

# Propiedades psicométricas de una tarea computarizada Stroop emocional caras y palabras

## Psychometric properties of a computerized emotional face-word Stroop task

### Propiedades psicométricas de una tarefa Stroop emocional computadorizada para rostos e palavras



ISSN 0124-0137  
EISSN 2027-212X

ARTÍCULO DE  
INVESTIGACIÓN  
Copyright © 2023  
by PsicoGente

#### Correspondencia de autores:

martha.jimenez@uptc.edu.co  
sergio.gonzalez02@uptc.edu.co  
angie.orozco01@uptc.edu.co

Recibido: 10-11-21  
Aceptado: 17-08-22  
Publicado: 01-01-23

Martha Cecilia Jiménez-Martínez , Sergio David González Caro ,  
Angie Nataly Orozco-Siabatto   
Universidad Pedagógica y Tecnológica de Colombia. Tunja- Boyacá, Colombia

#### Resumen

**Objetivo:** Determinar las propiedades psicométricas de una tarea computarizada de Stroop de reconocimiento emocional de caras y palabras en población colombiana.

**Método:** La investigación fue de naturaleza cuantitativa, con un diseño de investigación instrumental de tipo psicométrico y de desarrollo tecnológico. La muestra se seleccionó por conveniencia, estuvo conformada por 1172 participantes (56,6 % mujeres, 43,4 % hombres), con edades entre 18 a 25 años quienes ingresaban a un dominio para contestar la tarea después de leer y aceptar el consentimiento.

**Resultados:** Los resultados obtenidos demuestran una consistencia interna favorable con un KR-20=0,853 y un coeficiente Omega de McDonald=0,859, una validez de constructo, a partir de un análisis factorial exploratorio y análisis factorial confirmatorio, que presenta una estructura de seis factores y explica un 50,32 % de la varianza; además, de una validez convergente con un coeficiente de correlación Spearman positivo de 0,522 %; los índices de confiabilidad y validez obtenidos en el presente estudio hacen de esta versión computarizada una prueba que puede ser usada en el ámbito de la evaluación de procesos de atención, memoria de trabajo e inhibición en población adulta colombiana no clínica.

**Conclusiones:** En conclusión, se observó efectivamente que la configuración del instrumento psicológico en un software informático optimiza la recolección de datos, pues hay más precisión en la medición de latencias de respuestas, así mismo facilita el análisis de los datos.

**Palabras clave:** Stroop, reconocimiento facial emocional, validación y propiedades psicométricas.

#### Abstrac

**Objective:** To determine the psychometric properties of a computerized Stroop task of emotional recognition of faces and words in a Colombian population.

**Method:** The research was quantitative in nature with an instrumental design of a psychometric type and technological development. The sample was selected for convenience. It was made up of 1172 participants (56,6 % women, 43 % men) between 18 and 25 years of age. They entered a domain to answer the task after reading and giving consent.

**Results:** The results showed a favorable internal consistency with a KR-20=0,853 and a McDonald's Omega coefficient=0,859. A construct validity was obtained from an exploratory factor analysis and a confirmatory factor analysis which presents a structure of six factors and explains 50,32 % of variance. In addition, it presents a convergent validity with a positive Spearman correlation coefficient of 0,522. The reliability and validity indices obtained in the present study make this computerized version a test that can be used in the field of evaluation of processes of attention, working memory and inhibition in non-clinical Colombian adult population.

**Discussions:** In conclusion, it was effectively observed that the configuration of the psychological instrument in computer software optimizes data collection since there is more precision in the measurement of response latencies, as well as facilitating data analysis.

#### Cómo citar este artículo (APA):

Jiménez-Martínez, M.C., González Caro, S.D. y Orozco-Siabatto, A.N. 2023. Propiedades psicométricas de una tarea computarizada Stroop emocional caras y palabras. *PsicoGente* 26(49), 1-25. <https://doi.org/10.17081/psico.26.49.5471>

**Keywords:** Stroop, facial emotion recognition, validation and psychometric properties.

### Resumo

**Objetivo:** Determinar as propriedades psicométricas de uma tarefa Stroop computadorizada de reconhecimento emocional de rostos e palavras em uma população colombiana.

**Metologia:** A pesquisa era de natureza quantitativa, com um projeto de pesquisa instrumental de tipo psicométrico e de desenvolvimento tecnológico. A amostra foi selecionada por conveniência e consistiu de 1172 participantes (56,6 % mulheres, 43,4 % homens), com idades entre 18 e 25 anos, que entraram em um domínio para responder à tarefa após a leitura e aceitação do termo de consentimento.

**Resultados:** Os resultados obtidos mostram uma consistência interna favorável com um coeficiente KR-20=0,853 e McDonald's Omega = 0,859, A validade da construção foi obtida a partir de uma análise exploratória de fatores e análise confirmativa de Fatores, que apresenta uma estrutura de seis fatores e explica 50,32 % da variação, e validade convergente com um coeficiente de correlação Spearman positivo de 0,522. Os índices de confiabilidade e validade obtidos no presente estudo fazem desta versão computadorizada um teste que pode ser usado na avaliação da atenção, memória de trabalho e processos de inibição na população adulta colombiana não-clínica.

**Conclusões:** Em conclusão, foi realmente observado que a configuração do instrumento psicológico em um software de computador otimiza a coleta de dados, pois há maior precisão na medição das latências de resposta, além de facilitar a análise dos dados.

**Palavras-chave:** Stroop, reconhecimento facial emocional, validação e propriedades psicométricas.

## 1. INTRODUCCIÓN

Los test y pruebas psicológicas en los últimos años se han considerado una de las herramientas más importantes y útiles para la investigación y la práctica profesional (Ramos, 2016). Por esto, es que la psicología requiere instrumentos de medición que permitan a los profesionales obtener información más precisa y objetiva, con la que puedan realizar una correcta evaluación y diagnóstico sobre las personas y grupos que estudian, con el fin de orientar y desarrollar estrategias eficaces para sus tratamientos. Esto es posible si se cuenta con instrumentos válidos y confiables, que ofrezcan una información precisa, pues de esta manera, como refiere Malo (2008), podrán hacerse mediciones adecuadas de las distintas variables psicológicas.

Así mismo, la medición cumple unas funciones esenciales en la investigación, que hace posible la descripción, la predicción, la inferencia, la contrastación de hipótesis y, además, establece y estudia relaciones entre variables (Meneses, 2013). Cuando la medición es sobre variables psicológicas se encuentra un mayor desafío, por ende, se debe demostrar empíricamente que el instrumento evalúa aquello para lo que se diseñó (Hernández, 2008). Por tanto, se apela a la comprobación de dos aspectos fundamentales de cualquier instrumento de medición: la validez y la confiabilidad. La validez (de contenido, convergente) se encuentra vinculada al criterio de equidad, que busca minimizar la varianza relacionada con las características individuales y contextos de aplicación de la prueba (Standards for Educational and Psychological, 2014), se plantea que la validez forma parte de un modelo

de medida tripartita de estandarización de pruebas, en la que se incluye: la validez, confiabilidad y equidad; esta última se considera una nueva propiedad psicométrica que debe revisarse si se pretende contar con un instrumento de medición apropiado (Ventura, 2016).

En cuanto a la medición de consistencia interna (confiabilidad), Cronbach (1951) desarrolló el coeficiente alfa de Cronbach ( $\alpha$ ), en donde solo es necesario una aplicación de la prueba siendo semejante a la medición de correlaciones por mitades, corregidas con la fórmula de Spearman-Brown (Quero, 2010). Este método se utiliza para pruebas que contienen reactivos de respuesta múltiple, y lo que se busca es determinar qué tan similar es el conjunto de datos analizado teniendo así, que entre más cercanos a 1 tienden a ser similares y entre más cercanos a 0 menos similares; no obstante, debe tomarse en consideración, que cuando el valor del coeficiente alfa es demasiado alto, como por ejemplo 0,95, puede ser debido a presencia de superfluidez entre los reactivos.

Otra forma de establecer la consistencia interna de una prueba es a través del coeficiente KR-20, desarrollado por Kuder y Richardson (1937) para estimar la confiabilidad de una medición; este se interpreta igual que la del coeficiente alfa de Cronbach para pruebas con tipo de respuesta dicotómica (Reidl, 2013). El coeficiente que se obtiene generalmente es menor que el de la confiabilidad por mitades, siempre y cuando los ítems no sean muy homogéneos.

Por lo tanto, se hace necesaria la construcción y desarrollo de instrumentos psicológicos que cumplan con los criterios de calidad técnica que permitan, a su vez, garantizar tanto la validez como la confiabilidad de los mismos, de manera que las prácticas en psicología se realicen con calidad y de forma ética y profesional y, no se recurra al uso indiscriminado de instrumentos que no estén adaptados y normalizados a una población (León, 2013, como se citó en Rodríguez *et al.*, 2016).

En este sentido, y revisando particularmente la literatura colombiana relacionada con el test de Stroop, cuyo propósito es evaluar la capacidad de control de una interferencia automática sobre un proceso controlado y voluntario en sus diferentes variaciones (Stroop, 1935), solo se cuenta con una versión validada de lápiz y papel del Stroop color palabra (Rodríguez *et al.*, 2016) y dos validaciones del Stroop emocional (Henaó *et al.*, 2010; Olea *et al.*, 2010; Toro *et al.*, 2017); no obstante, no se encuentran versiones compu-

tarizadas del test, que permitan estudiar el procesamiento de la información, la atención selectiva, la memoria y la inhibición atencional en población sana y en personas con diferentes psicopatologías en Colombia de manera más rápida, objetiva y masiva.

Una de las variaciones del efecto Stroop que ha tenido mayor uso en la última década, ha sido el efecto Stroop caras-palabras o word-face, porque posibilita el estudio de la interferencia causada entre el reconocimiento de expresiones afectivas y la lectura de palabras relacionadas con emociones, ha resultado útil para evaluar la depresión (Strand *et al.*, 2013), actitudes afectivas (Cothran *et al.*, 2012), interacciones emocionales multimodales (Ignacio *et al.*, 2019) y el juicio de las expresiones faciales influenciado por palabras emocionales (Lee *et al.*, 2019), el deterioro cognitivo en adultos mayores (Ros *et al.*, 2021) y la influencia del contexto en la producción de expresiones faciales deliberadas (Katembu *et al.*, 2022). Başgöze *et al.* (2015) reportan que la tarea Stroop word-face se ha administrado en varios idiomas como sueco, inglés, turco, chino, noruego y alemán, tanto en personas sanas como en personas con algún tipo de trastorno; también ha demostrado efectividad en la evaluación de la atención y la inhibición. Sin embargo, no se encuentra una versión disponible de esta tarea en el idioma español, lo cual puede ser una limitante toda vez que este es el segundo idioma con mayor número de hablantes a nivel mundial, dejando así sin la oportunidad de evaluación de estos procesos mediante la prueba stroop a la población hispanohablante (Lyons, 2017). En este sentido, en la presente investigación, se hipotetiza que el instrumento por medio del AFC comprueba la existencia de las seis fases principales propuestas en el instrumento (Línea base caras, Reconocimiento facial congruente, Reconocimiento Facial incongruente, línea base lectura, lectura congruente y lectura incongruente caras), que dé como resultado un modelo de seis factores, que sean equivalentes a las seis fases propuestas en la tarea computarizada.

En la actualidad gracias a la tecnología, los métodos tradicionales de la evaluación y obtención de datos se han venido reemplazando en diversos campos de la psicología, un fenómeno de alto impacto que ha introducido grandes avances en la psicometría y en las pruebas psicológicas (Olea *et al.*, 2010; Muñoz *et al.*, 2021), mostrando la necesidad de contar con instrumentos de medición estandarizados y computarizados que ofrecen un mayor control y precisión de los ítems a través del monitor, que permiten la adminis-

tracción de ítems complejos o interactivos, y que registran con mayor exactitud tanto el tiempo, como los procesos de respuesta (Sierra *et al.*, 2007).

Igualmente facilitan la corrección y almacenamiento de las respuestas (Olea y Ponsoda, 2013) y hacen posible la devolución rápida y eficaz de los resultados (Lozzia *et al.*, 2009); bondades con las que no cuentan los tradicionales instrumentos de lápiz y papel.

Adicionalmente, el contexto suscitado por la COVID-19, ha propiciado un reto para la evaluación psicológica en condición de “atención remota”, que demanda con urgencia contar con instrumentos válidos y confiables que puedan ser aplicados de manera virtual. Con esta investigación se busca proveer una versión computarizada en el idioma español del test Stroop caras-palabras que posibilite evaluar procesos ejecutivos, atencionales, emocionales y de inhibición en contextos clínicos, educativos y neuropsicológicos, debido a que el instrumento podría potencialmente facilitar los procesos evaluativos al proporcionar datos más precisos. Permitiendo que la evaluación sea más objetiva, se reduzcan costos en la aplicación y se disminuyan los errores en la calificación resultante de fallas atencionales por parte de los evaluadores.

## 2. MÉTODO

### 2.1. Participantes

La muestra estuvo compuesta por 1172 adultos jóvenes, 505 hombres (43,1 %) y 667 mujeres (56,9 %) con edades entre los 18 y 25 años ( $M=20,95$   $DE: 2,271$ ). Los participantes antes de ingresar a la prueba contestaban un consentimiento informado, así como un cuestionario que permitiera saber: su edad; el nivel educativo; el estrato socioeconómico; si había sido diagnosticado con alguna patología que afecte su motricidad, visión o funciones cognitivas en donde al marcar como afirmativo en alguna se solicitaba especificar cuál; igualmente permitía conocer si había descansado bien en la última semana, además de preguntar si estaba consumiendo algún tipo de medicamento y en dado caso especificar el nombre.

### 2.2. Instrumentos o técnicas

#### 2.2.1. Stroop palabra-color Victoria

La tarea Stroop Victoria (Mueller y Piper, 2014) es una tarea de control inhibitorio, basada en el paradigma Stroop clásico, pero desarrollada en una

versión informatizada. Este instrumento consta de tres fases, cada una con 26 estímulos. En la primera fase, la persona debe presionar los botones que indiquen el color correcto (azul, rojo, amarillo y verde) de los círculos presentados en el monitor. En la segunda fase, el participante debe presionar el botón correspondiente al color de la tinta en la cual está escrita la palabra, dichas palabras son diferentes del color por lo que no genera algún efecto de interferencia. En la tercera fase, el individuo debe inhibir la tendencia de respuesta automática, pues el color de la tinta de las palabras no coincide con el significado (Ramos, 2016). La validación de este instrumento se hizo por juicio de expertos en una población de estudiantes ecuatoriana se obtuvo un alfa de Cronbach de  $\alpha = 0,949$ .

### 2.2.2. Stroop cara-palabra (word-face)

Está basado en el principio del paradigma Stroop clásico (Stroop, 1935), este instrumento está compuesto principalmente por dos estímulos, una palabra referente a una emoción de felicidad o tristeza escrita en tinta negra y un rostro que expresa las dos emociones mencionadas, el participante tendrá que presionar una tecla (1 o 2), según la indicación dada, ya sea responder a la imagen o la palabra. Las imágenes usadas pertenecen a la base de datos de Minear & Park (2004), de igual manera la estructura del instrumento está basada en la prueba original propuesta por Agustí *et al.* (2017), autores que nos proporcionaron la base de imágenes para el presente estudio.

Para este proyecto se programó la tarea en el *software* Unity. En cuanto a su estructura, el test está dividido en ocho fases y compuesto por 42 ensayos en total (ver Figura 1), cada uno con un tiempo máximo de respuesta de 3000 milisegundos (ms). En las cuatro primeras fases, el participante deberá contestar teniendo en cuenta el rostro, estas son: línea de base caras con ocho ensayos, práctica de reconocimiento fácil con dos ensayos, reconocimiento facial congruente (cara y palabra referidas a la misma emoción) y reconocimiento facial incongruente (cara y palabra no coinciden); por otra parte, en las cuatro fases restantes se deberá contestar a partir de la palabra escrita, estas son: línea base lectura, práctica de lectura, lectura congruente y lectura incongruente. Cabe aclarar, que los ensayos de práctica tanto de reconocimiento facial como de lectura no se tienen en cuenta para los análisis estadísticos.



tarea a estudiantes por correo masivos gracias a la colaboración de docentes y estudiantes de diversas universidades del país; dentro del mensaje se invitaba a compartir el link con personas alfabetas adultas. Todos los participantes de este estudio accedieron de manera voluntaria a diligenciar el instrumento. En la primera parte del instrumento computarizado se propuso el consentimiento informado, luego se dispuso un apartado para diligenciar los datos sociodemográficos y un cuestionario en el que se realizaban seis preguntas para evidenciar si el participante sufría algún trastorno psicológico, si estaba con medicación, defecto visual o si presentaba alguna dificultad en el reconocimiento emocional. Al terminar, el participante accede a la tarea Stroop de reconocimiento emocional de caras y palabras (word-face), seguido por el Stroop Victoria. El tiempo total empleado para diligenciar los dos test es de 8 minutos aproximadamente.

### 3.1. Análisis de datos

El plan de análisis comenzó con una descripción de la muestra a partir de las variables sociodemográficas recabadas. Para establecer la normalidad de los datos se aplicó la prueba Kolmogorov-Smirnov (K-S), de igual manera, para determinar la validez de constructo del Stroop de reconocimiento emocional de caras y palabras, se realizó un análisis factorial exploratorio (AFE) dividiendo la muestra a la mitad de manera aleatoria (AFE) y un análisis factorial confirmatorio (AFC) con 586 participantes cada uno, también se corrió la prueba de Friedman y Wilcoxon para identificar las diferencias de latencias de respuesta entre ensayos congruentes e incongruentes. Así mismo, para hallar la validez convergente se aplicó el estadístico de coeficiente de correlación de Spearman y un análisis de contingencia para observar si la tarea computarizada presentaba validez convergente con el Stroop Victoria. Para determinar la confiabilidad de la prueba se halló el coeficiente KR-20 y el Omega de McDonald (1999). Los análisis se hicieron, por medio de los software estadísticos SPSS versión 28 y Jamovi.

## 4. RESULTADOS

A nivel descriptivo se encontró en cuanto a la actividad académica, que el 82,3 % cursa estudios universitarios, 5,2 % cursa un tecnólogo, 4,5 % un posgrado, 3,1 % un técnico, 3 % bachillerato y 2 % realizan otra actividad. Respecto al estrato socioeconómico, se evidenció que el 32,4 % de los participantes se encuentran en estrato uno, 54,2 % en estrato dos, 7,2 % en estrato

tres, 4,8 % en estrato cuatro, 1 % en estrato cinco y 0.3 % en estrato seis. Los participantes provienen de distintos departamentos del país, dentro de los cuales están Boyacá, Casanare, Santander, Cundinamarca, Tolima, Norte de Santander, Guaviare, Caldas, Bolívar, Meta, Antioquia.

#### 4.1. Prueba de normalidad

Para identificar el supuesto de normalidad de los datos del instrumento, se usó la prueba K-S y se encontraron valores de  $p < 0,05$ , que demostraron que los datos obtenidos no tenían una distribución normal (Kim, 2013).

#### 4.2. Validez de constructo

Inicialmente se llevó a cabo el AFE, comúnmente usado para evidenciar la validez de constructo de los instrumentos psicológicos, este análisis comprende un conjunto de estadísticos multivariados utilizados para indicar si los ítems de un cuestionario están altamente correlacionados y pueden ser agrupados en un número menor de factores (Hernández *et al.*, 2008). Antes de realizar un AFE es necesario estimar dos medidas, el índice de Kaiser, Meyer y Olkin (KMO) el cual debe asumir valores por encima de 0,7 y la prueba de esfericidad de Bartlett que debe tener valores significativos ( $p < 0,05$ ), de manera que se pueda verificar la viabilidad de la aplicación del análisis factorial (Zamora *et al.*, 2009). En este sentido, los resultados para el instrumento Stroop de reconocimiento emocional caras y palabras arrojaron un índice de Kaiser, Meyer y Olkin ( $KMO = 0,869$ ) y unos valores en la prueba de esfericidad de Bartlett de ( $X^2 = 16946$ ;  $p < 0,001$ ), lo que permitió evidenciar la posibilidad de realizar el AFE.

En este caso, se aplicó el método de extracción de mínimos residuales, “que es un método de estimación de parámetros para el que no está establecido que las variables deban seguir una distribución determinada” (Morata *et al.*, 2015, p.80); además, se recomienda su uso cuando no se cumple con el supuesto de normalidad y cuando existen variables categóricas (Jiménez y García, 2005). Igualmente, se utilizó la rotación Oblimin, una de las más recomendadas en la actualidad, independientemente del modelo teórico del que se parta (Lloret *et al.*, 2014).

A partir de este procedimiento se obtuvo un total de seis factores, el filtro aplicado sobre las cargas factoriales presentó como resultado la estructura expuesta a continuación (Tabla 1).

**Tabla 1.**  
Carga factorial

|             | 1     | 2     | 3     | 4     | 5     | 6     | UNICIDAD |
|-------------|-------|-------|-------|-------|-------|-------|----------|
| LBRostro1   |       |       | 0,721 |       |       |       | 0,524    |
| LBRostro2   |       |       | 0,911 |       |       |       | 0,256    |
| LBRostro3   |       |       | 0,821 |       |       |       | 0,185    |
| LBRostro4   |       |       | 0,643 |       |       |       | 0,358    |
| LBRostro5   |       |       | 0,548 |       |       |       | 0,403    |
| ImRostro2   |       |       |       |       | 0,493 |       | 0,725    |
| ImRostro3   |       |       |       |       | 0,773 |       | 0,434    |
| ImRostro4   |       |       |       |       | 0,669 |       | 0,524    |
| ImRostro5   |       |       |       |       | 0,469 |       | 0,718    |
| ImRostro6   |       | 0,593 |       |       |       |       | 0,600    |
| ImRostro7   |       | 0,765 |       |       |       |       | 0,402    |
| ImRostro8   |       | 0,798 |       |       |       |       | 0,353    |
| ImRostro9   |       | 0,895 |       |       |       |       | 0,204    |
| ImRostro10  |       | 0,757 |       |       |       |       | 0,427    |
| LBPalabra2  |       |       |       | 0,593 |       |       | 0,623    |
| LBPalabra3  |       |       |       | 0,807 |       |       | 0,333    |
| LBPalabra4  |       |       |       | 0,841 |       |       | 0,317    |
| LBPalabra5  |       |       |       | 0,614 |       |       | 0,600    |
| LBPalabra7  |       |       |       | 0,388 |       |       | 0,774    |
| ImPalabra2  |       |       |       |       |       | 0,395 | 0,761    |
| ImPalabra3  |       |       |       |       |       | 0,694 | 0,519    |
| ImPalabra4  |       |       |       |       |       | 0,500 | 0,689    |
| ImPalabra6  | 0,852 |       |       |       |       |       | 0,258    |
| ImPalabra7  | 0,810 |       |       |       |       |       | 0,351    |
| ImPalabra8  | 0,857 |       |       |       |       |       | 0,252    |
| ImPalabra9  | 0,887 |       |       |       |       |       | 0,220    |
| ImPalabra10 | 0,861 |       |       |       |       |       | 0,258    |

Nota. Método de extracción "mínimos residuales" y rotación "Oblimin".

"LBRostro" hace referencia a los ensayos de línea base de reconocimiento facial, "ImRostro" a las imágenes de los ensayos de reconocimiento facial, "LBPalabra" a la línea base de lectura y "ImPalabra" a las imágenes de los ensayos de lectura.

Como puede observarse, la estructura cuenta con seis factores que se relacionan con las fases propuestas en la estructura inicial del instrumento Stroop reconocimiento emocional caras y palabras. En el procedimiento del análisis factorial se eliminaron los ensayos LBRostro6, LBRostro7, LBRostro8, ImRostro1, LBPalabra1, LBPalabra8 y el ensayo ImPalabra1 porque cargaban en dos factores y generaban ambigüedad. Por otra parte, se encontró que el modelo resultante explica el 50,32 % de la varianza total (Tabla 2).

**Tabla 2.**

Varianza explicada por factor.

| FACTOR | SS CARGADO | % VARIANZA | % ACUMULADO |
|--------|------------|------------|-------------|
| 1      | 3,97       | 13,23      | 13,2        |
| 2      | 3,00       | 10,01      | 23,3        |
| 3      | 3,01       | 10,02      | 33,3        |
| 4      | 2,39       | 7,97       | 41,2        |
| 5      | 1,64       | 5,46       | 46,7        |
| 6      | 1,45       | 3,62       | 50,32       |

Nota. En esta tabla se explica la varianza por cada factor y la varianza total de la prueba. Este modelo explica un total de 50.32 % de la varianza.

Posteriormente, se realizó un análisis factorial confirmatorio (AFC), para determinar si el número de factores obtenidos y sus cargas correspondían con el modelo teórico propuesto previamente (Ferrando & Anguiano-Carrasco, 2010). Al iniciar este proceso se estableció un modelo con seis factores, que hacían referencia a las seis fases propuestas en la estructura inicial, para ello se incluyeron los treinta y seis ensayos establecidos inicialmente. Sin embargo, al determinar los índices de bondad de ajuste, se encontró que sus valores no estaban dentro de los rangos requeridos, no obstante, al contrastar estos hallazgos con la literatura fue posible identificar que estos índices son sensibles al tamaño de la muestra, por lo que en muestras grandes generalmente asumen valores no adecuados (Miranda-Zapata, *et al.*, 2014). En este sentido, se procedió a analizar los índices más comunes de medidas incrementales (CFI, TLI, SRMR Y RMSEA) para observar si el instrumento se ajustaba. Se observó que los valores obtenidos no estaban dentro de lo requerido, indicando así que el modelo inicial propuesto no se adecuaba en su totalidad. Por tanto, fue necesario hacer una revisión más rigurosa de los factores (variables latentes), de los ensayos (ítems) y de sus cargas factoriales. En este sentido, se analizaron los índices de modificación para determinar los cambios que debían implementarse en el modelo y así lograr un ajuste (Lloret-Segura, *et al.*, 2014). Luego de este proceso de revisión, se estableció un segundo modelo con los mismos factores, pero esta vez con el número de ensayos establecidos en el AFE (veintisiete) por medio del método de máxima verosimilitud. Encontrando que los índices de medidas incremen-

tales se ajustaron adecuadamente al modelo, señalando así que la estructura del instrumento se relaciona con lo propuesto teóricamente.

**Tabla 3.**

Medidas de ajuste global

| MEDIDA DE BONDAD DE AJUSTE                     | VALORES   | ACEPTABILIDAD OBTENIDA |
|--|-----------|------------------------|
| Chi-cuadrada                                   | $p=0,000$ | Bajo                   |
| Error de aproximación cuadrático medio (RMSEA) | 0,0431    | Aceptable              |
| Índice de error cuadrático medio (RMR)         | 0,0343    | Aceptable              |
| Índice de validación cruzada esperada (ECVI)   | 0,722     | Aceptable              |
| ÍNDICES DE MEDIDAS INCREMENTALES               | VALORES   | ACEPTABILIDAD OBTENIDA |
| Índice de ajuste comparativo (CFI)             | 0,950     | Aceptable              |
| Índice de ajuste no normado (TLI)              | 0,943     | Aceptable              |
| Índice de ajuste normativo (NFI)               | 0,927     | Aceptable              |
| Índice incremental de ajuste (IFI)             | 0,920     | Aceptable              |
| Índice de ajuste relativo (RFI)                | 0,918     | Aceptable              |
| MEDIDAS DE AJUSTE DE LA PARSIMONIA             | VALORES   | ACEPTABILIDAD OBTENIDA |
| PRATIO   | 0,917     | Bajo                   |
| PNFI   | 0,823     | Bajo                   |
| PCFI   | 0,844     | Bajo                   |
| HOELTER  | 340       | Aceptable              |

Nota. En la tabla se muestran los índices de las medidas incrementales usadas para el AFC.

Como se evidencia en la Tabla 3, los cuatro índices de las medidas incrementales asumieron valores que se ajustan a los requerimientos de cada uno. Aunque los índices de bondad de ajuste presentaron puntuaciones un poco alejadas de lo establecido, los resultados de las medidas incrementales muestran que el modelo presenta un ajuste favorable. Por tanto, se decide aceptar el modelo de los seis factores presentado a continuación.

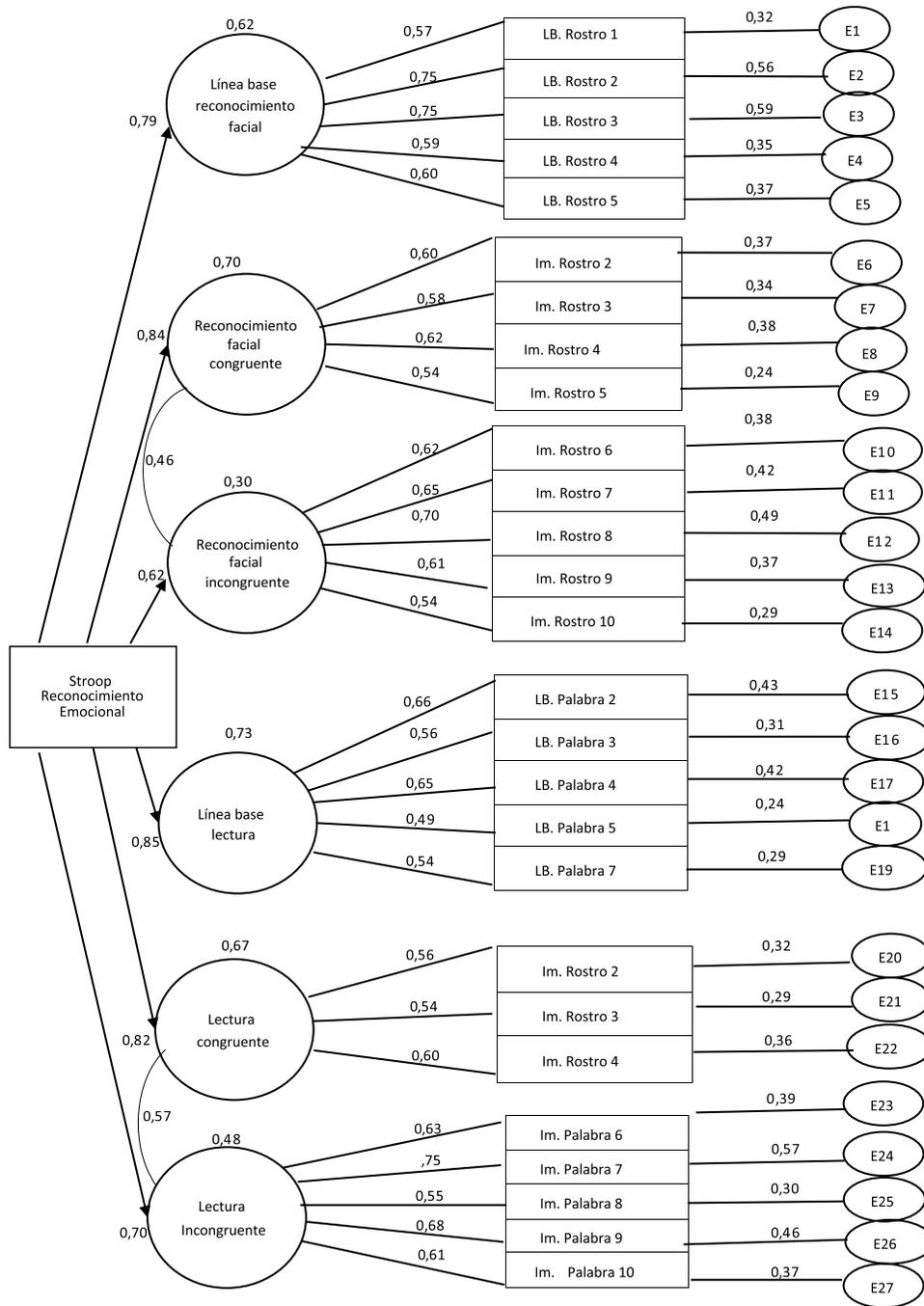


Figura 2.

Modelo confirmatorio de medida de seis factores.

Nota. En la figura se demuestra el modelo de seis factores encontrado en el AFC. Lo representado en óvalos hace referencia a las variables latentes o factores.

Otro hallazgo importante del AFC es la covarianza entre el factor dos y tres (fase de reconocimiento facial congruente y la fase de reconocimiento facial

incongruente) y el factor cinco y seis (fase de lectura congruente y fase de lectura incongruente), cuyos pesos factoriales son 46 % y 57 % respectivamente. Estas relaciones son posibles de explicar por la comunalidad que tiene las fases en su objetivo de respuesta, es decir, que la fase dos y tres comparten la instrucción de responder a la imagen de expresión facial, por otra parte, la fase cinco y seis a la instrucción de lectura de la palabra (respuestas automatizadas). De otra parte, se evidenció que el modelo final de seis factores se asemeja a la estructura del instrumento de reconocimiento emocional facial, propuesto por Rojas (2017). Con los resultados encontrados en el AFE y AFC fue posible establecer que el instrumento tiene validez de constructo, es decir, que evalúa la capacidad de control de una interferencia en el reconocimiento facial y la lectura emocional, lo que significa que la persona, en el desarrollo de la tarea, debe inhibir su respuesta automática ante estímulos incongruentes para tener un desempeño adecuado durante esta.

#### 4.3. Validez convergente

Este tipo de validez facilita estimar la correlación que existe entre los resultados de dos pruebas; por tanto, cuando hay una correlación positiva entre dos instrumentos se deduce que ambas pruebas evalúan el mismo proceso (Acuña *et al.*, 2017). Para hallar este tipo de validez, se usó el coeficiente de correlación de Spearman con el que se encontró una correlación positiva de  $r_s=0,522$  entre las fases de interferencia de las tareas computarizadas: Stroop de reconocimiento emocional de caras y palabras y Stroop Victoria, lo que hace posible inferir que las pruebas efectivamente presentan un efecto Stroop y evalúan el control inhibitorio; no obstante, lo hacen en distintos niveles, dado que en el primero se evalúa dicho control en el reconocimiento de colores y el segundo en el reconocimiento emocional de caras y palabras. En este sentido, es importante resaltar que estos procesos se diferencian claramente en su nivel de activación cerebral (Martínez *et al.*, 2009).

#### 4.4. Confiabilidad

Para la estimación de la confiabilidad se utilizó el coeficiente de KR-20, una forma especial del alfa de Cronbach para ítems dicotómicos. Este evalúa la consistencia interna de los instrumentos psicológicos, es decir, el grado de correlación que existe entre los ítems que componen una escala (Quero, 2010). También se utilizó el coeficiente Omega de McDonald ( $\omega$ ), recomendado como medida alternativa para el cálculo de confiabilidad, en razón de que presenta menos sensibilidad al número de ítems de la escala (Ventura y Caycho, 2017).

Para el caso que nos ocupa, se realizó una primera medición de la confiabilidad teniendo en cuenta los aciertos y errores de los ensayos. En este se obtuvo un valor general de  $KR-20=0,854$  y  $\omega=0,859$ , lo que indica una adecuada consistencia interna (Oviedo y Campo, 2005; Ventura y Caycho, 2017). De esta manera, en los seis factores del instrumento se hizo el análisis de coeficiente KR-20, encontrando en el primer factor  $KR-20=0,836$ , segundo  $KR-20=0,709$ , tercero  $KR-20=0,709$ , cuarto  $KR-20=0,755$ , quinto  $KR-20=0,668$  y el sexto  $KR-20=0,775$ , lo que evidencia una confiabilidad adecuada para cada uno de los factores. También se realizó una medición teniendo en cuenta la latencia de respuesta, allí se obtuvo  $\alpha=0,922$  y  $\omega=0,914$ , esto indica una adecuada consistencia interna (Oviedo y Campo, 2005; Ventura y Caycho, 2017).

Finamente, a manera descriptiva, se comparó el desempeño de los participantes entre ensayos congruentes e incongruentes, tanto en las fases de reconocimiento facial como en las de lectura. Para dar cuenta de esto se utilizó la prueba de Friedman, que es un estadístico no paramétrico que permite comparar tres o más mediciones de muestras relacionadas, con lo cual se identifica si las diferencias de las muestras son estadísticamente significativas (Núñez, 2018).

**Tabla 4.**

Estadísticos descriptivos de las fases congruentes e incongruentes

|        | RECONOCIMIENTO FACIAL CONGRUENTE | RECONOCIMIENTO FACIAL INCONGRUENTE | LECTURA CONGRUENTE | LECTURA INCONGRUENTE |
|--------|----------------------------------|------------------------------------|--------------------|----------------------|
| M      | 1,0803                           | 1,3462                             | 0,9033             | 1,0177               |
| DE     | 0,37962                          | 0,49977                            | 0,32100            | 0,40944              |
| Mínimo | 0,32                             | 0,13                               | 0,12               | 0,13                 |
| Máximo | 3,00                             | 3,00                               | 2,51               | 2,95                 |

Nota. En la tabla se encuentran algunos estadísticos descriptivos de las fases congruentes e incongruentes de la tarea computarizada Stroop de reconocimiento emocional caras y palabras. Los valores están representados en milisegundos.

Según los resultados encontrados en la prueba Friedman, se asume un  $p < 0,05$ , esto indica que, a nivel general, existe una diferencia estadísticamente significativa entre los tiempos de respuesta de los ensayos congruentes e incongruentes de la tarea computarizada. No obstante, este estadístico no especifica entre qué fases se ven las diferencias significativas. Por esto, se consideró pertinente correr la prueba de Wilcoxon para complementar los análisis de significancia (Berlanga y Rubio, 2012).

Los resultados encontrados evidencian la existencia de diferencias estadísticamente significativas entre las cuatro fases de congruencia-incongruencia del instrumento; con ello se comprueba que los ensayos de reconocimiento facial incongruentes tienen un mayor tiempo de latencia que los ensayos congruentes. Del mismo modo, se evidencia que los ensayos incongruentes de lectura presentan una mayor latencia de respuesta frente a los ensayos congruentes. Otro hallazgo importante es que se encuentra un mayor tiempo en la respuesta entre los ensayos incongruentes de reconocimiento facial ( $M=1,3462$ ) en comparación con los de lectura ( $M=1,0177$ ), esto indica la existencia de una mayor interferencia en la fase de reconocimiento facial con ensayos incongruentes. De acuerdo con esto, es posible establecer que el reconocimiento facial emocional es un proceso controlado, que implica un mayor nivel de procesamiento en comparación con el proceso de lectura que es automático.

## 5. DISCUSIÓN

El test Stroop ha sido ampliamente usado en los contextos de evaluación y medición psicológica, tomando especial relevancia en áreas como la psicología clínica, educativa, neuropsicología y psicofisiología (Martín *et al.*, 2012). Esta tarea ha tenido múltiples variaciones desde la creación de su paradigma clásico, con diversos estímulos y entradas sensoriales se han estudiado los distintos mecanismos subyacentes a la interferencia atencional que genera un proceso automático sobre uno controlado (Bauselas y Santos, 2016). Una de las variaciones que ha cobrado mayor importancia en la actualidad es la tarea Stroop de reconocimiento emocional caras y palabras, con la que se ha buscado evaluar la capacidad de reconocimiento emocional facial.

Como lo reportan Başgöze *et al.* (2015), la tarea Stroop caras y palabras suscita la investigación tanto de factores de congruencia e incongruencia así como emocionales de una manera simultánea, por consiguiente, ha incentivado la investigación; bajo este concepto, de trastornos como la depresión (Strand *et al.*, 2013), la ansiedad (Rocher y Pickering, 2017) y el alzheimer (Meléndez *et al.*, 2020). Sumado a esto, también se reportan estudios de neuroimagen que usan esta tarea para examinar las interacciones hemisféricas (Kavcic y Clarke, 2000), la especialización hemisférica (Anes y Kruer, 2004), diferencias de la activación neuronal en el procesamiento de la imagen en contraste de la palabra (Ovaysikia *et al.*, 2011) y coordenadas de redes neuronales en las diferentes variaciones del Stroop (Feng *et al.*, 2018); sin embargo, estas investigaciones tienen una connotación meramente experimental sin reportar propiedades psicométricas.

Como puede verse, aunque son varios los estudios relacionados con este paradigma, es posible identificar que no existe una validación de este instrumento a nivel nacional o internacional en población no clínica. A raíz de ese vacío encontrado en la literatura, nace el propósito de hallar las propiedades psicométricas de la tarea computarizada Stroop de reconocimiento emocional caras y palabras en población adulta joven de Colombia. Para dar respuesta a ese objetivo se realizaron una serie de análisis estadísticos; inicialmente se identificó la validez de contenido del instrumento a través de un análisis factorial exploratorio. Los resultados mostraron la existencia de seis factores principalmente, los cuales se agruparon en el mismo sentido que las seis fases establecidas en la estructura del Stroop de reconocimiento emocional. Luego se procedió a hacer un análisis factorial confirmatorio, para determinar si el número de factores obtenidos y sus cargas correspondían con el modelo teórico propuesto previamente (Ferrando & Anguiano-Carrasco, 2010).

Si bien, inicialmente, el paradigma de Stroop (1935) y la variación de Stroop imagen palabra, realizada por Ehri (1976), contemplaba tres fases (línea base, fase de congruencia y fase de incongruencia) para dar cuenta de la interferencia atencional y la capacidad de inhibición de los sujetos, solo se centraba en la identificación del color de la tinta o de la imagen, pero no incluía la lectura como objetivo de respuesta. Debido a esto, en investigaciones más recientes como la de Başgöze (2015), Agustí *et al.* (2017) y Rojas (2017), se propone una tarea Stroop caras y palabras, con dos tipos de instrucciones; en la primera el sujeto debe responder a la imagen y en la segunda responder a la palabra; no obstante, cada una de estas instrucciones contiene las tres fases propuestas originalmente por Stroop, por lo que el instrumento está compuesto por un total de seis fases; con esta estructura se busca ver la diferencia en las latencias de respuesta entre ensayos congruentes e incongruentes, tanto para el proceso de lectura como para el de reconocimiento facial. La estructura propuesta por los autores pudo ser corroborada estadísticamente gracias al AFE y al AFC en la presente investigación, en la que se comprueba que el modelo teórico se ajusta a las seis fases planteadas de la tarea computarizada Stroop caras y palabras.

En este sentido, los resultados mostraron, en el Stroop caras y palabras, la existencia de seis factores principales, los cuales se agruparon en el mismo sentido que las seis fases establecidas en la estructura del Stroop de reconocimiento emocional planteadas por Başgöze (2015), Agustí *et al.* (2017) y Rojas (2017) y sus cargas corresponden con el modelo teórico propuesto previamente por Ferrando y Anguiano (2010).

En cuanto a la validez convergente, se encontró una correlación positiva entre las tareas computarizadas Stroop de reconocimiento emocional caras y palabra y el Stroop Victoria; con estos resultados es posible deducir que ambos instrumentos evalúan el efecto Stroop que es la capacidad de control de interferencia. Otros aspectos importantes de la comunalidad entre estas dos tareas computarizadas Stroop, es que en su ejecución hay una participación de las funciones ejecutivas. En primera medida, se plantea que en ambos test la memoria de trabajo favorece la retención e implementación de las instrucciones en el desarrollo de cada una de las fases, de igual manera, la flexibilidad cognitiva juega un papel importante en la capacidad de respuesta al cambio de instrucciones de cada fase (Rojas, 2017). Finalmente, el control inhibitorio participa en la ejecución de la prueba, especialmente en las tareas que tienen una interferencia causada por la incongruencia de los estímulos (Martín *et al.*, 2012); sin embargo, esto se mide en distintos procesos y niveles, por un lado, el Stroop de reconocimiento emocional evalúa dicha interferencia en el proceso de reconocimiento de expresiones faciales y, por el otro, el Stroop Victoria en el proceso de identificación de colores (Mueller y Piper, 2014; Agustí *et al.* 2017 y Rojas, 2017).

Desde un punto de vista neurofisiológico, estas diferencias se pueden interpretar por el nivel de activación que presenta una tarea Stroop con caras y palabras, en contraste de la tarea clásica Stroop color palabra, dado que la primera posee elementos visuales con una carga emocional (Rostros con expresiones faciales y palabras con contenido emocional) que requieren un procesamiento visual más robusto y a su vez implica mayor participación de las estructuras cerebrales profundas (Anes y Kruer, 2004). Estas diferencias del nivel de activación, se ven más claramente en los estudios actuales de neuroimagen, pues en estos se evalúan y comparan diversos tipos de Stroop por el comportamiento de las redes neuronales activadas durante la tarea; en dichos estudios se refiere que existe una mayor activación de la corteza cingulada anterior durante la ejecución de una tarea Stroop con caras y palabras que en una tarea Stroop clásica de color palabra, demostrando que implica un mayor procesamiento neuronal en la primer tarea (cara-palabra) que en la segunda (color-palabra), a causa del peso cognitivo implicado, la interpretación de una expresión facial que a su vez puede traducirse en diferencias de desempeños entre las tareas (Ovaysikia *et al.*, 2011 y Feng *et al.*, 2018).

En cuanto al análisis de las latencias de respuesta del instrumento en las fases de congruencia e incongruencia, se encontró una diferencia estadísticamente

significativa entre las distintas fases de la tarea computarizada de Stroop de reconocimientos facial caras y palabras; esto comprobó que tanto los ensayos de reconocimiento facial como los de lectura incongruentes tienen un mayor tiempo de respuesta que los ensayos congruentes. Este resultado alude a la presencia de un conflicto emocional que se produce a raíz de una interferencia desencadenada por la incongruencia entre la expresión facial y la palabra sobrepuesta, dicha interferencia hace que sea necesario un procesamiento cognitivo mayor (Rojas, 2017).

De acuerdo con MacLeod (1991), este hallazgo se puede explicar desde la teoría conexionista propuesta por McClelland, (1988) y Rumelhart *et al.* (1986) en la cual se establece que existe una activación simultánea de información, tanto en la lectura de la palabra como en el reconocimiento facial emocional, las cuales coinciden en un mismo mecanismo de respuesta. Por consiguiente, ambos procesos compiten por alcanzar el umbral atencional; no obstante, cada proceso tiene una fortaleza diferente, pues gracias al entrenamiento la vía del procesamiento de la lectura es más rápida que la del reconocimiento facial emocional.

En consecuencia, en la literatura existe un debate en cuanto al nivel automático de este procesamiento, pues algunos estudios consideran que el reconocimiento facial es un proceso automático y la lectura un proceso controlado, contrario a lo reportado en otras investigaciones como la de Beall y Herbert (2008), quienes proponen que hay efectos de incongruencia más grandes en el proceso de lectura en comparación con el reconocimiento de rostros, sugiriendo que las caras afectivas se procesan más automáticamente que las palabras. Por el contrario, en los estudios de Ovaysikia *et al.* (2011), Agustí *et al.* (2017) y Rojas (2017) se reporta que los tiempos de respuesta de los ensayos que requerían que el participante nombrara la expresión facial eran significativamente más largos que los de lectura. En este orden de ideas, nuestros resultados se relacionan con las últimas propuestas, porque las latencias de respuesta fueron mayores en la fase incongruente de reconocimiento facial que en la de lectura, esto aporta evidencia empírica a que el reconocimiento facial emocional es un proceso controlado y de lectura, un proceso automático.

Por último, es menester mencionar que, durante la revisión de literatura de esta tarea, nos hemos percatado de que la mayor parte de estos estudios usaban un cronómetro manual, hecho que suscita la posibilidad de incurrir en un error en la toma y registro de tiempo, por tanto, es necesario recalcar la importancia de las pruebas computarizadas al momento de la medición

de latencias, pues como lo expresan Molinar *et al.* (2012) y Aguilar (2013), pruebas de este tipo aumentan la objetividad, reducen costos de aplicación, facilitan la medición simultánea de varias variables, reducen el tiempo de aplicación y fomentan las aplicaciones masivas, dando pie a la estandarización de instrumentos con grupos poblacionales más amplios, ya que, de esta manera, se obtiene una medida más precisa de los procesos psicológicos evaluados.

## 6. LIMITACIONES

Para próximas investigaciones, se recomienda hacer aplicaciones masivas con población clínica y no clínica, con el objetivo de realizar un trabajo experimentación y baremación con diferentes grupos etarios para determinar los desempeños en el reconocimiento emocional facial en las distintas etapas del ciclo vital, con el propósito de crear un sistema de calificación válido y confiable de la tarea Stroop emocional caras y palabras para nuestro contexto que incluya todos los grupos etarios.

## 7. CONCLUSIONES

Se concluye que la tarea computarizada de Stroop de reconocimiento emocional caras y palabras cuenta con propiedades psicométricas adecuadas para los adultos jóvenes (18-25 años) de población colombiana no clínica, puesto que muestra una confiabilidad favorable de KR-20 de= 0,853 y  $\omega=0,859$  teniendo en cuenta los aciertos y errores, de igual manera, presenta puntajes de  $\alpha= 0,922$  y  $\omega= 0,914$  por parte de los tiempos de latencia de respuesta lo que indica que su confiabilidad es adecuada. Por su parte, en cuanto a la validez se encontró en el AFE y en el AFC una estructura con los factores correspondientes a las seis fases planteadas desde la teoría y la construcción del instrumento.

Por otra parte, se evidencia que el instrumento guarda el principio básico de interferencia del efecto Stroop, pues en los ensayos de reconocimiento facial incongruente hubo una latencia de respuesta mayor que en las demás fases, lo que indica la existencia de interferencia atencional en la ejecución de esta tarea. Así mismo, se corroboró estadísticamente que el reconocimiento emocional facial es un proceso controlado y por lo tanto requiere más tiempo para su ejecución, a diferencia de la lectura que, al ser un proceso automático, su tiempo de respuesta es menor.

En cuanto a los resultados encontrados en la validez convergente, es posible identificar una correlación positiva pero moderada entre la tarea computa-

rizada Stroop de reconocimiento emocional de caras y palabras y el Stroop Victoria; esto permite deducir por una parte que los dos instrumentos presentan el efecto Stroop (una interferencia atencional) y por otra que en su aplicación requieren de la participación de las funciones ejecutivas. Teniendo en cuenta esto, ambos instrumentos evalúan interferencia, pero con demandas diferentes, dado que, en el Stroop Victoria la interferencia se evalúa en el procesamiento de color y en el Stroop de reconocimiento emocional se evalúa en el reconocimiento a expresiones faciales. En este sentido, se identifica que se procesan a distintos niveles e implican vías neuroanatómicas diferentes.

Se observó efectivamente que la configuración del instrumento psicológico en un software informático optimiza la recolección de datos, pues hay más precisión en la medición de latencias de respuestas, así mismo facilita el análisis de los datos y disminuye los errores humanos en el proceso de calificación.

**Agradecimientos:** Agradecemos a todos los colaboradores que participaron voluntariamente en el diligenciamiento de los instrumentos.

**Financiamiento:** El financiamiento se dio con recursos de los autores

**Conflicto de interés:** No existe conflicto de interés por parte de ninguno de los autores

## REFERENCIAS

- Acuña, I., Michelini, Y., Guzmán, J. y Godoy, J. (2017). Evaluación de validez convergente y discriminante en test computarizados de toma de decisiones. *Avaliação Psicológica*, 16(3), 375-383. <https://www.redalyc.org/pdf/3350/335056639015.pdf>
- Aguilar, J. (2013). Pruebas psicológicas sistematizadas. *PsicoAsiste Asistencia Electrónica de Información Psicológica*, 2-12. [https://www.academia.edu/8113908/Pruebas\\_Psicol%C3%B3gicas\\_Sistematizadas](https://www.academia.edu/8113908/Pruebas_Psicol%C3%B3gicas_Sistematizadas)
- Agustí, A., Satorres, E., Pitaraque, A. & Meléndez, J. (2017). An emotional task with faces and words. A comparison of young and older adults. *Consciousness and Cognition*, 53(1), 99-104. <https://doi.org/10.1016/j.concog.2017.06.010>
- Anes, M. & Kruer, J. L. (2004). Investigating hemispheric specialization in a novel face-word Stroop task. *Brain and Language*, 89(1), 136-141. [https://doi.org/10.1016/S0093-934X\(03\)00311-0](https://doi.org/10.1016/S0093-934X(03)00311-0)
- Başgöze, Z., Gönül, A., Baskak, B. & Gökçaya, D. (2015). Valence-based Word-Face Stroop task reveals differential emotional interference in patients with major depression. *Psychiatry Research*, 229(3), 960-967. <https://doi.org/10.1016/j.psychres.2015.05.099>
- Bauselas, E. & Santos, J. (2016). Utilidad del Stroop en la psicología clínica. *Avances en Salud Mental Relacionada*, 5(1). <https://psiquiatria.com/bibliopsiquis/utilidad-del-stroop-en-la-psicologia-clinica/>

- Beall, P. & Herbert A. (2008). The face wins: Stronger automatic processing of affect in facial expressions than words in a modified Stroop task *Cognition and Emotion*, 22(8), 1613-1642. <https://doi.org/10.1080/02699930801940370>
- Berlanga, V. y Rubio, M. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *Revista d'Innovació i Recerca en Educació*, 5(2), 101-113. <https://www.raco.cat/index.php/REIRE/article/download/255793/342836>
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. [http://psych.colorado.edu/~carey/Courses/PSYC5112/Readings/alpha\\_Cronbach.pdf](http://psych.colorado.edu/~carey/Courses/PSYC5112/Readings/alpha_Cronbach.pdf)
- Cothran, D., Larsen, R., Zelenski, J. & Prizmic, Z. (2012). Do emotion words interfere with processing emotion faces? Stroop-like interference versus automatic vigilance for negative information. *Imagination, Cognition and Personality*, 32(1), 59-73. <https://doi.org/10.2190/IC.32.1.e>
- Ehri, L. (1976). Do words really interfere in naming pictures? *Child Development*, 47(2), 502-505. <https://doi.org/10.2307/1128808>
- Feng, C., Becker, B., Huang, W., Wu, X., Eickhoff, S. B. & Chen, T. (2018). Neural substrates of the emotion-word and emotional counting Stroop tasks in healthy and clinical populations: A meta-analysis of functional brain imaging studies. *NeuroImage*, 173, 258-274. <https://doi.org/10.1016/j.neuroimage.2018.02.023>
- Ferrando, J. y Anguiano, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33. <https://www.redalyc.org/pdf/778/77812441003.pdf>
- Henao Arboleda, E., Muñoz, C., Aguirre-Acevedo, D., Lara, E., Pineda, D. y Lopera, F. (2010). Datos normativos de pruebas neuropsicológicas en adultos mayores en una población colombiana. *Revista Chilena de Neuropsicología*, 5(3), 214-226. <https://www.redalyc.org/pdf/1793/179318868005.pdf>
- Hernández, R., Fernández, C. y Baptista, P. (2008). *Metodología de la Investigación (4a ed.)*. McGraw-Hill Interamericana.
- Ignacio, D., Gerkens, D., Aguinaldo, E., Arch, D. y Barajas, R. (2019). Music-like stimuli affect the emotional processing of words and facial expressions. *Psychology of Music*, 1-19. <https://doi.org/10.1177/0305735619875800>
- Katembu, S., Xu, Q., Rostami, H. N., Recio, G., & Sommer, W. (2022). Effects of Social Context on Deliberate Facial Expressions: Evidence from a Stroop-like Task. *Journal of Nonverbal Behavior*, 1-21. <https://doi.org/10.1007/s10919-022-00400-x>
- Kavcic, V. & Clarke, J. (2000). Hemispheric interactions during a face-word Stroop-analog task. *Neuropsychology*, 14(4), 579-587. <https://doi.org/10.1037/0894-4105.14.4.579>
- Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52-54. <https://doi.org/10.5395/rde.2013.38.1.52>
- Kuder, G.F. y Richardson, M. (1937). La teoría de la estimación de la fiabilidad de las pruebas. *Psychometrika*, 2(3), 151-160. <https://doi.org/10.1007/BF02288391>
- Lee, J., Yang, H. & Lee, D. (2019). Context Modulation Effect by Affective Words Influencing on the Judgment of Facial Emotion. *Korean Journal of the Science of Emotion and Sensibility*, 22(2), 37-48. <https://doi.org/10.14695/KJSOS.2018.22.2.37>
- León, F. (2013). *Desarrollo de un modelo de estándares de Prueba para uso de pruebas Psicológicas en Colombia*. [Proyecto Tesis de Maestría, Universidad Nacional de Colombia]. Citado en Rodríguez, L., Pulido, N. y Pineda, C. (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>
- Lloret, S., Ferreres, A., Hernández, A. y Tomás, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>.
- Lozzia, G., Abal, F., Blum, D., Aguerri, M., Galibert, M. & Attorresi, H. (2009). Tests informatizados: Nuevos desafíos prácticos y éticos para la evaluación psicológica. *Summa Psicológica UST*, 6(1), 135-148. <https://doi.org/10.18774/448x.2009.6.57>
- Lyons, D. (2017). ¿En qué países se habla español y cuántas personas lo hablan? *La Revista de Babel*. <https://es.babel.com/es/magazine/cuantas-personas-hablan-espanol-y-donde-se-habla/>
- McClelland, J. (1988). Modelos conexionistas y evidencia psicológica. *Revista de Memoria y Lenguaje*, 27(2), 107-123. [https://doi.org/10.1016/0749-596X\(88\)90069-1](https://doi.org/10.1016/0749-596X(88)90069-1)
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: Lawrence Erlbaum Associates, Inc.
- MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203. <https://doi.org/10.1037/0033-2909.109.2.163>
- Malo, S. (2008). La medición como herramienta y como reflexión ética en el ejercicio del psicólogo. *Psicogente*, 11(19), 46-51. <http://revistas.unisimon.edu.co/index.php/psicogente/article/view/2688>
- Martín, R., Hernández, S., Rodríguez, C., García, E., Díaz, A. y Jiménez, J. (2012). Datos normativos para el Test de Stroop: patrón de desarrollo de la inhibición y formas alternativas para su evaluación. *European Journal of Education y Psychology*, 5(1), 39-51. <https://www.redalyc.org/pdf/1293/129324775004.pdf>
- Martínez, R., Tuya, L., Martínez, M., Pérez, A. y Cánovas, A. (2009). El coeficiente de correlación de los rangos de Spearman caracterización. *Revista Habanera de Ciencias Médicas*, 8(2). [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1729-519X2009000200017](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1729-519X2009000200017)
- Meléndez, J. C., Satorres, E. & Oliva, I. (2020). Comparing the Effect of Interference on an Emotional Stroop Task in Older Adults with and without Alzheimer's Disease. *Journal of Alzheimer's Disease*, 1-9. <https://pubmed.ncbi.nlm.nih.gov/31929162/>
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L., Turbany, J. y Valero, S. (2013). *Psicometría*. Editorial UOC.
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, y Computers*, 36, 630-633. <http://dx.doi.org/10.3758/BF03206543>.
- Miranda, E., Riquelme, E., Cifuentes, H. y Riquelme, P. (2014). Análisis factorial confirmatorio de la Escala de habilidades sociales en universitarios chilenos. *Revista Latinoamericana de Psicología*, 46(2), 73-82. <https://www.redalyc.org/pdf/805/80532608001.pdf> [https://doi.org/10.1016/S0120-0534\(14\)70010-X](https://doi.org/10.1016/S0120-0534(14)70010-X)
- Molinar, J., Escoto, C., García, R., y Bautista, E. (2012). Evaluación Computarizada de Pruebas Psicológicas mediante el Procesamiento Digital de Imágenes. *Enseñanza e Investigación en Psicología*, 17(2), 415-426. <https://www.redalyc.org/pdf/292/29224159006.pdf>
- Morata-Ramírez, M., Holgado-Tello, F., Barbero-García, I y Méndez, G. (2015). Análisis factorial confirmatorio: recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA. *Acción Psicológica*, 12(1), 79-90. <https://dx.doi.org/doi.org/10.5944/ap.12.1.14362>

- Mueller, S. & Piper, B. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, 222, 250-259. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3897935/>
- Muñoz, J., Torres, M., y Zavala, Y. (2021). Validación de Experimentos Cognitivos con PEBL y Wundts Lab. *Psicología desde el Caribe*, 38(3), 368-392. <https://doi.org/10.14482/psdc.38.3.153>
- Núñez, C. (2018). Análisis de varianza no paramétrica: un punto de vista a favor para utilizarla. *Acta Agrícola y Pecuaria*, 4(3), 69-79. <https://doi.org/10.30973/aap/2018.4.3/1>
- Olea, J. & Ponsoda, V. (2013). *Tests adaptativos informatizados*. Editorial UNED.
- Olea, J., Abad, F. & Barrada, J. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31(1), 97-107. <http://www.papelesdelpsicologo.es/pdf/1800.pdf>
- Ovaysikia, S., Chan J., K. Tahir, K. & DeSouza, J. (2011). Word wins over face: Emotional Stroop effect activates the frontal cortical network. *Frontiers in Human Neuroscience*, 4(1), 234-238. <https://doi.org/10.3389/fnhum.2010.00234>.
- Oviedo, H. y Campo, A. (2005). Metodología de investigación y lectura crítica de estudios. Aproximación al uso del coeficiente Alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572-580. <https://www.redalyc.org/pdf/806/80634409.pdf>
- Quero, M. (2010). Confiabilidad y coeficiente Alfa de Cronbach. *Telos*, 12(2), 248-252. <https://www.redalyc.org/pdf/993/99315569010.pdf>
- Ramos, C. (2016). Adaptación del test Stroop Victoria en estudiantes ecuatorianos. *Revista Iberoamericana de Diagnóstico y Evaluación*, 44(2), 57-64. [https://www.aidep.org/sites/default/files/2017-09/R44-Art5\\_1.pdf](https://www.aidep.org/sites/default/files/2017-09/R44-Art5_1.pdf)
- Reidl, L. (2013). Confiabilidad en la medición. *Investigación en Educación Médica*, 2(6), 107-111. <https://www.redalyc.org/pdf/3497/349733227007.pdf>
- Rocher, A. & Pickering, A. (2017). Trait anxiety, infrequent emotional conflict, and the emotional face Stroop task. *Personality and Individual Differences*, 111, 157-162. <https://doi.org/10.1016/j.paid.2017.02.017>
- Rodríguez, L., Pulido, N. & Pineda, C. (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>
- Rojas, F. (2017). *Adecuación y análisis comportamental del test de reconocimiento de expresiones faciales con conflicto emocional, versión en español (TREFACE-E) en una muestra de estudiantes universitarios* [Tesis de pregrado, Universidad Pontificia Bolivariana, Colombia]. [https://biblioteca.bucaramanga.upb.edu.co/docs/digital\\_35523.pdf](https://biblioteca.bucaramanga.upb.edu.co/docs/digital_35523.pdf)
- Ros, L., Satorres, E., Fernández-Aguilar, L., Delhom, I., López-Torres, J., Latorre, J. M., & Melendez, J. C. (2021). Differential effects of faces and words in cognitive control in older adults with and without major depressive disorder: An emotional Stroop task study. *Applied Neuropsychology: Adult*, 1-10. <https://doi.org/10.1080/023279095.2021.1927037>
- Rumelhart, D., Hinton, G. & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. <https://doi.org/10.1038/323533a0>
- Sierra, F., Valdelamar, J., Hernández, F., & Sarmiento, L. (2007). Test adaptativos informatizados. *Avances en Medición*, 5, 157-162. [http://www.humanas.unal.edu.co/psicometria/files/9513/7036/5476/Test\\_Adaptativos\\_Informatizados.pdf](http://www.humanas.unal.edu.co/psicometria/files/9513/7036/5476/Test_Adaptativos_Informatizados.pdf)
- Strand, M., Oram, M. & Hammar, Å. (2013). Emotional information processing in major depression remission and partial remission: faces come first. *Applied*

*Neuropsychology: Adult*, 20(2), 110-119. <https://doi.org/10.1080/09084282.2012.670159>

- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643-662. <https://doi.org/10.1037/h0054651>
- Toro, R., Avendaño, B., Arias, A., y Sarmiento López, J. (2017). Propiedades psicométricas de un Stroop emocional computarizado para evaluar vulnerabilidad cognitiva. *Revista CES Psicol*, 10(2), 103-115. <http://dx.doi.org/10.21615/cesp.10.2.7>
- Ventura, J. (2016). Breve historia del concepto validez en psicometría. *Revista Peruana de Historia de la Psicología*, 2, 89-92. [https://www.researchgate.net/publication/315110740\\_Breve\\_historia\\_del\\_concepto\\_validez\\_en\\_psicometria](https://www.researchgate.net/publication/315110740_Breve_historia_del_concepto_validez_en_psicometria)
- Ventura, J. y Caycho, T. (2017). El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627. <https://www.redalyc.org/pdf/773/77349627039.pdf>
- Ximénez, M. y García, A. (2005). Comparación de los métodos de estimación de máxima verosimilitud mínimos cuadrados no ponderados, en el análisis factorial confirmatorio mediante simulación de Monte Carlo. *Psicothema*, 15(3), 528-535. <https://www.redalyc.org/pdf/727/72717327.pdf>
- Zamora, S., Monroy, L. y Chávez, C. (2009). *Análisis factorial: una técnica para evaluar la dimensionalidad de las pruebas*. Cuaderno técnico 6. México: Centro Nacional de Evaluación para la Educación Superior, A.C.

**Esta obra está bajo:** Creative commons attribution 4.0 international license. El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas siempre y cuando reconozca y cite la obra de la forma especificada por el autor o el licenciente.

