p= 0.002$) solamente entre el nivel socioeconómico alto y el bajo.

Diferencias según inteligencia y problemas de aprendizaje

En cuanto a validez de criterio, Espinosa (2008) realizó un trabajo de tipo psicométrico para validar y estandarizar el test de Stroop en una muestra de 381 niños, niñas y adolescentes, en edades comprendidas entre 7 y 16 años. La autora utilizó la versión del test de Stroop de Golden (1999). Los participantes fueron diagnosticados y divididos en diferentes grupos: (a) coeficiente intelectual (CI) alto; (b) CI normal; (c) problemas de aprendizaje; (d) problemas de atención, y (e) problemas emocionales. Mediante Anova se encontraron diferencias significativas en P, con puntuación media más alta en personas con CI alto y más baja para personas con problemas de aprendizaje; en C, las puntuación media más alta la obtuvo el grupo con CI alto y la más baja el grupo de problemas emocionales; para PC, las puntuación media más alta la obtuvo el grupo con CI alto y la más baja el grupo de problemas emocionales; para INT no se encontraron diferencias entre los grupos.

Actualmente, y a pesar del uso extendido del Stroop entre los neuropsicólogos colombianos, no se encontró un estudio cuyo objetivo fuera la validación de la prueba en población colombiana no clínica. El objetivo de la presente investigación fue validar el test de Stroop en población colombiana sana en edades comprendidas entre 7 y 80 años.

Método

Diseño

En esta investigación se utilizó un estudio instrumental de carácter psicométrico, de acuerdo a los conceptos y técnicas que se utilizan en esta área (Montero & León, 2007).

Participantes

Criterios de inclusión-exclusión

Para la selección de los participantes se tuvo en cuenta que se encontraran en un rango de edad

comprendidas entre 7 y 80 años de edad, supieran leer, que participaran voluntariamente y que firmaran el consentimiento informado. En menores de edad, consentimiento firmado por padres o representantes legales; no presentar déficit cognitivo, alteraciones perceptivo visuales (incluyendo el daltonismo), déficits auditivos o motores que dificultaran o impidieran realizar la prueba, y no haber estado medicado (psicofármacos, Ritalina o anticonvulsivos).

Población y muestra

La población la conformaron hombres y mujeres entre 7 y 80 años de edad, vinculados a colegios, universidades, empresas y programas del adulto mayor o pensionado en entidades públicas, privadas o mixtas que pertenecieran a los estratos comprendidos entre uno y cinco de la ciudad de Tunja. No obstante, los participantes provenían de diversas regiones del país como lugar de nacimiento.

Para el cálculo del tamaño muestral se tomó como referencia la proyección del Departamento Administrativo Nacional de Estadística (DANE) para el 2013 de infancia, adolescencia, juventud y tercera edad de la ciudad de Tunja, según la proyección por grupo poblacional 2011-2015 que es de 84.565 personas, dando como resultado $n= 1054$ participantes con 95% de confianza y error de 3%.

La muestra inicial encuestada estuvo conformada por 1357 participantes de los cuales 26 no se incluyeron en la investigación; 18 niños no llevaron el consentimiento informado firmado por los padres o representante legal y dos no asistieron cuando estaban citados; de los adultos, uno era daltónico, tres eran analfabetas, uno tenía retardo mental y uno tenía más de 80 años. La muestra final quedó conformada por 1332 participantes, 561 varones (42.1%) y 771 mujeres (57.9%) entre 7 y 80 años de edad (Media= 29.80 DT= 17.95).

En relación a la edad de los participantes, tomada en años cumplidos, se agruparon en rangos de 7 a 15, 16 a 44, 45 a 64 y 65 a 80 años, de acuerdo a los criterios del Stroop, test de colores y palabras para la adaptación española (Golden, 1994); para

este estudio, el mayor porcentaje se concentró en el rango de 16 a 44 años (48.5%), seguidos de los de 7 a 15 (28.1%), 45 a 64 (18.1%) y la menor cantidad de participantes en el rango de 65 a 80 (5.2 %). Para el estrato, 86 hacían parte del estrato uno, 424 del dos, 589 del tres, 172 para el cuatro y 61 para el cinco. El estrato se agrupó en bajo (1 y 2), medio (3) y alto (4 y 5).

El muestreo que se utilizó fue de tipo no probabilístico por conveniencia, que constituye una muestra seleccionada por medio de métodos no aleatorios (Cerdá, 1978; Malhotra, 2004); para Casal & Mateu (2003) consiste en la elección por métodos no aleatorios de una muestra cuyas características sean similares a las de la población objetivo.

En cuanto a la escolaridad, se encontró que el 22.8% de la muestra tenía estudios secundarios, seguidos de pregrado (16.9%), primaria (16.4%), profesionales (15.6%), bachilleres (10.1%), posgrado (10.3%) y finalmente, con el menor porcentaje, los técnicos (7.8%).

Instrumentos

Los instrumentos utilizados fueron el Stroop, test de colores y palabras en la adaptación realizada por Golden en 1994 y el Trail Making Test forma B (TMT- B) (Reitan, 1992). A continuación, se describe cada una de las pruebas.

Stroop, test de colores y palabras (TEA, 1994)

Instrumento adaptado para la población española por de TEA ediciones S.A. en el año de 1994 (Golden, 2007). El test permite indagar la atención a través de la integración de los subpuntuajes: palabra (P), color (C), palabra-color (PC), y dentro de este proceso psicológico, la capacidad de flexibilidad cognitiva, resistencia a la interferencia y grado de complejidad cognitiva (Chapi, 2011). Golden & Malloy (2001, 2000, citados por Soares, 2009) indican que la prueba evalúa las dimensiones asociadas a la flexibilidad mental, habilidades de inhibición y a la resistencia a la interferencia en relación a estímulos externos.

Generalidades de la prueba. El test consta de tres láminas en tamaño de papel A4, cada una con cinco columnas y veinte elementos intercalados, para un total de 100 en cada una. El participante cuenta con 45 segundos para leer cada lámina. En la primera lee una lista de palabras que señalan los colores “rojo”, “verde”, “azul”, escritas en tinta negra; en la segunda lee solo el color en que están escritas una serie de cuatro letras “XXXX”, “XXXX”, “XXXX” que incluyen los colores de la primera lámina; en la última, la persona debe nombrar el color en que está escrita la palabra que no corresponde a la palabra “rojo”, “azul”, “verde”. Se deben seguir las instrucciones de aplicación descritas en el manual.

La prueba de Stroop permite obtener las siguientes puntuaciones: (a) primera lámina, número correcto de palabras leídas (palabra “P”); (b) segunda lámina, número correcto de colores leídos (color “C”), y (c) tercera lámina, número correcto de ítems leídos (palabra-color “PC”). Para calcular la interferencia “INT” se deben aplicar las siguientes formulas: a) $PC' = P \times C / P + C$ y b) $INT = PC - PC'$.

El Trail Making Test (TMT)

La prueba del TMT (Reitan, 1992) fue desarrollada en 1938 y Partington y Leiter (1949) la adaptaron del original titulándola Partington’s Pathways Test. Se dio a conocer con el nombre de *Distributed Attention Test* y en español como Test de atención distribuida; la prueba también se conoce como Test del trazo. Inicialmente formó parte de la *Army individual Test Battery* (1944) y posteriormente de la batería neuropsicológica de Halstead Reitan (Reitan & Wolfson, 1993, citados por Burin, Drake & Harris, 2007).

Este instrumento consta de dos formas, A y B, para niños de 9 a 14 y adultos de 15 a 89 años. La parte A mide habilidades motoras, visoespaciales de búsqueda visual y atención sostenida. La forma B evalúa de forma adicional la flexibilidad cognitiva y atención dividida; esta parte B exige mayor esfuerzo en la realización de la tarea. En el modelo de atención planteado por Mirsky (1996, citado por Burin, Drake, & Harris, 2007), la tarea del Stroop respondería al componente focalizador–ejecutor. El

TMT consta de dos partes, la parte A y la B, con un ensayo que se debe pasar antes de administrar la prueba que se califica. En la A (para niños y adultos), los círculos van de 1 al 8 y en la B (para niños y adultos), los números del 1 a 4 y las letras de A a D. Se explica la tarea hasta que se esté seguro de que las instrucciones fueron comprendidas por la persona a la que se le aplicará la prueba. Se entrega un lápiz negro para realizar los trazos.

TMT-B: para los niños son 15 círculos del 1 al 8 y las letras de la A a la G; en adultos consta de 25 círculos pero esta vez van del 1 al 13 y las letras de la A a la L, la tarea consiste en que se deben unir de forma alternada un número y una letra (1A, 2 B, 3C) y así sucesivamente hasta terminar la secuencia en el círculo con el número 8 para los niños y en 13 para los adultos. Se debe tener en cuenta el orden ascendente para los números y las letras. Se registra el tiempo que emplea el participante en realizar la tarea (en segundos).

En la presente investigación se eligió el TMT forma B para evaluar validez convergente por ser una prueba corta, de fácil administración y principalmente porque, al igual que el Stroop, evalúa atención, flexibilidad cognitiva e interferencia como parte de las funciones ejecutivas (Portellano, 2009).

Procedimiento

El proyecto fue desarrollado en varias fases. Se realizó una búsqueda bibliográfica en bases de datos como Proquest, Ebsco, ScienceDirect, Scopus, Google Académico, entre otras, para construir el marco teórico. En la segunda fase se seleccionaron los participantes de la muestra de acuerdo a las edades propuestas (7 a 80 años). Las edades son el criterio para ubicar los posibles sitios de aplicación: se eligen, entonces, para el rango de 7 a 18 años cuatro colegios, dos de carácter público y dos privados; de 18 a 64, dos universidades públicas y cinco privadas (estudiantes, docentes y administrativos), así como seis empresas privadas y tres públicas (nivel administrativo y operativo), y, finalmente, para 65 a 80 años, los programas de la tercera edad

o pensionados en programas de cuatro empresas públicas y una mixta.

En la tercera fase se pidieron autorizaciones en cada uno de los sitios tentativamente seleccionados, inicialmente con citas para presentar la propuesta y posteriormente con solicitudes escritas para formalizar las autorizaciones y realizar visitas a las instalaciones físicas para establecer los requisitos de aplicación de la prueba en aspectos como luz, privacidad y ausencia de ruido. Se establece la dinámica para que los participantes se retiren del aula de clase o sitio de trabajo de manera que las personas que participen no entorpezcan el trabajo de clase o en las empresas.

En la cuarta fase se hicieron capacitaciones, entrenamientos y simulación para la aplicación. Se contrataron un psicólogo, tres psicólogas y cuatro estudiantes de último semestre de Psicología para la aplicación de las pruebas. Cada uno fue capacitado por la investigadora principal para aplicar la prueba y para realizarla teniendo en cuenta las instrucciones y pasos en la aplicación. En los colegios, para los niños menores de edad se entregaron los consentimientos informados a las directoras de curso para que los niños lo llevaran a la casa y una vez que los devolvieron firmados, se preparó la actividad. Cada aplicador lleva los consentimientos, las pruebas, formatos de registro, cronómetro, lápiz y esfero en paquetes ordenados para cada participante.

La quinta fase, de aplicación, se hizo de manera individual. Las sesiones estaban establecidas de manera que todos los aplicadores las realizaran de la misma forma. Una vez recibido, se procedía a dar las instrucciones y cuando se estaba seguro(a) que se comprendía la tarea, se daba inicio y se cronometraron los 45 segundos establecidos para cada lámina. La segunda aplicación se realizó en un lapso de tiempo de diez minutos, teniendo en cuenta que esta tarea no es susceptible de aprendizaje, pero sí de cambios madurativos o fisiológicos asociados a la edad. Para la aplicación del Trail Making Test forma B se pasa primero un ensayo en el que se da un ejemplo de la tarea en una forma corta, si es necesario se corrige cualquier error, cuando el ensayo termina y la tarea se finaliza con éxito, se entrega entonces la hoja que se califica, se cronometra el

tiempo que utiliza para realizarlo, se anota (en minutos) y se registran las calificaciones en el formato creado para tal fin.

Al finalizar las aplicaciones se procedió a la fase seis: la elaboración de una base de datos en Excel y SPSS versión 19. En dicha fase se transcribieron los formatos de registro y calificaciones en Excel, realizando varias revisiones para verificar la correcta transcripción de los datos o posibles errores de digitación.

Aspectos éticos

Se siguieron los lineamientos de la Ley 1090 de 2006 y la Resolución 008430 de 1993. Se contó con el aval del comité de ética de la Dirección de Investigaciones (DIN) de la UPTC, a quienes se socializó la propuesta y el consentimiento informado y se realizaron los ajustes respectivos de acuerdo a las recomendaciones de dicho comité.

Análisis estadísticos

Se aplicaron las fórmulas para obtener las puntuaciones PC' y la interferencia de acuerdo a lo señalado en el manual y el tiempo del TMT-B es transformado de minutos a segundos, para cada uno de los datos. Se realizaron las transformaciones especificadas en el manual según los rangos de edad. Se calcularon los análisis psicométricos de confiabilidad y la validez que corresponde al objetivo de este estudio. La confiabilidad usando el método test-retest por el método de coeficiente de correlación intraclase CCI con intervalo de confianza del 95%. Es importante resaltar que comúnmente se hace uso del r de Pearson en el cálculo de la confiabilidad test-retest. Sin embargo, este mide únicamente la fuerza de la asociación lineal entre dos variables y no el acuerdo entre ellas. Para Serra & Peña (2006) la r de Pearson "es un coeficiente interclase más que intraclase" y por ello tiende a sobreestimar la confiabilidad. El CCI evalúa el grado de variación entre un tiempo 1 y un tiempo 2, si la diferencia es pequeña, el CCI será grande (cerca de 1). Además, el CCI mide la relación entre variables que comparten la misma métrica y varianza. Para la validez

convergente se utilizaron los puntajes de la prueba TMT-B de Reitan (1992) y se correlacionan con los puntajes del Stroop para P, C y PC mediante la r de Pearson. Para la validez de constructo se utilizó el método de extracción de componentes principales analizando el porcentaje de varianza explicado y los distintos componentes extraídos. Una confirmación del número de componentes es establecida mediante el Análisis paralelo de Horn (1965).

Se realizó un análisis descriptivo mediante medias y desviación estándar según grupo etario. Se revisó el cumplimiento de los supuestos de normalidad (prueba de Kolmogorov-Smirnov), se verificó el supuesto de igualdad de varianzas mediante la prueba de Levene y se realizaron los tratamientos estadísticos inferenciales de acuerdo a las diferentes variables contempladas, como el sexo, la edad, la escolaridad y el estrato. Usando la prueba t student se verificó si había diferencias significativas entre hombres y mujeres de cada uno de los rangos de edad para cada una de las puntuaciones transformadas de P, C y PC. Seguidamente, usando la prueba F de Fisher se verificó si había diferencias significativas según el nivel

educativo y el estrato de cada uno de los rangos de edad para cada uno de las puntuaciones transformadas de P, C y PC.

Resultados

De acuerdo a los objetivos del presente trabajo, los resultados serán presentados primero en lo relacionado al estudio psicométrico del Stroop, test de colores y palabras de Golden (2007) respecto de la confiabilidad y la validez. Seguidamente se presentan los estadísticos descriptivos por cada rango de edad. Posteriormente se verifica el cumplimiento de los supuestos de normalidad y homocedasticidad en las variables P, C y PC, luego de lo cual se presentan los análisis inferenciales.

Análisis de los resultados de la confiabilidad del Stroop

La estimación de la confiabilidad se realizó utilizando test-retest calculando el coeficiente de correlación intraclass CCI entre las dos aplicaciones de cada subprueba a un nivel de 95% para el intervalo

TABLA 1.
Resultados de confiabilidad del Stroop mediante test-retest

Subescala	CCI	IC 95%	p valor
P	0.91	(0.90;0.92)	0
C	0.87	(0.83;0.90)	0
PC	0.78	(0.51;0.88)	0

Fuente: elaboración propia.

TABLA 2.
Resultados correlacionales entre las escalas del Stroop y TMT-B

Resultados correlacionales entre las escalas del Stroop y TMT-B	
Lámina 1: (P)	0.416**
Lámina 2: (C)	0.355**
Lámina 3: (PC)	0.404**
Interferencia (INT)	0.140**
Significancia bilateral	0
n	860

Nota: **La correlación es significativa a nivel 0.01

Fuente: elaboración propia.

de confianza. El CCI mide tanto la consistencia en el tiempo como entre evaluadores. Indica en qué grado diferentes evaluadores puntuarán de manera idéntica la ejecución en una tarea.

La tabla 1 muestra los coeficientes de correlación intraclase encontrados en cada subescala P, C y PC, valores que son muy buenos (superiores a 0.78 para las subescalas P, C y PC). Esto es congruente con la consistencia reportada en los estudios colombianos, españoles y norteamericanos (Henao-Arboleda et al., 2010).

Análisis de los resultados de validez del Stroop

En el presente estudio se hallaron validez convergente y de constructo. La validez convergente se correlacionó a través de *r* de Pearson, los puntajes de las pruebas Stroop (Golden, 1994) y el TMT-B (Reitan, 1992).

Resultados validez convergente

Los resultados mostraron una correlación negativa y significativa con valores entre -0.35 y -0.41 para las tres subescalas P, C y PC con las puntuaciones en el TMT-B, lo que indica que a mayor número de ítems leídos en la primera, se emplea menor tiempo en la segunda prueba; las tres primeras escalas presentan correlaciones significativas y aceptables, en tanto están por encima de 0.30 que, según Norman & Steiner (1995, p. 105), se consideran correlaciones

importantes. Por su parte, la correlación INT – TMT-B es débil porque no cumple con este supuesto (véase tabla 2).

Resultados validez de constructo

El procedimiento estándar para dar cuenta de la validez de constructo de una prueba psicológica es el análisis factorial. Permite explicar la varianza de un conjunto de variables a un número más reducido de factores o componentes. Golden (1994) utilizó el Análisis factorial exploratorio (AFE) con el fin de saber cuál de las puntuaciones del Stroop permitía aportar una información única y original (p. 15).

Es importante aclarar que durante muchos años ha existido una fuerte controversia entre los que afirman que el método más adecuado para realizar un AFE es el análisis factorial común (Factorización de ejes principales - FEP) y los que aconsejan el uso del Análisis de componentes principales (ACP) (Henson & Roberts, 2006; Widaman, 2007).

De acuerdo con Widaman (2007), “ambas técnicas tienen grandes diferencias y algunas similitudes. Lo común es que los métodos FEP y ACP no permitan obtener errores típicos de los parámetros que producen. La diferencia más importante es que el método FEP separa la varianza común de la varianza específica, mientras que el ACP incluye en la solución dimensional tanto una como otra, y tiende a sobredimensionar la matriz de componentes interpretativa”.

TABLA 3.
Varianza total explicada por los tres componentes del ACP

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3.742	46.777	46.777	3.742	46.777	46.777
2	1.865	23.312	70.088	1.865	23.312	70.088
3	1.340	16.755	86.844	1.340	16.755	86.844
4	0.773	9.665	96.509			
5	0.276	3.454	99.963			
6	0.003	0.037	100			

Método de extracción: análisis de componentes principales.

Fuente: elaboración propia.

Debido a que en el presente estudio no se cumplieron los supuestos para el AFE mediante FEP, se procedió a establecer el número de componentes y la explicación de su varianza total. En consecuencia, se hablará de componentes y no de factores.

Como se observa en la tabla 3, conforme la recomendación Kaiser y Guttman, los tres componentes encontrados con autovalores mayores a 1 explican el 86.84% de la varianza total (Guttman, 1953; Kaiser, 1960). Sin embargo, y pese a la polémica acerca del número de factores o componentes a retener en un análisis factorial, hemos realizado el procedimiento de Análisis paralelo de Horn (1965), el cual ha ganado bastante terreno en la literatura psicométrica al ser más objetivo que la regla de Kaiser como procedimiento para determinar el número de factores a retener en un AFE (Ruiz & San Martín, 1992). El análisis paralelo compara los valores propios de los datos del estudio real con valores propios aleatorios. El método consiste en aceptar aquellos factores o componentes siempre que se cumpla que el valor propio del estudio real supere el valor propio del estudio aleatorio o simulado (Brett, Brown, & Onsmann, 2010).

El resultado encontrado confirmó los mismos tres componentes encontrados en el ACP (véase tabla 4). Los componentes encontrados tienen la agrupación presentada en la tabla 5 y confirman una solución componencial similar a la encontrada por Soares (2009), lo cual indica que la estructura del Stroop encontrada en la muestra portuguesa es similar a la estructura encontrada en la muestra colombiana. El primer componente, con un va-

lor propio de 3.74, incluye ítems con saturaciones entre 0.74 y 0.98 y explica el 47% de la varianza total, refiriéndose al número de palabras leídas en la primera página del test (P), al número de colores denominados en la página dos del test (C), a la suma de esta dos puntuaciones (C1+P1) y al número de respuestas correctas en la página tres del test que contiene la interferencia (PC1).

El segundo componente, con un valor propio de 1.86, está constituido por ítems que contribuyen a explicar el 23% de la varianza total y cuyas saturaciones fluctúan entre 0.64 y 0.92. Este componente incluye lo relacionado esencialmente con medidas de interferencia, observando que las puntuaciones PC1-PC'1 y PC1-C1son medidas de interferencia "pura", corregidas por factores de velocidad (Soares, 2009).

Análogamente, el tercer componente hallado en el presente estudio con un valor propio de 1.34 incluye un ítem con saturación de 0.84, explicando el 16% de la varianza total. El ítem que constituye este factor está relacionado con la relación entre el número de palabras leídas en la primera página del test y los colores nombrados en la página dos del test (P1/C1).

El presente estudio confirma también al propio Golden cuando afirma que "es el segundo componente el que mejor representa las medidas de interferencia pura (PC1-PC'1 y PC1-C1) corregidas por factores de velocidad" (Golden, 1994, p. 15).

TABLA 4.

Valores propios de los datos reales, medias y percentiles de los valores propios simulados

Variable	Datos en bruto	Medias	Percentiles	Decisión
1	3.74	1.11	1.15	Aceptar
2	1.86	1.07	1.09	Aceptar
3	1.34	1.04	1.06	Aceptar
4	0.77	1.01	1.03	Rechazar
5	0.27	0.98	1	Rechazar
6	0.002	0.95	0.97	Rechazar
7	0	0.92	0.94	Rechazar
8	0	0.89	0.91	Rechazar

Fuente: elaboración propia.

PROPIEDADES PSICOMÉTRICAS DEL STROOP, TEST DE COLORES Y PALABRAS EN POBLACIÓN COLOMBIANA NO PATOLÓGICA

TABLA 5.
Matriz de componentes encontrados mediante ACP

	Componente		
	1	2	3
PC1	0.746		
C1	0.963		
P1	0.869		
C1+P1	0.982		
PC1-PC'1		0.928	
PC1-C1		0.646	
PC1/C1			0.598
P1/C1			0.843

Fuente: elaboración propia.

La distribución por rangos de edad en la muestra final estuvo compuesta así:

Niños (7 a 15 años)	375
Adultos jóvenes (16-44 años)	647
Adultos (45-64 años)	241
Adultos mayores (65-80 años)	69

TABLA 6.
Resultados descriptivos del Stroop en población colombiana, corregidos por edad

	P	C	PC	INT
	Media (DE)	Media (DE)	Media (DE)	Media (DE)
Población general	103 (15)	69 (12)	41 (8)	-0.21 (6.53)
Varones	103 (15)	67 (12)	40 (8)	-0.14 (6.80)
Mujeres	103 (15)	71 (11)	42 (8)	-0.47 (6.32)
Niños (7 a 15 años)	106 (13)	70 (11)	41 (8)	-0.51 (5.65)
Adultos jóvenes (16-44 años)	104 (14)	71 (12)	41 (9)	-0.61 (6.70)
Adultos (45-65 años)	99 (18)	65 (12)	38 (8)	-0.56 (6.45)
Adultos mayores (65-80 años)	91 (20)	63 (12)	43 (7)	6.29 (6.42)

Fuente: elaboración propia.

TABLA 7.
Pruebas de normalidad para la subprueba P en los diversos grupos etarios

		NIÑOS	ADULTOS JOVENES	ADULTOS	ADULTOS MAYORES
Diferencias más extremas	Absoluta	0.136	0.054	0.109	0.224
	Positiva	0.136	0.054	0.109	0.044
	Negativa	0.034	-0.052	-0.001	-0.224
Z de Kolmogorov-Smirnov		1.313	0.682	0.844	0.919
Sig. asintót. (bilateral)		0.064	0.741	0.474	0.367

Fuente: elaboración propia.

Análisis descriptivo del Stroop en la muestra colombiana, 2013

La versión Stroop de TEA (1994) fue aplicada a un total de 1332 personas con rangos de edad entre 7 y 80 años con un promedio de 29.8 años ($DE = 17.94$). La muestra se compuso de 771 mujeres (57.88%) y 561 hombres (42.11%).

La tabla 6 muestra las medias y desviaciones estándar en cada uno de los rangos de edad para las subescalas P, C, PC e INT. Cada uno de estos estimadores fue calculado luego de haber realizado la corrección para los rangos de edad de niños, adultos y adultos mayores señalada en el manual de la prueba.

Con base en las puntuaciones en cada subescala P, C, PC e INT con respecto a los estadísticos descriptivos obtenidos, es preciso señalar lo siguiente:

- Los resultados descriptivos de la muestra colombiana son inferiores a los obtenidos en poblaciones españolas en las cuatro puntuaciones del Stroop, Golden (1994, p. 41).
- No se aprecian grandes diferencias entre varones y mujeres en cada una de las distintas sub escalas.
- Con relación a las puntuaciones de interferencia, se encuentra que los adultos mayores obtuvieron un mayor promedio, lo que se interpreta como mayor resistencia a la interferencia. No obstante, no se aprecian diferencias importantes en los puntajes promedio entre los niños, jóvenes y adultos.

Resultados del análisis de los supuestos de normalidad y homocedasticidad

Como se aprecia en la tabla 7, las pruebas de Kolmogorov-Smirnov confirman que los datos de la subprueba P provienen de poblaciones normalmente distribuidas; las demás subpruebas básicas C y PC arrojaron una distribución de probabilidad similar. Así mismo, mediante la prueba de Levene, confirmaron que se cumple el supuesto de homocedasticidad o igualdad de varianzas.

Análisis de las subescalas según variables sociodemográficas

El Anova de un factor arrojó diferencias significativas en todos los rangos de edad para todas las subescalas P, C, PC e Interferencia. Para el caso de la subescala P, $F(3.1328) = 13.25$; $p=0$; para la subescala C, $F(3.1328) = 18.36$, $p=0$; en la subescala PC, $F(3.1328) = 14.17$, $p=0$. Finalmente para el caso de la Interferencia se encontró $F(16.1315) = 1.88$, $p = 0.018$.

Conforme se aprecia en la tabla 8, las pruebas t para muestras independientes confirmaron que no hubo diferencias significativas con respecto al sexo en las subescalas P, PC e INT una vez estas se corrigieron por la edad. Únicamente la subescala C arrojó diferencias significativas según la variable sexo.

Igualmente, el Anova de un factor arrojó diferencias significativas según estrato socioeconómico sólo para la subescala P [$F(1326.4) = 2.77$, $p = 0.026$] y pa-

TABLA 8.*Prueba de homogeneidad de varianzas y prueba T entre hombres y mujeres para diferencia de medias*

		Prueba de Levene		Prueba T para la igualdad de medias según la variable sexo						
		F	Sig.	T	gl	Sig. (bil.)	Diferencia de medias	Error típ. de la diferencia		
Puntaje P corregido	Se han asumido varianzas iguales	0.91	0.34	1.28	1330	0.2	1.12	0.87		
Puntaje C corregido	Se han asumido varianzas iguales	0.17	0.68	4.65	1330	0	3.07	0.66		
Puntaje PC corregido	Se han asumido varianzas iguales	0.05	0.83	1.41	1330	0.16	0.65	0.46		
interferencia general	Se han asumido varianzas iguales	1.26	0.26	-1.73	1330	0.08	-0.63	0.36		

Fuente: elaboración propia.

ra la subescala PC [$F(1326.4) = 3.53, p = 0.007$]. Sin embargo, no se encontraron diferencias significativas en la subescala C ni en la Interferencia. En cuanto a P, mediante la prueba de Tukey, el análisis *post hoc* con $\alpha = 0.05$ arroja dos subconjuntos homogéneos así: para los estratos 1, 2, 3 y 5 niveles promedio entre 99 y 103; mientras que el estrato 4 posee un valor promedio en la escala P de 105. Análogamente, para la subescala PC los dos subconjuntos homogéneos son: uno conformado los estratos 1, 2 y 3 con promedios entre 39 y 41 y el otro conformado por los estratos 4 y 5 con niveles promedio entre 42 y 42.4.

Finalmente, con relación al nivel educativo se encontraron diferencias significativas en todos los niveles educativos. Para la subescala P, $F(16.1315) = 3.33; p = 0$; en la subescala C, $F(16.1315) = 5.89; p = 0$; en la subescala PC se obtuvo $F(16.1315) = 6.02; p = 0$, y finalmente para la escala de Interferencia se obtuvo $F(16.1315) = 1.88; p = 0.018$.

Conclusiones y discusiones

Concluimos que el Stroop, test de colores y palabras de Golden (2007), versión en español de TEA (1994), posee adecuadas propiedades psicométricas en diversos grupos etarios de población colombiana no clínica. Esta versión ha sido una de las más aplicadas en forma individual en diversos estudios de evaluación neuropsicológica en todo el mundo de habla hispana, principalmente junto a otras pruebas y no de manera aislada.

Con respecto a confiabilidad, el test mide con precisión en el tiempo y se confirman los hallazgos de estudios previos como el de Henao-Arboleda et al. (2010) en Medellín con valores de coeficiente de correlación intraclase muy similares a los encontrados en la presente investigación para P, C y PC (todos superiores a 0.70). La confiabilidad, usando el método test-retest encontró puntuaciones mayores a las de la adaptación española en P, C y PC, pero más baja en INT, Golden, 1994.

Con relación a la validez de constructo, el análisis de componentes principales confirma una estructura similar a la encontrada por Soares (2009) en Portugal (con un 96.16% de varianza explicada por tres factores con autovalores ma-

yores que 1). En nuestro estudio se encuentran los mismos tres factores, resultado que pudimos confirmar a través de Análisis paralelo de Horn, los cuales explican el 86% de la varianza total. En consecuencia, los resultados permiten confirmar que el constructo subyacente del Stroop, test de colores y palabras de Golden (2007) es el mismo que para el presente estudio.

En cuanto a la validez de convergente, los resultados mostraron una correlación negativa entre las puntuaciones del Stroop (Golden, 2007) y el TMT-B (Reitan, 1992) que indica que entre mayor es el número de ítems correctos en las láminas del Stroop, menor es el tiempo en la tarea del TMT-B.

Sin duda, un hallazgo importante de la presente investigación es que las medias y desviaciones estándar para cada una de las subescalas son inferiores en la población colombiana respecto de la población española, lo que justifica la importancia de estudios de validación que den cuenta de los parámetros en la población en la cual se use el instrumento.

En el desempeño de la prueba se evidenciaron puntuaciones directas que en promedio decrecen con el avance de la edad en P y en C, pero no de manera clara en PC e INT. Por ejemplo, encontramos que en las subescalas PC e INT el grupo etario entre 16 a 44 tienden a puntuar más alto que los demás grupos etarios. A partir de los 45 años estos puntajes tienden a decrecer en estas dos escalas. Respecto a estos resultados, es importante destacar que otros estudios (Wright & Wanley, 2003) resaltan la importancia de la edad que en la infancia es una variable significativa en el rendimiento de las pruebas neuropsicológicas y especialmente las que evalúan habilidades ejecutivas, abstracción y atención. “Es a partir de los 44 años que el rendimiento en las puntuaciones disminuye debido a que éste se ve afectado por la edad y de manera especial en lo relacionado con la atención y las funciones ejecutivas” (Saltohuse, 1996; Wecker et al., 2005; Zahr, 2009, citado por Soares, 2009. La mayoría de las tareas de desempeño disminuyen a medida que aumenta la edad (Perea Bartolomé & Ladera Fernández, 1995).

Además, los estudios previos que hemos citado realizaron normalizaciones y lo hicieron únicamente en algunos rangos de edad, de 7 a 16 (Espinosa, 2008; Agudelo, 2012); de 6 a 12 años en España (Martín, 2012); de 8 a 12 años en Chile (Conca & Ibarra, 2004); de 18 a 49 años (Rognoni et al., 2011) y de 24 a 81 (Van der Elst et al., 2006) por lo que se toman como referentes teóricos, pero dificultan la realización de comparaciones ya que no cumplen las mismas condiciones de edad que el presente estudio.

En este estudio se encuentra que, en general, no hay diferencia significativa entre hombres y mujeres, al igual que en el estudio de la adaptación de la muestra española, el cual encontró diferencias sólo para INT en varones con 3.63 y mujeres 1.88, pero no hubo diferencias significativas para P, C y PC en la variable sexo. Resultados similares fueron hallados por Conca & Ibarra (2004), Espinosa (2008) y Soares (2009). Sin embargo, en el presente estudio se encontró que el Stroop discrimina adecuadamente por estrato y nivel educativo, confirmando hallazgos como los de Perea Bartolomé & Ladera Fernández (1995), Soares (2009) y Van der Elst et al. (2006).

La principal conclusión del presente estudio es que el Stroop, posee adecuadas características de validez y confiabilidad y, en consecuencia, puede ser usado en evaluación neuropsicológica de población colombiana sana teniendo como base los estadísticos descriptivos (medias y desviaciones típicas) de la presente validación y no las normas de la versión española.

Referencias

- Agudelo, E., & Guzmán, C. (2012). *Normalización y Características Psicométricas del Stroop, Test de Colores y Palabras en una muestra de escolares entre 7 y 16 años de la ciudad de Tunja*. Tunja: Uptc.
- Bauer, R.M., Tobias, B. & Valenstein, E. (1993). Amnesic disorders. En Heilman, K.M., & Valenstein, E. (Eds.), *Clinical Neuropsychology* (pp.214-228). New York: Oxford University Press.
- Bausela, E., & Santos, J. (2006). Utilidad Del Stroop en la Psicología Clínica. *Revista Internacional On-line*, 5(1) Recuperado de <http://www.psiquiatria.com/revistas/index.php/asmr/article/view/912/>
- Brett, W., Brown, T., & Onsmann, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 3(8). Recuperado de <http://ro.ecu.edu.au/jephc/vol8/iss3/1>
- Burin, D. I., Drake, M. A., & Harris, P. (2007). *Evaluación neuropsicológica en adultos*. Buenos Aires: Paidós.
- Casal, J. & Mateu, E. (2003). Muestreo por conveniencia. Slideshare. Recuperado de <http://www.slideshare.net/selene1524/muestreo-por-conveniencia#btnPrevious>
- Cerdá, E. (1978). *Psicometría general* [segunda edición]. Barcelona, España: Editorial Herder.
- Chapi, J. (2011). Rendimiento neuropsicológico de personas con esquizofrenia pertenecientes a un programa de rehabilitación integral. *Revista Electrónica de Psicología Iztacala*, 14(4), 136-159. Recuperado de <http://www.medigraphic.com/pdfs/psicologia/epi-2011/epi114h.pdf>
- Comalli Jr., P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology*, 100(1), 47-53.
- Conca Binfa, B., & Ibarra González, M. (2004). *Estandarización de la prueba de colores y palabras de STROOP en niños de 8 a 12 años para la Región Metropolitana*. Tesis de grado inédita. Santiago: Universidad de Chile. Recuperado de <http://repositorio.uchile.cl/handle/2250/113385>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Dodrill, C. B. (1978). A neuropsychological battery for epilepsy. *Epilepsia*, 19(6), 611-623.
- Espinosa, A. (2008). *Validación y estandarización del Test de Colores y Palabras STROOP con una muestra de niños y adolescentes Bogotanos*. Tesis inédita. Bogotá: Universidad Nacional de Colombia.
- Golden, C. J. (2007). *Stroop test de colores y palabras, manual* (5° Ed.). Madrid: TEA Ediciones.
- Golden, J. (2001). *Stroop test de colores y palabras, manual* (3° Ed.). Madrid: TEA Ediciones.
- Golden, C. J. (1994). *Stroop: test de colores y palabras: manual*. Madrid: TEA Ediciones.

- Golden, C. J. (1975). A group version of the Stroop Color and Word Test. *Journal of personality assessment*, 39(4), 386-388.
- Golden, C. J. (1974). Sex differences in performance on the Stroop Color and Word Test. *Perceptual and Motor Skills*, 39(3), 1067-1070.
- Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18, 277-296.
- Henaó-Arboleda, E., Muñoz, C., Aguirre-Acevedo, D. C., Lara, E., Pineda, D. A., & Lopera, F. (2010). Datos normativos de pruebas neuropsicológicas en adultos mayores en una población Colombiana. *Revista Chilena de Neuropsicología*, 5(3), 214-226.
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: common errors and some comments on improved practice. *Educational and Psychological Measurement*, 66, 393-416.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Houston, B. K., & Jones, T. M. (1967). Distraction and Stroop Color Word Performance. *Journal of Experimental Psychology*, 74, 54-56.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- León, F. (2013). *Proyecto Tesis de Maestría: Desarrollo de un modelo de estándares de Prueba para uso de pruebas Psicológicas en Colombia*. Bogotá: Universidad Nacional de Colombia.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163-203.
- Malek, A., Hekmati, I., & Amiri, S. (2013). The Standardization of Victoria Stroop Color-Word Test among Iranian Bilingual Adolescents. *Archives of Iranian Medicine*, 380.
- Malhotra, N. (2004). *Investigación de mercados un enfoque aplicado* (cuarta edición). México: Pearson educación.
- Martín, R., Hernández, S., Rodríguez, C., García, E., Díaz, A., & Jiménez, J. E. (2012). Datos normativos para el Test de Stroop: patrón de desarrollo de la inhibición y formas alternativas para su evaluación. *European Journal of Education & Psychology*, 5(1), 39-51.
- Maureira, F., Aravena, C., Gálvez, C., & Flores, E. (2014). Propiedades Psicométricas y datos normativos del test de stroop y del test torre de Hanoi en estudiantes de educación física de Chile. *Rev GPU*, 10(3), 344-349.
- Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Norman, G.R. & Streiner D.L. (1996). *Bioestadística*. Madrid: Mosby-Doyma libros
- Ostrosky F, Ardila A, Rosselli M (1991). *Evaluación neuropsicológica breve en español-NEUROPSI*. México: Publingenio S.A.
- Partington, J.E. & Leiter, R.G.(1949). Partington's Pathway Test. *The Psychological Service Center Bulletin*. 1, 9-20
- Perea Bartolomé, M., & Ladera Fernández, V. (1995). Rendimientos neuropsicológicos: edad, educación y sexo. *Psicothema*, 7(1), 105-112.
- Pérez, M. (2009). *Manual de Neuropsicología Clínica*. Madrid: Pirámide.
- Portellano, A. (2009). *ENFEN Evaluación Neuropsicológica de las Funciones Ejecutivas en Niños*. Madrid: TEA Ediciones.
- Regard, M. (1981). *Cognitive rigidity and flexibility: A neuropsychological study*. Unpublished Ph.D. dissertation, University of Victoria
- Reitan, R. M. (1992). *Trail Making Test Manual for Administration and Scoring*. Tucson: Reitan Neuropsychology Laboratory.
- Rognoni, M. T., Casals-Coll, G., Sánchez-Benavides, M., Quintana, R. M., Manero, L., Calvo, R., Palomo, F., Aranciva, F., Tamayo, J., & Peña-Casanova. (2013). Spanish normative studies in young adults (NEURONORMA young adults project): Norms for Stroop Color-Word Interference and Tower of London-Drexel University tests. *Neurología (English Edition)*, 28(2), 73-80. Recuperado de <http://www.elsevier.es/es-revista-revista-generica-295-articulo-estudios-normativos-espanoles-poblacion-adulta-90193606>
- Ruiz, M. A., & San Martín, R. (1992). Una simulación sobre el comportamiento de la regla K1 en la estimación del número de factores [The behavior of the

- K1 rule estimating the number of factors: A study with simulated data]. *Psicothema*, 4(2), 543-550.
- Serra, A. & Peña, J. (2006). Fiabilidad test-retest e interevaluador del Test Barcelona. *Neurología*, 21(6), 277-281.
- Soares, S. (2009). *Adaptación del test de colores y palabras de stroop en una muestra portuguesa. Influencia de la reserva cognitiva en la función ejecutiva de sujetos sanos y con enfermedad tipo alzheimer de inicio tardío*. Tesis inédita. Salamanca: Universidad Salamanca.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643-662.
- Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Stroop Color-Word Test influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, 13(1), 62-79.
- Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. En R. Cudeck, & R. C. MacCallum (Eds.). *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: LEA.
- Wright, B. C., & Wanley, A. (2003). Adults' versus children's performance on the Stroop task: Interference and facilitation. *British Journal of Psychology*, 94(4), 475-485.