



Psicología y Ciencias Sociales

Doctorado en Psicología

Tesis doctoral:

**Comparación de dos versiones del Test de Palabras y Colores Stroop**

Presentada por:

**María Luisa Palencia Avendaño**

Dirigida por:

**Jesica Formoso**

Buenos Aires, 2020

### **Agradecimientos**

Llega el momento de dar cierre a esta tesis, ardua tarea que me propuse hace algunos años y que me ha traído alegrías, desencuentros y grandes aprendizajes. Por todo ello me gustaría reconocer y dar las gracias a las personas que han participado directa o indirectamente.

Primero a mi familia, especialmente a mi madre doña Bernarda y mi hermana Ángela, por su amor incondicional, sostén y guía siempre. A mi hermoso hijo Manuel Felipe, por su comprensión y paciencia.

A mis amigos y colegas, por su inspiración, sus risas, sus reflexiones y palabras de aliento. Adriana, Néstor, María Isabel, Darío, a pesar de las distancias siempre presentes. Candela & Emmanuel, Luisa, Carol, Karoll & Sergio, por acogernos en esta ciudad. Y a María Laura, por recorrer a mi lado el camino del doctorado.

A mis amigas y compañeras de LAFUN, por compartir los desafíos del trabajo diario, en especial a Julieta y Selva, entrañables a prueba de todo. Al Dr. Máximo, por la confianza depositada en mí.

A la Dra. Irene Injoque-Ricle, por su acompañamiento y guía en el inicio de este proceso, su orientación siempre acertada y cálida hasta el último momento. Y a la Dra. Jesica Formoso, quien recibió esta tarea con la misma entereza y empeño, dándome el aliento necesario para concluir.

Finalmente, un agradecimiento sincero a los participantes y sus familias, por hacer posible esta investigación.

**Índice de contenido**

	Pág.
<b>INTRODUCCIÓN</b>	11
Vigencia del Stroop	12
Uso del Stroop en Diversos Campos	12
El Stroop como Objeto de Estudio	13
Definición de Términos	14
Propósitos y Alcances de la Investigación	16
Capitulado	17
<b>PARTE I. MARCO TEÓRICO</b>	19
Antecedentes en el Estudio de la Atención	20
Procesos y Componentes de la Atención	21
Sobre la Definición de Atención	22
Aproximación Neurocognitiva	23
Redes Atencionales	23
Orígenes del Test Stroop	27
Observaciones Iniciales	27
El Efecto Stroop	30
El Efecto Stroop como Proceso Cognitivo	31
Modelos del Efecto Stroop	32
Velocidad de Procesamiento	32
Automaticidad	33
Procesamiento Distribuido Paralelo	35
Atención Selectiva	37
Competencia de Información	38
Mecanismos de Control Inhibitorio: Paradigma Stroop	42
Desarrollo del Control Inhibitorio	44
Trastornos de Neurodesarrollo	46
Proceso de Normatización	48
La prueba CWST de Golden (1978)	48
Adaptaciones en Hispanoamérica	49
Variantes en Otros Idiomas	50
Observaciones sobre la Proliferación de Versiones	52

	Pág.
Stroop Emocional, Pictórico, Numérico y Otras Adaptaciones	53
Propiedades Psicométricas del Stroop	56
Hallazgos de Confiabilidad del Stroop	57
Hallazgos de Validez del Stroop	58
Estado del Arte	61
Introducción de Formas Computarizadas	61
Antecedentes de Comparación entre Formas	63
Resultados del Stroop en Grupos Clínicos	65
Líneas de Investigación Actuales	66
Mediciones en el Stroop	68
Puntajes Básicos	68
Parámetros de Latencias	68
Parámetros de Aciertos	71
Parámetros de Errores	73
Métodos de Cuantificación de la Interferencia	74
Índice de Diferencia ( $I_D$ )	74
Índice de Razón ( $I_R$ )	75
Índice de Interferencia de Golden ( $I_G$ )	76
Índices de Tempo Personal ( $I_{READ}$ e $I_{NAME}$ )	77
Otros Índices de Interferencia	78
<b>PARTE II. METODOLOGÍA</b>	<b>81</b>
Planteamiento del Problema	82
Preguntas de Investigación	86
Objetivos	87
Hipótesis	88
Tipo y Diseño de Investigación	89
Determinación de la Muestra	89
Criterios de Inclusión y Exclusión	91
Participantes	91
Instrumentos	93
Stroop Versión Tradicional	93
Operacionalización de Variables	94
Stroop Versión Computarizada	94

	Pág.
Operacionalización de Variables	95
Consideraciones Éticas	96
<b>PARTE III. RESULTADOS</b>	97
Análisis de Datos	98
Desempeño en el Stroop Tradicional	100
Latencias de Respuesta Estimadas en el Stroop Tradicional	100
Contrastes por Factor	101
Efectos e Interacción de los Factores y Análisis Post-Hoc	102
Número de Aciertos en el Stroop Tradicional	104
Contrastes por Factor	105
Efectos e Interacción de los Factores y Análisis Post-Hoc	108
Número de Errores en el Stroop Tradicional	109
Contrastes por Factor	110
Efectos e Interacción de los Factores y Análisis Post-Hoc	111
Índices de Interferencia Calculados en el Stroop Tradicional	114
Sensibilidad y Especificidad	114
Análisis por Edad y Condición Clínica	115
Capacidad Discriminante de la Tarea PC	116
Desempeño en el Stroop Computarizado	118
Latencias Media de Aciertos en el Stroop Computarizado	118
Contrastes por Factor	119
Efectos e Interacción de los Factores y Análisis Post-Hoc	120
Número de Aciertos en el Stroop Computarizado	123
Contrastes por Factor	124
Efectos e Interacción de los Factores y Análisis Post-Hoc	125
Número de Errores en el Stroop Computarizado	127
Contrastes por Factor	128
Efectos e Interacción de los Factores y Análisis Post-Hoc	130
Índices de Interferencia Calculados en el Stroop Computarizado	133
Sensibilidad y Especificidad	133
Análisis por Sexo, Edad y Condición Clínica	134
Capacidad Discriminante de la Tarea PC	134
Stroop Tradicional <i>versus</i> Stroop Computarizado	137

	Pág.
Comparación General de las Mediciones Básicas ( $n=240$ )	137
Comparación entre Grupos Clínicos ( $n=180$ ) y Controles ( $n=60$ )	138
Análisis de Correlaciones de las Mediciones Básicas	142
Análisis de Equivalencia de las Mediciones Básicas	142
Comparación de los Índices de Interferencia	143
Análisis de Correlaciones de los Índices de Interferencia	145
Análisis de Equivalencia de los Índices de Interferencia	145
Capacidad Discriminante de PC en Ambas Versiones	145
<b>PARTE IV. DISCUSIÓN</b>	<b>148</b>
Semejanzas y Diferencias entre las Versiones del Stroop	149
Efecto Stroop	149
Cuestiones Relativas a las Mediciones Básicas	150
Sobre las Latencias	151
Sobre los Aciertos	152
Sobre los Errores	154
Examen de la Equivalencia Considerando las Mediciones Básicas	154
Hallazgos por Sexo	157
Hallazgos por Rango Etario	158
Hallazgos entre Grupos Clínicos y Grupo Control	160
Cuestiones relativas al Cálculo de la Interferencia	162
Hallazgos en $I_D$	162
Hallazgos en $I_R$	163
Hallazgos en $I_V$	164
Hallazgos en $I_{READ}$ , $I_{NAME}$ e $I_{REL}$	165
Hallazgos en $I_G$	165
Examen de la Equivalencia Considerando los Índices Calculados	166
<b>PARTE V. CONCLUSIONES</b>	<b>168</b>
Conclusiones	169
Limitaciones y Prospectiva	171
<b>REFERENCIAS</b>	<b>173</b>
<b>ANEXOS</b>	<b>204</b>

**Índice de figuras**

	Pág.
1. Representación de las áreas corticales y subcorticales implicadas en la atención	25
2. Arquitectura de red del modelo de Cohen, Dunbar y McClelland (1990)	36
3. Modelo de codificación de palabras por activación y verificación de Roelofs (1992)	39
4. Diseño de módulos y redes neuronales de Kaplan et al. (2007)	40
5. Modelo de difusión de Stattford y Gurney (2007)	41
6. Latencias de respuesta estimadas en el Stroop tradicional ( $n=240$ )	100
7. Discriminación de las latencias de respuesta estimadas en la versión tradicional, por rango etario y grupo diagnóstico	103
8. Número de aciertos en el Stroop Tradicional ( $n=240$ )	105
9. Discriminación del número de aciertos en la versión tradicional, por rango etario y grupo diagnóstico	108
10. Número de errores en el Stroop tradicional ( $n=240$ )	110
11. Discriminación del número de errores en la versión tradicional, por rango etario y grupo diagnóstico	112
12. Latencias medias de aciertos en el Stroop computarizado ( $n=240$ )	118
13. Discriminación de las latencias de respuesta en la versión computarizada, por rango etario y grupo diagnóstico	121
14. Número de aciertos en el Stroop computarizado( $n=240$ )	123
15. Discriminación de los aciertos en la versión computarizada, por sexo y grupo diagnóstico	125
16. Discriminación de los aciertos en la versión computarizada, por rango etario y grupo diagnóstico	126
17. Número de errores en el Stroop computarizado ( $n=240$ )	128
18. Discriminación de los errores en la versión computarizada, por sexo y grupo diagnóstico	130
19. Discriminación de los errores en la versión computarizada, por rango etario y grupo diagnóstico	131
20. Comparación del desempeño entre las versiones ( $n=240$ )	132
21. Comparación de latencias en ambas versiones del Stroop	137
22. Comparación de número de aciertos en ambas versiones del Stroop	140
23. Comparación de número de errores en ambas versiones del Stroop	141
24. Puntos de corte para los puntajes brutos obtenidos en la tarea PC de ambas versiones	142

## Índice de tablas

	Pág.
1. Hallazgos de confiabilidad en estudios comparativos entre distintas versiones del Stroop	64
2. Antecedentes de comparación del desempeño en la tarea de conflicto palabra-color (PC) entre grupos clínicos y controles	65
3. Reporte de tiempos de respuesta para las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)	69
4. Reporte de respuestas correctas obtenidas en las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)	72
5. Observación de Golden (1978, 1999) sobre los puntajes <i>T</i> en el Stroop	73
6. Reporte de errores para las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)	74
7. Número y porcentaje parcial de participantes en cada grupo	92
8. Latencias de respuesta estimadas en el Stroop tradicional y comparación por factores	101
9. Número de aciertos en las tareas del Stroop tradicional y comparación por factores	106
10. Número de errores en las tareas del Stroop tradicional y comparación por factores	111
11. Resultados generales y análisis de curvas ROC de los índices de interferencia calculados en el Stroop tradicional	114
12. Análisis de curvas ROC para los índices de interferencia calculados en cada rango etario	116
13. Capacidad de clasificación de las mediciones de la tarea PC en el Stroop tradicional	116
14. Análisis de curvas ROC para las mediciones básicas en la tarea PC en cada rango etario	117
15. Latencias de respuesta en las tareas del Stroop computarizado y comparación por factores	119
16. Aciertos en las tareas del Stroop computarizado y comparación por factores	124
17. Número de errores en las tareas del Stroop computarizado y comparación por factores	128
18. Tipo de error arrojado por el Stroop computarizado en la tarea de conflicto palabra-color	129
19. Resultados generales y análisis de curvas ROC de los índices de interferencia calculados en el Stroop computarizado	133
20. Análisis de curvas ROC para los índices de interferencia calculados considerando los factores analizados	134



21. Capacidad de clasificación de las mediciones de la tarea PC del Stroop computarizado	135
22. Análisis de curvas ROC para las mediciones básicas en la tarea PC del Stroop computarizado, considerando los factores analizados	136
23. Comparación de la capacidad de clasificación de los índices de interferencia	144
24. Comparación de la capacidad de clasificación de las mediciones básicas en la tarea PC	146
25. Ejemplo de puntuaciones de aciertos en las tareas de la versión tradicional del Stroop	154

**Índice de anexos**

	Pág.
1. Formatos de presentación del Stroop	205
2. Autorización institucional y Consentimiento informado	206
3. Descriptivos generales de latencias estimadas en la versión tradicional y pruebas de ajuste	210
4. Análisis de varianza factorial o univariante para las mediciones de latencias estimadas en las tareas de la versión tradicional	213
5. Descriptivos generales del número de aciertos en la versión tradicional y pruebas de ajuste	215
6. Análisis de varianza factorial o univariante para las mediciones de aciertos en las tareas de la versión tradicional	218
7. Descriptivos generales de número de errores en la versión tradicional y pruebas de ajuste	220
8. Análisis de varianza factorial o univariante para las mediciones de errores en las tareas de la versión tradicional	223
9. Índices de interferencia en el Stroop tradicional	225
10. Descriptivos generales de latencias medias de aciertos en la versión computarizada y pruebas de ajuste	228
11. Análisis de varianza factorial o univariante para las mediciones de latencias media de aciertos en las tareas de la versión computarizada	231
12. Descriptivos generales del número de aciertos en la versión computarizada y pruebas de ajuste	233
13. Análisis de varianza factorial o univariante para las mediciones de aciertos en las tareas de la versión computarizada	236
14. Descriptivos generales del número de errores en la versión computarizada y pruebas de ajuste	238
15. Análisis de varianza factorial o univariante para las mediciones de errores en las tareas de la versión computarizada	242
16. Índices de interferencia en el Stroop computarizado	244
17. Análisis de correlaciones entre las versiones del Stroop ( $n=240$ )	247
18. Comparación general del desempeño en las versiones tradicional y computarizada entre los grupos clínicos (TDAH-DA, TDAH-C y TAPZ) y el grupo control	250

## INTRODUCCIÓN

### **Vigencia del Stroop**

El test de Stroop (Stroop, 1935; Golden, 1978) es una de las pruebas psicométricas más utilizadas en los estudios sobre procesos atencionales, particularmente en lo que se refiere al control inhibitorio. Un profuso número de publicaciones, desde inicios del siglo XX hasta la fecha, dan cuenta del interés que sigue generando el Stroop, definido como el instrumento “*Gold Standard*” de las medidas atencionales (MacLeod, 1992, 2015).

El Stroop permite evaluar la capacidad de inhibir la interferencia cognitiva que se produce cuando el procesamiento de un atributo específico del estímulo presentado impide el procesamiento simultáneo de un segundo atributo, o también la demora producida al responder a estímulos de doble entrada o información conflictiva, fenómeno conocido como Efecto Stroop (Jensen, 1965; Golden, 1978; MacLeod, 1991a).

Sus antecedentes pueden rastrearse hasta los primeros laboratorios de psicología y los cuestionamientos sobre discriminación perceptual (Cattell, 1886; Bergström, 1894; James, 1909; Shepard & Fogelsonger, 1913). No obstante, el Stroop no es considerado un instrumento exclusivo del campo de la Psicología Experimental, sus fundamentos incluyen diversos modelos conceptuales.

De modo que, la vigencia y actualidad del Stroop, así como de los interrogantes sobre qué son y cómo operan los mecanismos atencionales, siguen convocando múltiples perspectivas.

### **Uso del Stroop en Diversos Campos**

En el marco de la Psicología Cognitiva, el test suele usarse en el estudio de los efectos de la interferencia, la inhibición de la respuesta automática, la velocidad de procesamiento de la información visual y la competencia de respuesta, entre otras líneas (Jensen & Rohwer, 1966; MacLeod, 1991b; Stafford & Gurney, 2007; Levin & Tzelgov, 2014).

Es también una herramienta fundamental en el campo Neuropsicología, dada su utilidad en la exploración y el diagnóstico de disfunción o daño cerebral, en pacientes con distintas condiciones clínicas, ya que aporta información acerca del funcionamiento de los mecanismos de control atencional, la memoria de trabajo y la flexibilidad cognitiva

(López- Villalobos et al., 2010; Ramírez-Benítez & Díaz-Bringas, 2011; Max et al., 2013; Duchek et al., 2013), tanto el test propiamente dicho como en los montajes experimentales que utilizan tareas tipo Stroop.

Con el desarrollo de técnicas de neuroimagen funcional y la aplicación de modelos computacionales, el uso del Stroop se ha posicionado también como una técnica central en el estudio de las bases neuroanatómicas de distintos aspectos atencionales. En el marco general de las Neurociencias, constituye un intento más por comprender lo que ocurre en determinadas regiones cerebrales al resolver tareas de interferencia (Adlerman et al., 2002; Compton & Banich, 2003; Herd, Banich & O'Reilly, 2006; Kaufmann et al., 2005; Coderre, Filipi, Newhouse & Dumas, 2008; entre otros).

De modo que resulta de suma importancia considerar su fundamento psicométrico, para indagar cuestiones propias de la construcción del test Stroop. La existencia y uso de numerosas adaptaciones y variantes plantea interrogantes sobre la pertinencia de ciertas formas de aplicación, así como respecto de los distintos métodos implementados para su calificación e interpretación (Homack & Riccio, 2004; Van der Elst, Van Boxtel, Van Breukelen & Jolles, 2006; Lansbergen, Kenemans & Van Engeland, 2007; Penner, Kobel, Stöcklin, Weber, Opwis & Calabrese, 2012; Scarpina & Tagini, 2017). Sin duda, este último aspecto es el más relevante en la actualidad.

### **El Stroop como Objeto de Estudio**

El diseño del Test de Palabras y Colores Stroop es muy sencillo y su administración se realiza en un tiempo breve, según el formato aplicado. En su versión original (Stroop, 1935; Golden, 1978) se presentan tres tareas: lectura de un listado de palabras que nombran colores, nominación del color de una secuencia de estímulos neutros y nominación del color en un listado de palabras pintadas en un color incongruente con su significado (adelante abreviadas como P, C y PC, respectivamente).

Múltiples versiones, adaptaciones y revisiones de la prueba han sido desarrolladas con los años (Jensen & Rohwer, 1966; Golden, 1978; MacLeod, 1991a; Scarpina & Tagini, 2017), no obstante, coinciden en el propósito de comparar el rendimiento en la condición conflictiva palabra-color, donde debe suprimir la respuesta automática de leer y

forzar la producción de una respuesta inusual, como es nombrar el color de la tinta en que está pintada cada palabra.

Si bien su uso es extenso, tanto en ámbitos clínicos como de investigación, no existe una versión única del Stroop con relación a los materiales, la administración o la puntuación e interpretación de sus resultados.

Las distintas investigaciones no sólo confirman la observación del fenómeno de interferencia, sino que dan cuenta de la robustez de las mediciones (Homack & Riccio, 2004; Ben-David, Nguyen & Van Lieshout, 2011), independientemente del idioma, las características de las muestras o las adaptaciones del formato de la prueba.

Es común observar que las adaptaciones disponibles, así como las puntuaciones utilizadas, son muy diversas, por lo que resulta primordial la revisión y el análisis de los aspectos propios de la construcción, adaptación, validación y aplicación del test.

De la misma forma, distintos autores han mencionado la urgencia de un debate sobre la equivalencia entre formatos de aplicación y el análisis de la pertinencia de los métodos para cuantificar el Efecto Stroop (Jensen & Rohwer, 1966; Lansbergen, Kenemans & Van Engeland, 2007; Scarpina & Tagini, 2017).

### **Definición de Términos**

A lo largo de este documento se menciona la expresión Stroop de distintas maneras, por lo que vale aclarar a qué hace referencia cada uno.

- *Test Stroop*: se refiere propiamente de la prueba, bien en la versión original de Stroop (1935) o en alguna de las adaptaciones y estandarizaciones, por ejemplo, Stroop Color and Word Test (Golden, 1978), Test de Colores y Palabras Stroop (Golden, 1999) o sus ediciones subsiguientes, utilizadas comúnmente como herramienta de evaluación neuropsicológica en España e Hispanoamérica, así como a las adaptaciones en otros idiomas.
- *Tareas de la prueba*: leer el listado de palabras, nombrar el color en estímulos neutros y nombrar el color en los estímulos palabra-color son las tres tareas que conforman tradicionalmente la aplicación de la prueba.
- *Tareas tipo Stroop*: en el caso de ciertos montajes experimentales que utilizan al menos la medición con estímulos incongruentes y una o varias tareas más con

estímulos neutros de referencia, aunque no necesariamente se consideran los baremos de la prueba tradicional.

- *Versiones Stroop*: adaptaciones del formato original de aplicación, pero que puede cambiar en su presentación (tareas incluidas, número de ítems, orden y diseño de los materiales) y modalidad de las respuestas (verbal, manual, mixta).
- *Variantes Stroop*: aquellas aplicaciones diseñadas desde el mismo principio de evaluación de la interferencia, pero considerando otras categorías semánticas distintas del estímulo palabra-color; es el caso de variantes como el Stroop emocional, Stroop numérico, Stroop de animales, entre muchas otras.
- *Efecto Stroop o Interferencia Stroop*: define el fenómeno de conflicto atencional observado ante ciertos estímulos que muestran información incongruente o que se presentan de modo simultáneo, dando lugar a una disminución en el desempeño en comparación con otras tareas neutras. Distintas estrategias se han sugerido estimar el Efecto Stroop mediante operaciones matemáticas conocidas de modo genérico como índices de interferencia.

### **Propósitos y Alcances de la Investigación**

La presente investigación propone la comparación de dos formas alternativas de aplicación en español, que corresponden a una versión similar a la prueba tradicional de Stroop (1935) y una adaptación computarizada de Etchepareborda y Comas (1994), en una muestra de niños y adolescentes con desarrollo típico y con diagnósticos de Trastorno de Déficit de Atención (TDAH), subtipos combinado (-C) e inatento (-DA), y Trastorno de Aprendizaje (TAPZ) sin déficit de atención secundario.

Específicamente, se presenta una descripción del desempeño general en cada una de estas versiones Stroop y en cada una de sus tareas, considerando los factores de edad, género y condición clínica de los participantes, así como las distintas puntuaciones que pueden derivarse de éstas, y se analiza la equivalencia entre formas de aplicación.

Adicionalmente, se examinan algunos de los más reconocidos métodos de cuantificación del Efecto Stroop, aplicados a los resultados de cada versión para comparar los resultados entre versiones y determinar cuál o cuáles de poseen mayor sensibilidad y especificidad.

Tal como se propone, el análisis psicométrico de estas dos versiones del Stroop no sólo permite reconocer su valor instrumental, sino que contribuye al examen de la pertinencia y rigurosidad en el tratamiento de los resultados de la prueba (MacLeod, 1992; Lansbergen, Kenemans & Van Engeland, 2007; Gualtieri & Johnson, 2006; Gutiérrez-Martínez et al., 2018). Estos aspectos no suelen ser presentados en los estudios de baremación, lo cual configura el principal aporte científico de la investigación.

La propuesta metodológica desarrollada a continuación parte de una elección rigurosa de la muestra, el uso de contrabalanceo en la aplicación de las versiones tradicional y computarizada, así como el análisis de distintas mediciones arrojadas en ambas versiones y sus correspondientes índices de interferencia.

El estudio de las cuestiones psicométricas y teóricas en estos dos formatos del Stroop puede ser de gran utilidad tanto para psicólogos como para profesiones en disciplinas afines, a la hora de determinar qué versión de la prueba y qué método de cuantificación de la capacidad de resistencia a la interferencia Stroop puede resultar más pertinente.



La elección del objeto de estudio y la metodología desarrollada obedece a un interés personal y profesional. En nuestra labor cotidiana de evaluación y diagnóstico insisten los cuestionamientos respecto de la selección y uso de los distintos instrumentos aplicados, su utilidad, practicidad y adecuación, su beneficio y pertinencia, etc., más allá de los indicadores formales de validez y confiabilidad. De modo que intentar respuestas a todo ello resulta prioritario, más si se tiene en cuenta que los resultados representan un punto de partida en la toma de decisiones en la clínica neuropsicológica, para el tratamiento y rehabilitación.

### **Capitulado**

La presentación de la investigación está organizada en cinco apartados. En la primera sección aparece una revisión de los antecedentes teóricos y experimentales de la atención, se reseña brevemente el surgimiento del test, sus adaptaciones, variantes o formatos de aplicación, se presentan aspectos relativos a sus propiedades psicométricas reportadas en los estudios formales del Stroop, así como los hallazgos respecto de las distintas mediciones obtenidas y los cálculos que de ellas se derivan.

Se entrama aquí la revisión del estado del arte en la investigación y uso del Stroop, al señalar la aparición de formas computarizadas, los hallazgos de validez y confiabilidad en los antecedentes de comparación entre versiones, así como en estudios de casos clínicos y controles, y las más recientes líneas de trabajo.

La segunda parte se ocupa del desarrollo empírico y el método elegido para esta investigación, iniciando con el planteo de interrogantes sobre las mediciones del Stroop, los objetivos trazados aquí y las hipótesis de trabajo, para seguir con la descripción del tipo de estudio, la conformación de la muestra y la selección de las dos versiones del Stroop a analizar.

En la tercera parte, la presentación de los resultados comienza con la descripción de cada una de las variables y mediciones básicas consideradas en cada una de las tareas de ambas versiones del Stroop, así como la selección de los estadísticos correspondientes para los contrastes subsiguientes.

Los análisis comparativos entre las versiones y según cada factor permiten corroborar un desempeño diferenciado intrasujeto e intragrupo para las mediciones

básicas (latencias de respuesta, número de aciertos y número de errores), en tanto que el análisis de los métodos de cuantificación de la interferencia o estimación del Efecto Stroop permiten poner en discusión el procedimiento más convencionalmente usado (Golden, 1978) y defender la pertinencia de otras estimaciones (Caffarra et al., 2002; Chafetz & Matthews, 2004; Homack & Riccio, 2004; Van der Elst et al., 2006; Lansbergen, Kenemans & Van Engeland, 2007; López-Villalobos et al., 2010; Scarpina & Tagini, 2017).

La presentación de la discusión, en la cuarta parte, sigue el orden de los análisis realizados para cada una de las versiones del Stroop, considerando el impacto de los factores del sexo, la edad y la condición clínica de los participantes. Se corrobora el Efecto Stroop en ambos formatos de aplicación, aunque se advierten variantes respecto de las mediciones básicas y las cuantificaciones de la interferencia.

Si bien se comprueba una alta confiabilidad en las distintas mediciones, la discrepancia en los contrastes indica poca equivalencia entre las versiones, en tanto que cada una ofrece magnitudes distintas del mismo fenómeno observado. No obstante, en ambas versiones el análisis del desempeño en la tarea de conflicto, especialmente del registro del número de errores cometidos, aporta una adecuada sensibilidad y especificidad en la caracterización de los niveles de interferencia.

En la última parte se presentan las conclusiones y se señalan algunas limitaciones de esta investigación, así como las perspectivas de desarrollo de futuras líneas de investigación.

PARTE I:  
MARCO TEÓRICO

### **Antecedentes en el Estudio de la Atención**

Al realizar un seguimiento al estudio científico de la atención se identifican tres grandes momentos. El primero, situado a finales del siglo XIX que coincide con el nacimiento de los laboratorios experimentales en Psicología en distintos centros académicos, el segundo de la mano del renacimiento cognitivo de mediados del siglo XX y un tercer momento asociado al avance de las neurociencias de las últimas décadas.

A lo largo de los años la naturaleza de la atención ha sido una de las preocupaciones centrales de los investigadores experimentales. Originalmente Cattell (1886) describe los fenómenos de interferencia perceptual en la atención, cuestionando el modo de realizar las mediciones del tiempo de reacción y la influencia de la práctica y la fatiga en la respuesta a determinados estímulos.

Surgen definiciones de la atención como una actividad mental por la cual se regula la “nitidez” que las sensaciones, las imágenes y los sentimientos adquieren en la conciencia (Wundt, 1904), y en donde los procesos de focalización, concentración y conciencia constituyen su esencia (James, 1909).

Vale aquí un paréntesis para destacar brevemente otra perspectiva. La idea de que la atención es imprescindible e inherente a todo proceso mental aparece de forma enfática en los primeros textos freudianos, en los que se distingue entre una atención reflectoria, pasiva y automática, marcada por el contenido del universo sensorial, y una atención psíquica, en tanto mecanismo que induce al yo a seguir las percepciones y a influir sobre ellas. Freud (1895) señala que la atención está enfocada sobre las percepciones y excita la conciencia, es importante para todo el curso del desarrollo, guía al yo y hace posible las más complicadas funciones del pensamiento.

En este mismo sentido, Titchener (1909) hace un reconocimiento a la importancia fundamental de la atención y su estatus central en toda teoría psicológica. Señala que tanto la atención ejecutiva como la respuesta motriz ocurren de modo constante en la experiencia individual, asegurando que éstas son precursoras de toda elaboración mental.

Más adelante se describe el Sistema Reticular Activador Ascendente, ubicado en el tallo cerebral, denominado *arousal*, que permite mantener el nivel de alerta necesario para llevar a cabo ciertos procesos cognitivos de mayor complejidad (Moruzzi & Magoun, 1949). A partir de allí se sugiere que la atención responde a un estado elevado

de excitación, aunque se reconoce una limitación en la capacidad cognitiva destinada a diversas tareas (Kahneman, 1973).

Con los años se bifurcan las líneas de investigación sobre la atención. Se comienza a estudiar el efecto de la práctica y la experiencia previa en la función atencional (Hirst, Spelke, Reaves, Chaharack & Neisser, 1980), el papel de los procesos de control de la memoria operativa (Shiffrin & Schneider, 1977), la idea de la atención como foco (Tsal, 1983; Eriksen & St. James, 1986), la postulación de un sistema de supervisión de la atención asociado a la memoria (Norman & Shallice, 1986), entre otras funciones.

### **Procesos y Componentes de la Atención**

En un intento por precisar cuáles son los componentes principales, niveles o procesos implicados en la atención, surgen distintos modelos explicativos. Algunos de ellos derivados de fuentes clínicas y otros de fuentes experimentales y psicométricas.

Tal es el caso de Solhberg y Mateer (1989) quienes, basados en observaciones de pacientes con lesiones o anomalías encefálicas, definen un modelo jerárquico, que parte de la capacidad de estar despierto y mantenerse alerta y se complejiza gradualmente hacia atención focalizada, sostenida, selectiva, alterna y dividida.

Por su parte Mirsky y Duncan (1987, 2001) desarrollan un modelo multidimensional de la atención, que incluye cinco factores: focalización-ejecución, codificación, cambio, sostenimiento y estabilidad, a partir del estudio de ocho test neuropsicológicos, entre ellos el Stroop.

Mientras que Posner y Petersen (1990) señalan sólo tres componentes de la atención: orientación, vigilancia y atención ejecutiva, que se sustentan en el funcionamiento supramodal de determinadas redes cognitivas y neuroanatómicas.

Particularmente el componente ejecutivo implica la idea de la coordinación voluntaria de los recursos cognitivos con un fin (Posner & Rothbart, 2000) que, a su vez, involucra otros componentes o niveles de la atención: el cambio, el control inhibitorio, la detección de errores, etc.

### **Sobre la Definición de Atención**

Si bien no existe todavía consenso sobre la definición de atención, se reconoce la participación de componentes del nivel neurofisiológico, cognitivo, e incluso afectivo (Roselló, Munar, Obrador & Cardell, 2007).

Para Red Hunt y Ellis (2007) los hallazgos sobre atención explican dos aspectos generales del funcionamiento mental: la selectividad y la capacidad atencional, por cuanto sólo una porción de los estímulos que traspasa los receptores sensoriales llega a ser percibida de modo consciente y puede ser procesada. Dada la limitación de la capacidad atencional, o consecuentemente sus restricciones, para ejecutar dos tareas en forma simultánea, este proceso requiere del control voluntario e intencionado (Cohen, Dunbar & McClelland, 1990; Bugg, Jacoby & Toth, 2008).

Se dice además que la atención es diferenciable de los sistemas de entrada sensorial, de procesamiento y de respuesta motora, pero está asociada a éstos, favorece el comportamiento selectivo ante contextos o tareas específicas, obedeciendo al funcionamiento de determinadas redes neuronales (Posner & Rothbart, 2009; Ball et al., 2011; Steven & Bavelier, 2011; Petersen & Posner, 2012).

La concepción más actual considera que la atención no es un proceso unitario y aislado, sino un estado cognitivo dinámico (Kahneman, 1973; García-Oqueta, 2001; Cohen & Swerdlik, 2001; Roselló, Munar, Obrador & Cardell, 2007; Fernández, 2014), un mecanismo complejo implicado en los procesos de selección, distribución y mantenimiento de la actividad mental (Anderson, 2002; Petersen & Posner, 2012).

Se han utilizado muchas y muy diferentes pruebas o tareas para examinar la naturaleza de la atención. Entre los más estudiados están las tareas de búsqueda, los procedimientos de facilitación o *priming* y el test Stroop (Tirapu-Ustárrroz, García-Molina, Ríos-Lago & Ardila-Ardila, 2012).

Las tareas con estímulos conflictivos, superpuestos o dicotómicos, como presenta el Stroop o variantes de este, han sido usadas a lo largo de estos tres momentos para intentar responder a interrogantes sobre la discriminación atencional, atención selectiva, procesos lectores, memoria semántica, interferencia y control atencional, entre otros procesos (Reed Hunt & Ellis, 2007; Bausela Herreras & Santos Cela, 2006).

Distintos autores sugieren que instrumentos como el Stroop ofrecen claves en la comprensión de la atención (Cohen, Dunbar & McClelland, 1990; MacLeod, 1991b, 1992), y más particularmente de control atencional como una medida del funcionamiento fronto-ejecutivo (Caffarra, Vezzadini, Francesca, Zonato & Veneri, 2002; Chafetz & Matthews, 2004; Scarpina & Tagini, 2017).

### **Aproximación Neurocognitiva**

Una de las más conocidas teorías de la atención es la de Posner (1971) y sus revisiones subsiguientes (Posner & Petersen, 1990; Fernández-Duque & Posner, 2001; Posner & Rothbart, 2009; Petersen & Posner, 2012), que identifica tres principales funciones: orientar a los estímulos sensoriales, mantener el estado de alerta y orquestar las acciones voluntarias. Estas son descritas como redes cognitivas supramodales especializadas en la respuesta de orientación, vigilancia y la función ejecutiva, respectivamente.

Tal concepción coincide en parte con el esquema de las unidades funcionales de Luria (1970, 1974, 1995), al distinguir tres sistemas fisiológica y anatómicamente diferenciables, pero interrelacionados: un sistema inicial de regulación del tono y vigilia, situado en estructuras subcorticales, y otros dos considerados de nivel superior, ubicados en regiones corticales posteriores y anteriores.

### **Redes Atencionales**

El primer sistema, el más primitivo en el desarrollo, es responsable tanto del nivel general de excitación interna, como de la respuesta de alerta ante los cambios en el medio ambiente. Comprende el Sistema Reticular Activador, el *locus caeruleus* y las proyecciones mesencefálicas al hipotálamo y el tálamo, entre otras estructuras. Se encarga del aspecto intensivo de la atención, los estados de alertamiento (*arousal*), la vigilancia y la atención sostenida, así como cambios fásicos de la alerta, esto es, de orientación a estímulos novedosos o sorprendidos.

El segundo, de atención posterior, está asociado al funcionamiento de los lóbulos parietales, tiene que ver con la dirección de la atención en el espacio, la respuesta de orientación visual y la atención focalizada. La red posterior involucra la activación del

córtex parietal posterior, los colículos superiores y el núcleo pulvinar del tálamo, principalmente. Su correlato con áreas primarias y secundarias de la corteza temporal y occipital permite recibir, analizar y almacenar información.

Y, el tercero, conocido como red anterior, está integrado por regiones de los lóbulos frontales, particularmente con el córtex cingulado anterior, regiones prefrontales mediales, prefrontales dorsolaterales y el núcleo caudado del neocórtex (Fuster, 1999, 2002).

A su vez, el nivel más alto se relaciona con la capacidad de mantener la atención en objetos o hechos específicos a pesar de la presencia de otros estímulos, por lo que se dice que es un sistema para programar y verificar la actividad, esto es, de regulación para funciones de orden superior, más recientemente englobadas en el término de funciones ejecutivas (Tirapu-Ustárriz et al., 2008).

En este sentido, para Cabeza y Nyberg (2000), esta red anterior está implicada en la selectividad atencional y el control de la acción, iniciación e inhibición de la respuesta (zonas anteriores y adyacentes al giro cingulado), la regulación de la atención, necesaria para la planificación, regulación deliberada y flexibilidad de las demandas cambiantes del contexto (córtex prefrontal, principalmente).

No obstante, el grado de activación de estas áreas encefálicas puede variar según la modalidad sensorial predominante, sea visual, auditiva, somestésica, etc. Fuster (2008) y Fernández (2014) señalan que la atención visual implica estructuras subcorticales (colículos superiores, pulvinar, caudado y sustancia nigra) y estructuras corticales occipitales (correspondientes a las áreas de Brodmann 17, 18 y 19), temporales (en especial las áreas 22 y 37), parietales (7, 39 y 40) y regiones prefrontales dorsolaterales y cinguladas mediales (9, 46, 6, 32 y 24).

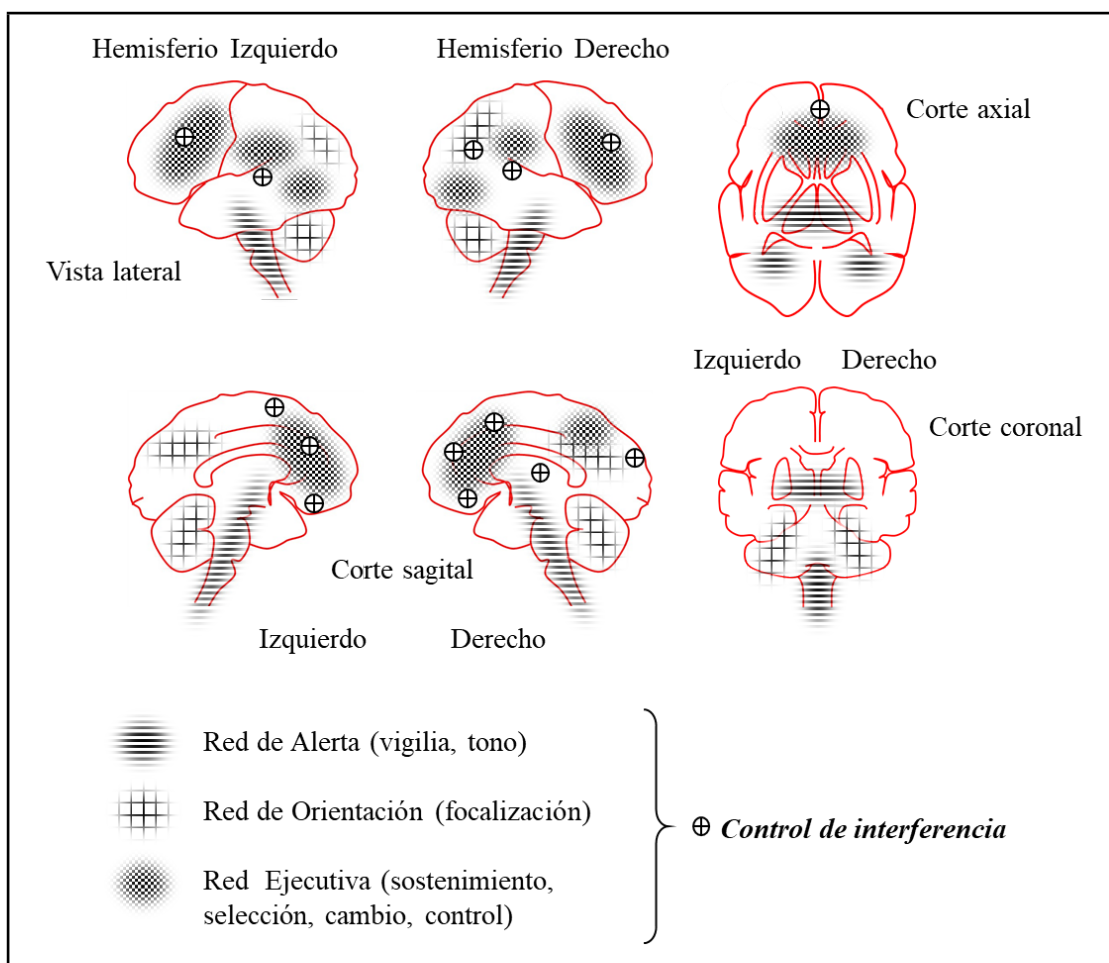
Para Tirapu-Ustárriz et al. (2008), los procesos ejecutivos de la atención involucran además al área motora suplementaria, la corteza orbitofrontal, así como secciones de los ganglios basales y el tálamo. Al mismo tiempo, la atención opera bilateralmente involucrando la activación de ambos hemisferios cerebrales (Posner & Petersen, 1990; Mesulam, 1990; Posner & Rothbart, 2000; Fuster, 2002).

A continuación, la Figura 1 presenta una demarcación aproximada de algunas de las regiones involucradas en procesos atencionales.



**Figura 1**

*Representación de las áreas corticales y subcorticales implicadas en la atención*



*Adaptado de Cabeza & Nyberg (2000) y Meneses Ortega (2001)*

Es interesante observar cómo las áreas asociadas al control de interferencia comparten la disposición de las redes de alertamiento, orientación y atención ejecutiva. Como se indica en la figura anterior, aquellas regiones cerebrales que muestran activación durante la resolución del Test o de tareas tipo Stroop.

La activación funcional de áreas específicas, así como de las redes córtico-subcorticales involucradas, ha sido observada mediante estudios realizados con potenciales evocados, resonancia magnética, tomografía, tractografía, etc. (Bush et al., 1998; Cabeza & Nyberg, 2000; Leung, Skudlaski, Gatenby, Peterson & Gore, 2000; Stuss, Floden, Alexander, Levine & Katz, 2001; Ravnkilde et al., 2002; Harrison, 2005; Kaufmann, 2005; Lansbergen, Van Hell & Kenemans, 2007).



## Orígenes del Test Stroop


Los orígenes empíricos de la prueba suelen situarse puntualmente en dos publicaciones: "*Grundformen menschlichen Seins*" (Jaensch, 1929) y "*Studies of interference in serial verbal reactions*" (Stroop, 1935). En ambas se presentan resultados experimentales de la evaluación de la interferencia que ejercen ciertos estímulos sobre otros. Dado el contexto idiomático, histórico y político, el trabajo de Stroop logra mayor difusión y reconocimiento, mientras que a Jaensch se le atribuye una visión especulativa y tendenciosa respecto de las diferencias individuales y socioculturales.

Ambos trabajos tienen como antecedente diferentes estudios experimentales, conocidos ya desde finales del siglo XIX, que se ocuparon de describir los fenómenos de interferencia perceptual (Cattell, 1886) o de asociación y discriminación perceptual, de la disposición en los procesos de reconocimiento de color en contraste con la lectura y del estudio de la inhibición perceptual, entre otros (Bergström, 1894; Shepard & Fogelsonger, 1913).

Al igual que otros investigadores, Stroop (1935) intenta comprender por qué algunas asociaciones son más fuertes que otras, o cómo se explica que una persona tarde más tiempo al nombrar el color de una serie de estímulos simples que al leer las palabras impresas en color. Su trabajo experimental examina y analiza estas diferencias mediante un montaje de tareas de lectura y nominación de color. Para ello utiliza cinco palabras y colores: *red, blue, green, brown, purple*, y descarta el uso del amarillo por falta de contraste.

### Observaciones Iniciales

En un primer experimento, Stroop (1935) examina el efecto de interferencia de los estímulos de color sobre la lectura serial de nombres de colores. Esto es, evalúa el tiempo que tarda la lectura de un listado de 100 palabras donde el color de la impresión y la palabra son diferentes (**RED**: decir "red") y la lectura de otro listado con palabras impresas en tinta negra (**RED**: decir "red").

En el segundo experimento, Stroop (1935) analiza el efecto de la interferencia de la palabra sobre la denominación serial de los colores, al nombrar 100 colores presentados en barras de color sólidas en lugar de palabras (: decir "red"), en

comparación con la tarea de nombrar el color de la tinta en estímulos incongruentes, que corresponden a palabras pintadas en un color que no concuerda con el significado de la palabra (**GREEN**: decir “red”).

En ambas experiencias se observa una mayor demora al resolver la tarea con estímulos de doble entrada de información. En el primer caso, Stroop (1935) encuentra que leer los 100 estímulos de palabras en tinta negra se realiza 2.3 segundos más rápido en promedio, una diferencia pequeña que no alcanza un nivel de significancia. En el segundo caso, la tarea de nombrar el color en los estímulos incongruentes fue 47.0 segundos más demorada que nominar el color en los estímulos neutros, lo que indica un incremento significativo en el tiempo medio empleado para responder al conflicto palabra-color (74.2%).

Entonces, Stroop (1935) sugiere que el análisis de estas diferencias permite una estimación de la interferencia perceptual tanto en tareas de lectura como en tareas de nominación del color. Siendo el segundo procedimiento el más comúnmente utilizado en estudios de interferencia.

A partir de aquí, la demora que producen los estímulos incongruentes comienza a postularse como una medida central de la atención (MacLeod, 1991a, b) y se discuten distintas aproximaciones en la comprensión de sus resultados y el cálculo de la interferencia (Jensen, 1965, 1966).



Stroop (1935) atribuye este fenómeno al hecho de que, en un sujeto “normal”, un estímulo como una palabra atrae una sola respuesta (la lectura), mientras que un estímulo conflictivo atrae una variedad de posibles respuestas, de modo que la elección de la respuesta causa una evidente desaceleración en el rendimiento.

Para Cohen, Dunbar y McClelland (1990), los efectos observados en el Stroop proporcionan una ilustración clara de la capacidad de las personas para seleccionar entre informaciones distintas, así como la capacidad de algunos estímulos para escapar del control atencional.

En su tercer experimento, Stroop (1935) examina el efecto de la práctica en estas tareas y los cambios observados cuando se realiza en orden inverso, encontrándose que la práctica aumentaba o disminuía la variabilidad del grupo, dependiendo de la naturaleza del material utilizado, aunque la interferencia no llega a desaparecer completamente.

Debe destacarse el trabajo de Klein (1964) que incorpora varias condiciones nuevas al formato original y observa que, a pesar de las instrucciones para ignorar la palabra, resulta difícil nombrar el color de manera fluida. La condición palabra-color estándar (**BLUE**) registra la mayor interferencia, mientras que en palabras relacionadas con el color (**SKY**), palabras comunes no asociadas (**HAND**), palabras raras (**ABJURE**) y sílabas de difícil pronunciación (**GSXRQ**), progresivamente la tarea de nombrar la tinta de las letras se hace más fácil. De modo que cuanto mayor es la relevancia de la palabra, mayor es la interferencia resultante.

No obstante, tal como advierte Stroop (1935), es razonable concluir que la diferencia en la velocidad en la lectura de los nombres de colores y la denominación de los estímulos de color está asociada al entrenamiento que se tenga en estas dos actividades.

Van Maanen, Van Rijn y Borst (2009) subrayan tres hallazgos centrales en las experiencias de Stroop (1935) y los trabajos que le siguen: a) la observación de la interferencia semántica, por cuanto nombrar el color en un estímulo neutral (por ejemplo: ) es más rápido que hacerlo en condiciones incongruentes (**BLUE**), b) la observación de una cierta facilitación semántica, cuando nombrar el color en estímulos congruentes (**RED**) resulta más rápido que sólo decir el color en estímulos neutros () y c) la desaparición o atenuación de la interferencia y de la facilitación, si la tarea es simplemente leer en vez de nombrar la tinta (**RED**), lo que se ha denominado Asincronía Stroop y suele explicarse por la automatización de la lectura.

Los efectos de facilitación, asincronía e interferencia Stroop pueden evidenciarse al utilizar otras categorías semánticas distintas del color (véanse: Bush et al., 1998; Risko, Schmidt & Besner, 2006; Van Maanen, Van Rijn & Borst, 2009; Augustinova, Parris & Ferrand, 2019).

## El Efecto Stroop

Las investigaciones que siguieron al trabajo de Stroop (1935) han sido consistentes en la observación sobre las diferencias entre las tareas con estímulos neutros y conflictivos (Jensen, 1965; Kahneman, 1973; Golden, 1978,1999; Cohen, Dunbar & McClelland, 1990), ya que las latencias de respuesta aumentan de modo significativo cuando la palabra es incompatible con el color de la tinta, que si es congruente o neutra.

Su experiencia se ha extendido del campo experimental a la Psicología Cognitiva, Neuropsicología, Neurofisiología y Psicopatología, entre otras disciplinas, tanto en investigación de la capacidad atencional en tareas de interferencia, como en evaluación clínica.

La definición más sencilla del Efecto Stroop se adhiere a la demostración del fenómeno de ralentización en el tiempo de reacción al enfrentar información conflictiva (Stroop, 1935; Golden, 1978, 1999; Cohen, Dunbar & McClelland, 1990; McLeod, 1991; Welsh et al., 1991; Besner, Stolz & Boutilier, 1997; Bugg, Jacoby & Toth, 2008; Steinhäuser & Hübner, 2009; Ramírez-Benítez & Díaz-Bringas, 2011; Martín, Hernández, Rodríguez, García, Díaz & Jiménez, 2012, entre otros). En otras palabras, el efecto de interferencia se evidencia en una ejecución más dificultosa y lenta, con mayor número de errores y menos aciertos durante la tarea de nominación del color en los estímulos incongruentes (**ROJO AZUL NEGRO VERDE**, por ejemplo).

Como se menciona antes, esto mismo ha sido observado incluso cuando las palabras cuyos colores deben nombrarse no son nombres de colores. Por ejemplo, al usar la palabra **SAD**, en la adaptación del Stroop emocional de Compton et al. (2003), o también la palabra **HUMO**, en la variante de la prueba de riesgo tabáquico de Calleja & Hernández-Pozo (2009), entre otras variantes del Stroop.

La discrepancia entre el tiempo que se requiere para la nominación del color y la lectura de los nombres de colores se ha convertido en una medida de referencia en los estudios de interferencia (Kahneman, 1973; MacLeod, 1991a, 1992).

No obstante, una de los procedimientos más difundidos y aceptados para la cuantificación de la interferencia Stroop es el método propuesto por Golden (1978), en el que se opera con el número de respuestas correctas, en vez de tiempos de respuesta, estableciendo además un ajuste por edad.

### **El Efecto Stroop como Proceso Cognitivo**

Los estudios del Efecto Stroop se han enfocado en la capacidad y el esfuerzo cognitivo (Hunt & Ellis, 2004; Kaplan, Sengor, Gurvit & Guzelis, 2007), implicados en las tareas de lectura y de nominación del color, así como en la estimación de las discrepancias entre estas cuando implican competición de estímulos, información dual o información incongruente, lo que produce efectos neuronales, cognitivos y comportamentales.

Kahneman (1973) distingue entre la interferencia de la capacidad y la interferencia estructural. La primera sugiere que existe interferencia incluso cuando las dos actividades no compartan los mismos mecanismos de percepción (*visual versus* auditiva, por ejemplo) o mecanismos de respuesta (*verbal versus* motora). Esto es, tiene efecto sobre la capacidad atencional y depende de las demandas de las actividades en competencia.

La segunda ocurre porque se ocupan los mismos mecanismos de percepción o de respuesta; es estructural en el sentido de la organización del sistema nervioso, por lo que tiene efecto sobre sus centros de procesamiento cognitivo, sus vías aferentes y eferentes.

Asimismo, Kahneman (1973) discute cuáles son las condiciones, disponibilidad y velocidad de la respuesta al conflicto, señalando que una persona suele estar más activa y alerta mientras resuelve un problema, mira una imagen o responde a una de las láminas del Stroop, que durante un periodo de reposo vigil, mostrando una disposición diferenciable.

Lo anterior coincide con la observación de Jensen y Rohwer (1966) sobre los efectos comportamentales (proxémicos) durante la tarea de conflicto en el test Stroop, al describir cómo las personas tensan hacia adelante el torso, toman la expresión de fatiga visual, hacen gestos con los brazos y las manos, muestran mayor énfasis vocal.

### **Modelos del Efecto Stroop**

La visión general de Stroop (1935) tiene en cuenta dos explicaciones: de un lado, la automaticidad de la lectura y su efecto sobre la nominación del color en el estímulo conflictivo y, de otro, el tiempo relativo a la resolución de las tareas.

Ambos modelos tienen una amplia difusión y aceptación. Sin embargo, otras posturas ofrecen alternativas en la comprensión de qué ocurre durante las tareas del Stroop, cómo se relacionan, cuándo y por qué varían, dónde ocurre el procesamiento de la información congruente e incongruente, etc.

### **Velocidad de Procesamiento**

Tiempo de reacción y velocidad de procesamiento son dos constructos utilizados indistintamente en los estudios que analizan la rapidez y eficacia con que se responde a las tareas del Test Stroop, así como las diferencias temporales entre estímulos neutros y conflictivos.

El primero se define por el lapso entre la presentación del estímulo, captado por los órganos sensoriales, y el inicio de la respuesta motora (Kahneman, 1973); en tanto que el segundo considera el inicio de la respuesta previo a la ejecución motora, el momento en que ocurre una activación en los centros de procesamiento cortical (Dunbar & MacLeod, 1984).

La velocidad de procesamiento es una medida de la eficiencia de la función cognitiva y suele expresarse en términos de latencia de respuesta, esto es, el tiempo requerido para completar una serie de operaciones o el número de elementos logrados en un período establecido; por lo que puede comprender la sumatoria de los tiempos para percibir una información y procesarla, así como para preparar y ejecutar una respuesta (Ríos Lago & Periañez Morales, 2010).

La evidencia de los tiempos de reacción sugiere que el análisis de señales concurrentes ocurre en paralelo, pero con una tasa de enlentecimiento en el foco atencional, lo que parece corresponderse con la dificultad para monitorear distintos mensajes dentro de la misma modalidad sensorial, en especial si son conflictivos (Etchepareborda, 2004).

Posner y Snyder (1975) presentan tres conclusiones respecto de la velocidad relativa de procesamiento en el Stroop: en primer lugar, el efecto de interferencia surge debido a la competencia entre respuestas; segundo, la dirección de la interferencia depende de las relaciones de tiempo involucradas, en tanto que las palabras se leen más rápido de lo que nombran los colores, como señala Cattell (1886); y, por último, las



palabras a menudo facilitan la salida vocal a los colores con los que comparten un nombre común. También advierten que las tareas de leer y nominar el color continúan en paralelo y sin interferencia hasta cerca de la salida.

Por su parte, Cohen, Dunbar y McClelland (1988) prefieren hablar de una velocidad de desempeño y advierten que no se trata de una variable binaria, por cuanto la práctica produce un decremento gradual en los tiempos de respuesta.

Según MacLeod (1991a), tal diferencia se considera particularmente crítica cuando compiten dos respuestas potenciales (la palabra o el color, por ejemplo), siendo la interferencia el costo de demora de esta competencia.

Sin embargo, aunque es convincente y ampliamente aceptada la velocidad de procesamiento como modelo teórico del Stroop, en su revisión más reciente MacLeod (2015) señala que esta no proporciona una explicación suficiente de la interferencia. Algunas variantes y manipulaciones de las tareas presentadas en la prueba (letras invertidas, uso de formas y colores, *priming*, etc.) pueden automatizarse luego del entrenamiento, afectando la celeridad de la respuesta.

### **Automaticidad**

Desde la teoría de la capacidad atencional (Kahneman, 1973) se abre la discusión sobre aquellas tareas que pueden ejecutarse de manera automática, utilizando pocos recursos cognitivos, en comparación con aquellas que requieren mayor control atencional.

El concepto de respuesta automatizada se introduce para exponer cómo ciertas actividades pueden lograrse sin demandar capacidad de procesamiento central (Posner & Snyder, 1975; Stirling, 1979), esto es, procesos automáticos que pueden ocurrir sin intención y que no requieren de la observación consciente.

Una actividad habitual como leer logra automaticidad debido a la repetición en los años iniciales de la escolarización (Logan, 1988), siendo la práctica “el ingrediente crucial en la automatización” (Hunt & Ellis, 2004). Si bien la lectura comienza con el reconocimiento individual de cada letra, una vez se aprende a leer es posible comprender automáticamente el significado de las palabras como unidades.

No obstante, durante el reconocimiento y la nominación de estímulos de color puede aparecer cierta vacilación al responder, por lo que se presume que se trata de un proceso no automatizado o, en todo caso, de un menor entrenamiento.

En la tarea de conflicto del Stroop, nombrar el color de la tinta en el estímulo en lugar de decir el significado de la palabra, requiere la anulación del impulso inicial de leer automáticamente. Esta idea está basada en la premisa que la lectura automática no necesita atención controlada (Monahan, 2001), pero utiliza suficientes recursos atencionales para reducir la cantidad de atención accesible para el procesamiento de la información del color.

Cuando se trata de estímulos congruentes, las tareas de leer o nombrar el color se realizan con poco esfuerzo cognitivo, mientras que con estímulos novedosos o conflictivos se requiere del control atencional, por cuanto el significado de la palabra puede afectar la capacidad para nombrar el color de la tinta en esas palabras (Posner & Boies, 1971).

Para Cohen, Dunbar y McClelland (1990), la distinción tradicional entre procesos automáticos y controlados requiere una reconsideración. En el marco de las teorías de procesamiento distribuido paralelo, proponen que los procesos automáticos emergen gradualmente con la práctica, esto es, mediante el fortalecimiento de las conexiones entre las unidades cognitivas.

Por su parte, Besner, Stolz y Boutiuer (1997) destacan que el Efecto Stroop implica el componente semántico de las palabras al que se retorna (al menos de modo inconsciente) para procesar las palabras, dado que el contexto literal resulta más poderoso que la instrucción de ignorar la palabra. Coinciden con Klein (1964) en la noción de un “gradiente semántico” dentro del efecto de interferencia.

Dicho contexto restablece la lectura en el estímulo dual, permitiendo un conjunto de estructuras mentales y el camino que las conecta. Desde el modelo de *set* mental (Besner, Stolz & Boutiuer, 1997), la baja automaticidad de la lectura explica la aceleración del procesamiento de los estímulos incongruentes, al reducir la interferencia que genera el conflicto palabra-color.

En este sentido, Lazarus, Ludwig y Aberson (1984) sostienen que el Stroop permite un registro breve pero efectivo para diferenciar entre lectores hábiles y aquellos que presentan problemas en la lectura.

Finalmente, Protopapas, Archonti y Skaloumbakas (2007) advierten que la automaticidad es una noción compleja: si bien se asocia comúnmente a la velocidad de la respuesta, no es equivalente a fluencia, y deben considerarse aspectos como la intencionalidad (inicio, control y terminación de la tarea), los requerimientos del recurso cognitivo (atención como esfuerzo) y vigilia consciente (atención focalizada).

### **Procesamiento Distribuido Paralelo**

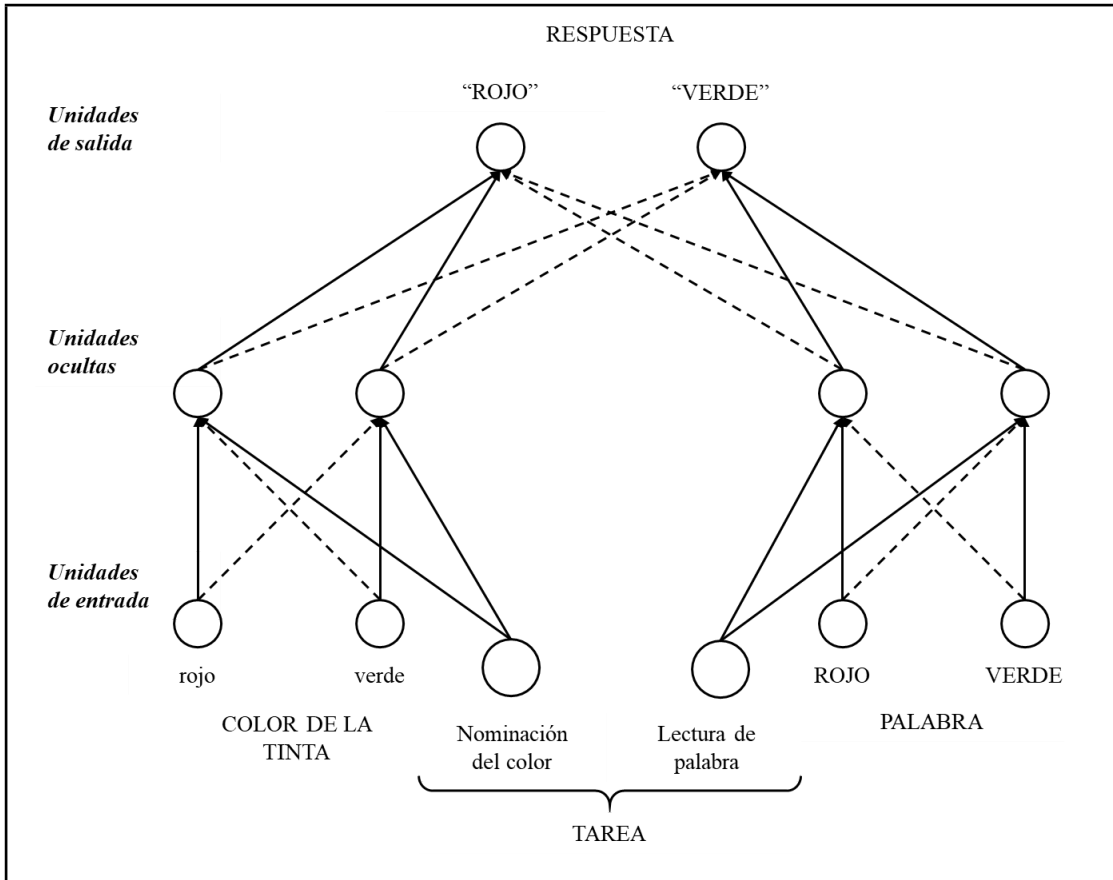
La teoría del procesamiento distribuido paralelo propone que toda la información disponible en el campo sensorial se procesa de manera simultánea, y no de modo secuencial, al tiempo que los componentes de cada estímulo se procesan junto con su totalidad (Dunbar & MacLeod, 1984; Mesulam, 1990).

El modelo de Cohen, Dunbar y McClelland (1990) integra una arquitectura de nodos y vías en la representación de la información y el procesamiento. Fue diseñado para probar una hipótesis sobre los mecanismos de respuesta automática y controlada.

En la Figura 2, el flujo de la información del modelo se orienta de abajo hacia arriba, de la tarea solicitada a la respuesta. Las líneas representan conexiones unidireccionales, las punteadas son de carácter inhibitorio y las continuas son excitatorias.

Figura 2

Arquitectura de red del modelo de Cohen, Dunbar y McClelland (1990)



Adaptado de Cohen, Dunbar y McClelland (1990) y MacLeod (1991a)

En este modelo conexionista se destaca el rol de la atención selectiva y la variabilidad en los mecanismos de respuesta y considera la automaticidad en términos de un *continuum* basado en la fuerza del procesamiento, ante las demandas de cada tarea (Cohen, Dunbar & McClelland, 1990). Aunque también implica la modulación atencional de dicho procesamiento.

A medida que el cerebro analiza la información, construye vías específicas para realizar cada tarea; el procesamiento toma lugar en módulos interconectados y ocurre por la propagación de señales de un módulo a otro, fortaleciendo determinadas rutas. Cuando se activan dos vías simultáneamente se producen interferencias entre la ruta de lectura (más fuerte o de mayor conectividad) y la ruta de nominación (más débil o de menor conectividad).

Por ejemplo, ante el estímulo **AZUL**, la palabra azul es procesada por el lexicón mental y el color rojo se procesa en la memoria semántica. Con la tercera tarea del

Stroop, decir el color de las letras implica una translación entre módulos cognitivos (Sugg & McDonald, 1994; Monahan, 1997, 2001), esto es, del nodo de memoria semántica (identificación de la tinta roja) al lexicón mental (la palabra rojo).

### **Atención Selectiva**

Kahneman (1973) advierte que el monitoreo de varios canales de información es posible, pero menos efectivo que atender a una única información. La teoría de la atención selectiva plantea que se necesita utilizar más atención para reconocer un color que para reconocer la palabra en el estímulo conflictivo, implicando un mayor gasto del recurso atencional (Mead et al., 2002; Lamers, Roelofs & Rabeling-Keus, 2010; Wright, 2017).

En este marco, la selección de la respuesta en el estímulo incongruente puede ser el resultado de una excitación de las representaciones del estímulo atendido (color de las letras) o de una mejor inhibición de distractores que no son respuestas apropiadas (Wühr & Waszak, 2003; Zarghi et al., 2012), de modo que el enlentecimiento de la respuesta o presencia de errores da cuenta de una dificultad de selección.

Algunos estudios señalan que si una persona logra enfocarse exclusivamente en el color de la fuente mientras ignora el contenido léxico, no habrá diferencia en los tiempos de reacción entre las tareas de nominación en estímulos conflictivos y neutros; por el contrario, si no se logra atender selectivamente el color de la fuente, responder al conflicto palabra-color le tomará más tiempo (Ben-David, Nguyen & Van Lieshout, 2011).

La generación de la respuesta en la tarea Stroop requiere la inhibición intencional de la respuesta habitual de leer, así como la selección y reclutamiento de los centros de procesamiento adecuados a la ejecución. Según Pardo, Pardo, Janer y Raichle (1990), la activación cortical del cíngulo anterior media dicho procesamiento de selección atencional.

Sin embargo, según Mead et al. (2002), otras regiones como el surco precentral inferior, el *pars opercularis* y la región premotora ventral se encuentran involucradas críticamente en la condición de interferencia.

Así, puede advertirse cómo distintas áreas corticales participan de modo conjunto en la atención selectiva y la mediación de las demandas articulatorias competitivas para producir las respuestas durante el Stroop. Como se señala antes, el procesamiento de estímulos incongruentes comparte una disposición similar a las redes de alerta, orientación y atención ejecutiva (Cabeza & Nyberg, 2000; Meneses Ortega, 2001).

### **Competencia de Información**

Desde la teoría de competencia, el estímulo palabra-color puede elicitar tanto una respuesta correcta como una incorrecta. El efecto de interferencia ocurre principalmente cuando la palabra impresa provoca una respuesta de codificación que es relevante para la tarea y, por tanto, es el resultado de la competencia en el nivel de codificación. Para Kahneman (1973) algunas respuestas están ya preparadas por la tarea, a modo de *priming*, y la provocación de tales respuestas por un estímulo irrelevante causa la interferencia.

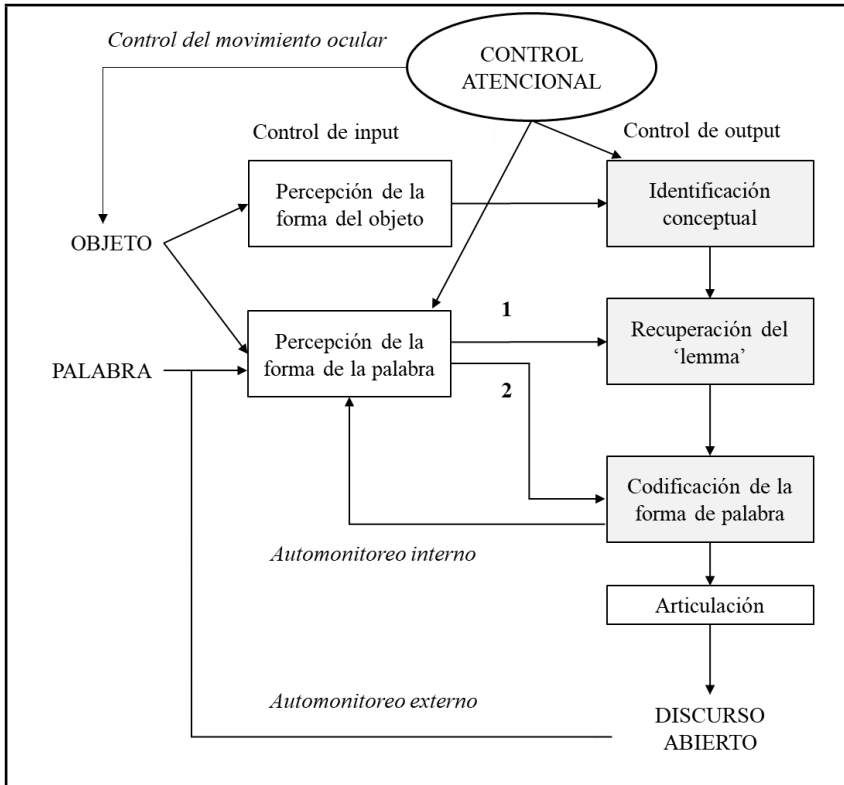
Repovs (2004) y Washburn (2016) sugieren que la información proporcionada por el sistema sensorial comienza a construir una activación preferencial de la respuesta, cuanto menor es la relación de la intención competitiva con el conjunto de respuestas, más rápido se puede establecer la diferencia y se inicia la respuesta. Un tiempo de reacción más lento ocurre debido a la interferencia producida por la doble entrada del estímulo, o un retraso en el procesamiento, causado por la competencia de la información conflictiva.

Existen distintos mecanismos para nombrar el color y las palabras, en función de las diferencias en la arquitectura de producción del lenguaje, de modo que la interferencia palabra-color puede ser resultante del acceso a la memoria declarativa.

En este sentido, el modelo de Roelofs (1992) supone que la red comprende tres estratos principales, que corresponden a las etapas de preparación conceptual, recuperación del lemma (nivel sintáctico) y codificación de la forma de la palabra. Tal como presenta la figura 3, la lectura de palabras puede involucrar o no la selección de lemas (rutas 1 o 2).

**Figura 3**

*Modelo de codificación de palabras por activación y verificación de Roelofs (1992)*



*Adaptado de Roelofs (1992, 2003)*

Roelofs (2003) señala que el Efecto Stroop surge de las interacciones de procesamiento dentro de la arquitectura de producción de lenguaje y el control explícito referido a objetivos, por lo que el conflicto crucial de la denominación de color ocurre durante el acceso al *lemma*.

Otras variantes de estos modelos tienen en cuenta el conocimiento sobre redes y módulos neuronales que interactúan conjuntamente durante las tareas del Stroop.

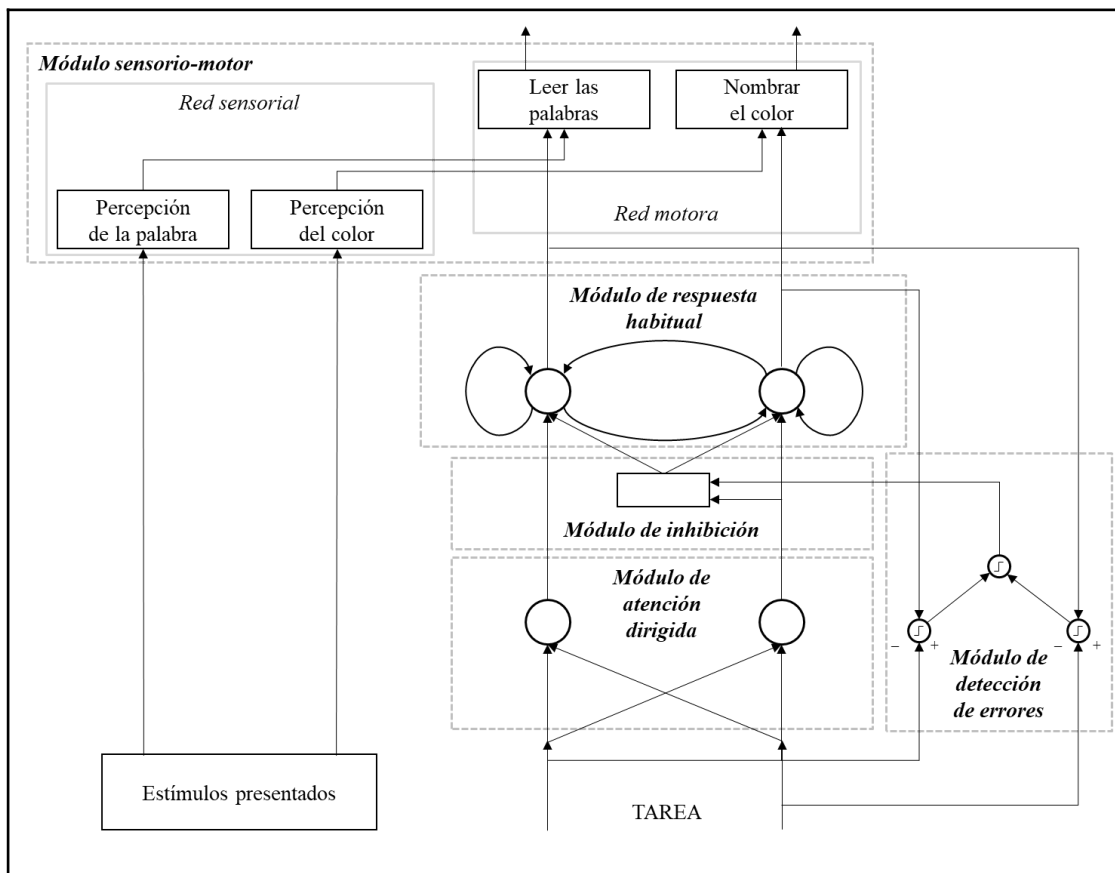
El modelo integrativo de redes neuronales, propuesto por Herd, Banich y O'Reilly (2006) se basa en la difusión de la activación de distintas regiones cerebrales, no únicamente frontales. Resultados de resonancia durante la ejecución del Stroop muestran además la activación del córtex superior del lóbulo parietal izquierdo y regiones temporales cuando la consigna es ignorar la palabra, y de regiones ventrales del lóbulo occipital cuando la consigna es ignorar el objeto.

El modelo de Kaplan et al. (2007) ilustra cómo los procesos cognitivos de orden superior entran en el Stroop, sobre la base de la interacción funcional de distintos

módulos (Figura 4). Cuando la tarea es nombrar colores, el módulo de inhibición se torna para inhibir la tendencia del módulo de respuesta habitual y transmite la selección de acciones relevantes para la tarea. Cuando el módulo de respuesta habitual hace una elección errónea debido a una inhibición insuficiente, el módulo de detección de errores verifica la relevancia de la decisión y reactiva el módulo de inhibición.

**Figura 4**

*Diseño de módulos y redes neuronales de Kaplan et al. (2007)*



*Adaptado de Kaplan et al. (2007)*

Dichos módulos no son propiamente análogos en términos fisiológicos o anatómicos, sino más bien una aproximación computacional al sistema prefrontal para la inhibición cognitiva. En especial las regiones operculares frontales parecen tener un papel decisivo en el monitoreo y control atencional (Kaplan et al., 2007).

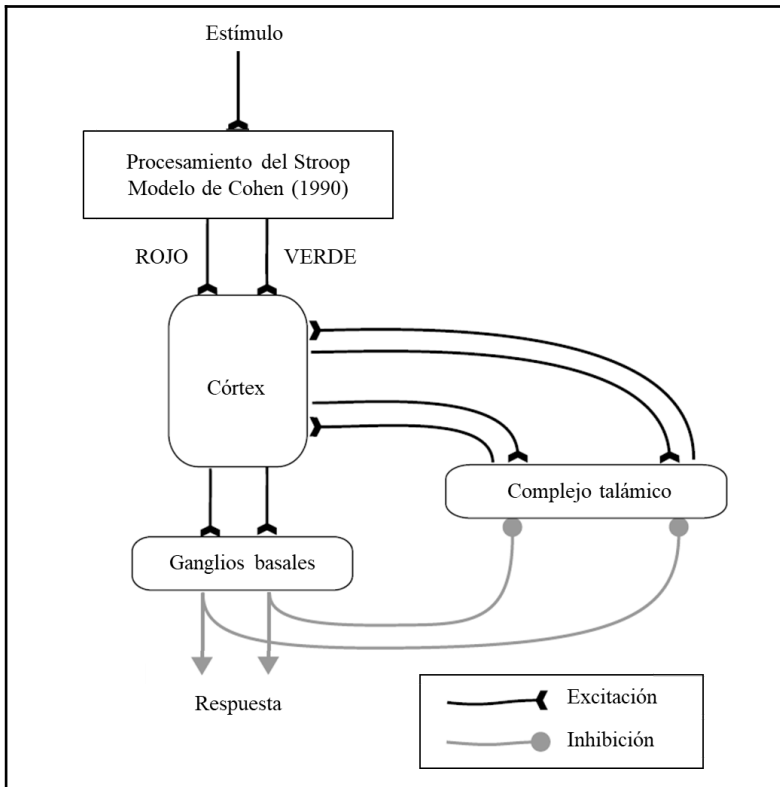
Por su parte, Statford y Gurney (2007) proponen una arquitectura funcional que involucra al mismo tiempo la participación de estructuras subcorticales. Como muestra la figura 5, los autores combinan el modelo de Cohen, Dunbar y McClelland (1990) con el



mecanismo de respuesta de los ganglios basales y circuitos tálamo-corticales, que sirven como un "interruptor" central o árbitro entre las solicitudes de acción.

**Figura 5.**

*Modelo de difusión de Stattford y Gurney (2007)*



*Adaptado de Stattford y Gurney (2007)*

En el marco del monitoreo del conflicto y la activación de redes neuronales, otros modelos han sido propuestos (Van Maanen, Van Rijn & Borst, 2009; Levin & Tzelgov, 2014). Las investigaciones más recientes argumentan que no hay un único y estático *locus* para el efecto Stroop y destacan las diferencias observadas a nivel cognitivo y neurofisiológico (Parris, Augustinova & Ferrand, 2020).

### **Mecanismos de Control Inhibitorio: Paradigma Stroop**

A la par con estas distintas aproximaciones, es importante revisar aquellas explicaciones del Efecto Stroop en el marco de la investigación del neurodesarrollo, y desde los hallazgos entre distintos grupos clínicos y controles, en los estudios de las funciones ejecutivas.

Como modelo de conocimiento aceptado por la comunidad científica, aún desde distintas posturas teóricas, el Stroop se postula como herramienta central en la investigación de los mecanismos de control inhibitorio, que están directamente relacionados con la regulación atencional y la monitorización de orden superior.

Según Anderson (2002), los mecanismos de control inhibitorio permiten ejecutar los planes y los programas en un orden determinado, identificar oportunamente los errores y alcanzar las metas. Se definen también por la capacidad de retrasar una respuesta, interrumpir un proceso que ya ha comenzado o evitar la interferencia (Tamm, Menon & Reiss, 2002), que resulta fundamental para una función ejecutiva exitosa, la regulación del comportamiento y la adaptación social.

El control inhibitorio comprende tres procesos interrelacionados (Barkley, 1997, 2012). El primero de ellos corresponde a la inhibición de la respuesta prepotente inicial a un evento, es decir, una respuesta para la que se dispone de refuerzo inmediato (positivo o negativo) o con el cual se ha asociado un refuerzo en el pasado; el segundo es la detención de la respuesta en curso o del patrón de respuesta, lo cual provoca un retraso en la decisión para responder o continuar respondiendo y, en tercer lugar, el control de la interferencia, definida como la protección del período de retraso y de las respuestas autodirigidas que se producen en su interior.

En esta misma línea, Etchepareborda (2004) considera que la inhibición comprende tres habilidades coexistentes: control de la espera, control de impulsos y control de la interferencia. El primero corresponde a la habilidad para aguardar una situación de expectativa, para responder luego de la aparición de un estímulo *target*; depende del desarrollo de circuitos neocorticales de la corteza prefrontal de ambos hemisferios y es observable a partir de los 2 años de edad.

El control de impulsos se define por la habilidad para analizar un estímulo (señal o situación) antes de actuar en consecuencia. Implica retrasar la respuesta prepotente y

detener el curso de acción (Etchepareborda, 2004). Su desarrollo puede observarse desde los 4 años aproximadamente y está asociado a la maduración de regiones diencefálicas y de la corteza orbitofrontal (Fuster, 2002). Puede registrarse también mediante las comisiones simples en los test de vigilancia y atención sostenida, como por errores de comisión *non-target*, por pérdida de la meta propuesta o fallo en la discriminación del estímulo (Palencia Avendaño, Mogollón Rincón & Etchepareborda, 2016).

Por otro lado, el control de interferencia se define por la capacidad de restringir estímulos irrelevantes de otras áreas asociativas y modalidades sensorio-perceptivas (Etchepareborda, 2004), que se inicia a partir de los 6 a 7 años y va aumentando a lo largo de la infancia y la adolescencia.

El grado de control de interferencia puede apreciarse particularmente en el desempeño ante la tercera tarea Stroop, tanto por las latencias de respuesta en contraste con las otras dos tareas, como por las respuestas correctas y los errores.

Respecto de la comisión de errores, Etchepareborda y Comas (1994) distinguen entre los errores congruentes y las imprecisiones (o errores incongruentes). Definen como error congruente a aquel que se comete cuando se invierte el criterio de selección de respuesta en el Stroop, respondiendo con lectura y no con el color de las letras. Dicho error presupone un pobre control inhibitorio y por tanto un desbalance noradrenérgico-dopaminérgico (Palencia Avendaño, Mogollón Rincón & Etchepareborda, 2016).

Mientras que el error de imprecisión ocurre cuando no se respeta criterio de selección alguno, que puede interpretarse como indicador básico de impulsividad (Etchepareborda, 2004). En este caso, la imprecisión puede asociarse a un bajo nivel dopaminérgico (inadecuado alertamiento), o también un fallo en la discriminación del estímulo, por bajo nivel noradrenérgico (Tamm, Menon & Reiss, 2002; Paiva Barón, Saona Olivera, & Perna Ramos, 2009; Palencia Avendaño, Mogollón Rincón & Etchepareborda, 2016).

### **Desarrollo del Control Inhibitorio**

Distintos autores plantean que las habilidades para inhibir o retrasar de manera activa una respuesta dominante se adquieren en el curso del neurodesarrollo (Roselli et

al., 2002; Tirapu-Ustárrroz et al., 2012); al tiempo que tienen implicaciones en el temperamento y la emoción, la modificación consciente del comportamiento y la cognición (Barkley, 1997; Faraone & Biederman, 1998; Posner & Rothbart, 2000).

Investigaciones en distintas muestras reportan variaciones del Efecto Stroop asociados a la edad, lo que sugiere un progresivo desarrollo del control inhibitorio. Se ha observado un pico de mayor interferencia entre los 7 y 8 años, que decrece en los años siguientes (Comalli, Wapner & Werner, 1962; Golden, 1878).

Así también se halla un incremento en la interferencia Stroop en adultos, mayores de 65 años aproximadamente (Houx, Jolles & Vreeling, 1993; West & Alain, 2000; Van der Elst, Van Boxtel, Van Breukelen & Jolles, 2006). En tanto que en niños pequeños (alrededor de 6 años) no se halla la interferencia palabra-color (Fuentes et al, 2003).

No obstante, según León-Carrión, García-Orza & Pérez-Santamaría (2004), los hallazgos sobre la trayectoria del desarrollo mediante estudios en los que fue usado el Test Stroop algunas veces son contradictorios.

Por ejemplo, para Welsh et al. (1991), la maduración del control inhibitorio se logra alrededor de los 10 a 12 años; mientras que Comalli, Wapner y Werner (1962) sugieren que el incremento de la inhibición prosigue hasta los 16 a 18 años; y para Golden (1999) la adecuada resistencia a la interferencia Stroop predomina en el rango de 16 a 44 años.

Esta discrepancia puede darse por dos razones: a) el uso de las mediciones de la tarea de lectura en el cálculo de la interferencia, según la prueba clásica del Stroop que presupone el logro de los procesos lectores, b) el uso de variantes del test original, en tanto que pueden ocurrir que se examinen componentes distintos del mecanismo inhibitorio.

Como señalan Ludwig, Borella, Tettamanti y Ribaupierre (2010), el formato del test utilizado y las manipulaciones que se hagan de las mediciones pueden incidir sobre la interpretación del Stroop.

Los estudios del desarrollo y los cambios corticales observados durante la ejecución del Stroop igualmente dan cuenta de diferencias significativas asociadas a la edad.

Adleman et al. (2002) encuentran que los jóvenes (18 a 22 años) muestran una mayor activación cortical que los niños (7 a 11 años), en las áreas del cíngulo anterior y regiones parietales y parietoccipitales del hemisferio izquierdo, así como del giro medio frontal izquierdo. Igualmente, los jóvenes y los adolescentes (12 a 16 años) muestran una mayor activación general del córtex parietal.

Otros autores proponen que los resultados del test Stroop pueden relacionarse con el desarrollo funcional del lóbulo parietal en la adolescencia, mientras que el córtex prefrontal continúa su desarrollo en la adultez (Houx, Jollesb& Vreeling, 1993; Bugg, Delosh, Davalos & Davis, 2007; Cabeza & Nyberg, 2000; Fuster, 2008).

Tanto la inhibición como la capacidad de demorar una respuesta son esenciales del sistema ejecutivo, por lo que un acto impulsivo puede entenderse como un fallo del funcionamiento ejecutivo (Barkley, Fischer, Smallish & Fletcher, 2002; Barkley, 2012), debido a lesión cortical, disfunción o alteraciones en la neuromaduración.

Es importante mencionar que la observación de un rendimiento disminuido en el Stroop en grupos clínicos. Este ha sido definido como un déficit en el procesamiento de la información, caracterizado por un pobre control inhibitorio, enlentecimiento de los tiempos de respuesta y mayor comisión de errores para las tareas conflictivas (Schachar, Tannock, Marriott & Logan, 1995; Weiler, Bernstein, Bellinger & Waber, 2002; Van Mourik et al 2009; Christiansen & Oades, 2010)

En algunos casos, dicha disminución del rendimiento está asociado a un deterioro en los mecanismos de inhibición frontal, necesarios para el procesamiento de la información y el control atencional (Pliszka, Liotti & Woldorff, 2000).

Se ha señalado que en las tareas del Stroop, particularmente ante estímulos incongruentes, se activa inicialmente la corteza cingulada anterior, seguida de la corteza temporoparietal izquierda, posiblemente relacionada con la necesidad de un procesamiento adicional del significado de las palabras (Liotti, Woldorff, Pérez & Mayberg, 2000), lo que genera un incremento en los tiempos de respuesta, independientemente de la modalidad verbal o manual.

De modo que, una posible explicación para las dificultades de alertamiento, orientación y atención ejecutiva, identificadas en los trastornos de neurodesarrollo, es el reconocimiento de una maduración cortical más tardía, particularmente de regiones

cinguladas y de circuitos frontotemporales, entre otras, que pueden incidir en las mediciones del Stroop en determinadas muestras clínicas (Hervey et al., 2006; Depue et al., 2010; Mullane et al., 2010; Sjowall, Roth, Lindqvist & Thorell, 2012).

### **Trastornos de Neurodesarrollo**

Los trastornos de neurodesarrollo comprenden aquellos problemas de inicio temprano, vinculados a un funcionamiento deficitario, alterado o inmaduro del sistema nervioso (Artigas-Pallarés, 2009). Estos pueden agruparse en: a) trastornos debido a alteraciones genéticas bien identificadas, b) casos de retraso mental de causa desconocida, c) trastornos con una causa ambiental conocida (alcoholismo fetal, desnutrición, etc.). y d) trastornos que pueden definirse a partir de déficit cognitivos o conductuales.

En este último se incluyen a su vez los Trastornos por Déficit Atencional e Hiperactividad (TDAH), los Trastornos de aprendizaje (TAPZ), los Trastornos del Espectro Autista (TEA) y los Trastornos del lenguaje y la comunicación (TEL).

En el CIE-10 (OMS, 1995) dichos trastornos aparecen separados en una de dos categorías más generales: “Trastornos del desarrollo psicológico” y “Trastornos emocionales y del comportamiento que aparecen habitualmente en la niñez o en la adolescencia”. Por su parte, en el DSM-IV (APA, 2002) se los define dentro de los “Trastornos de inicio en la infancia, la niñez o la adolescencia” y, más recientemente, en el DSM-V (APA, 2014), en la categoría de los “Trastornos del desarrollo neurológico”.

Los avances en neurociencias sitúan el TDAH como uno de los problemas de neurodesarrollo más prominentes, cuyas manifestaciones se observan especialmente en la organización funcional del sistema nervioso central, en los sistemas de alerta, atención y regulación, afectando la conducta y las relaciones interpersonales en distintos ámbitos (Barkley, 1997a; Faraone & Biederman, 1998; Etchepareborda, 2000; Castellanos & Acosta, 2002, 2004; APA, 2002, 2014; Martínez-Morga, Quesada-Rico, Bueno & Martínez, 2018).

En tanto que los TAPZ representan la consecuencia de una disfunción cerebral, en general de carácter más leve, pero con consecuencias de considerable importancia, ya que perturban el aprendizaje y se manifiestan principalmente en el desempeño escolar (APA,

2002, 2014). Se describe como dominante para la lectura, la escritura o las aptitudes matemáticas, al tiempo que compromete tanto la comprensión como la expresión, aun cuando exista una capacidad intelectual normal (Campos-Castelló, 2000; Artigas-Pallarés, 2009).

Las dificultades de aprendizaje con frecuencia aparecen asociadas al déficit atencional, al tiempo que las dificultades cognitivas propias del TDAH sin duda interfieren en el aprendizaje.

La tasa de prevalencia para los trastornos de neurodesarrollo es un tema controvertido. En el DSM-V (APA, 2014) se establece para el TDAH una ocurrencia del 5 % y en el TAPZ un intervalo entre 5 y 15 % en niños de edad escolar de diferentes lenguas y culturas. No obstante, distintos autores admiten que la variabilidad de estos trastornos es muy elevada (Polanczyk, Silva de Lima, Lessa Horta, Biederman & Rohde, 2007; Cardo et al., 2010; Jiménez, 2012).

En la población española y sudamericana se informa una prevalencia del 11 al 17 % para los trastornos del neurodesarrollo, de los cuales el TDAH, los trastornos del lenguaje y los trastornos del aprendizaje constituyen la mayor parte (Pineda, Lopera, Henao Palacios & Castellanos, 2001; Sociedad Argentina de Pediatría, 2017; Carballal et al., 2018)

### Proceso de Normatización del Test Stroop

Las primeras normas de la prueba fueron presentadas por Stroop (1935) en una versión editada y distribuida por Stoeling Company durante 10 años aproximadamente. En esta se utiliza como referente psicométrico los resultados de sus experimentos, aunque en la presentación de los materiales hace una abreviación y considera sólo el segundo experimento, aquél que había mostrado una diferencia mayor entre los estímulos de color neutros y los estímulos incongruentes o conflictivos.

La prueba comprende tres láminas o tareas (lectura de nombres de colores, nominación del color presentado en recuadros pintados y nominación del color en estímulos de conflicto palabra-color). Cada una con 100 ítems dispuestos en columnas, en las que se alternan cinco colores (*red, blue, green, brown, purple*).

Nuevas versiones aparecieron posteriormente, reafirmando la interpretación del Efecto Stroop, que se define como fenómeno de demora al resolver el estímulo incongruente palabra-color, en donde las palabras impresas en colores parecen activar una respuesta de verbal automática que interfiere con la instrucción para nombrar los colores.

Jensen y Rhower (1966) hacen un primer recorrido por las variaciones del test Stroop surgidas entonces en distintos centros académicos, en Estados Unidos principalmente, y destacan las posibilidades que ofrecen las adaptaciones y réplicas del test original.

La adaptación de Thurstone de 1944 (citado por Jensen, 1966 y MacLeod, 1991a) es la más próxima a la forma estándar y hace parte de una de las primeras baterías de evaluación, cuyo rigor comprende un análisis factorial múltiple para variables cognitivas operacionalizadas mediante esta y otras pruebas. Su versión usa tres láminas, aunque en fondo negro para incrementar el contraste, alternando cuatro colores (*red, green, yellow, blue*).

### La prueba CWST de Golden (1978)

Unas décadas después aparece la adaptación CWST (Color-Word Stroop Test) de Golden (1978). La prueba introduce algunas modificaciones: utiliza tres colores (*blue, red, green*), modifica la secuencia de rectángulos de color, para presentar en su lugar una seguidilla de letras (**XXXX XXXX XXXX**), y propone también una limitación al tiempo



de respuesta de 45 segundos para cada lámina, registrando total de ítems logrados, número de aciertos y número de errores.

Mantienen este mismo esquema las actualizaciones posteriores las adaptaciones del CWST en español que fueron presentadas por la Editorial TEA en 1994, 1999, 2001, 2007 y 2010, para su difusión en España e Hispanoamérica.

La norma de corrección e interpretación de resultados que suele usarse en nuestro contexto corresponde a la segunda edición de Golden (1999), o las revisiones siguientes (Golden, 2001, 2007), cuyo proceso de normatización y tipificación se hizo con una muestra de población española ( $n=254$ ), con un rango amplio de edad, entre 7 y 80 años.

Estudios posteriores en otras muestras españolas concuerdan con estos resultados y suelen seguir el método de cuantificación de la interferencia propuesto en el manual de la prueba, en el que se combinan los resultados de las tres tareas, previa corrección de las puntuaciones brutas a las que se suma un valor asignado por edad.

Tal es el caso de Martín, Hernández, Rodríguez, García, Díaz y Jiménez (2012), que presentan los datos normativos en niños y niñas sanos de 6 a 12 años ( $n=1003$ ), utilizando la versión española de Golden (2001), y el de Rognoni, et al. (2012), que presentan análisis de los test Stroop y Torre de Londres, como parte del estudio en jóvenes españoles ( $n=179$ ).

### **Adaptaciones en Hispanoamérica**

En el continente americano han sido publicados algunos estudios de normatización del Stroop en español. En una revisión cronológica pueden citarse, por ejemplo, a Busigó, Martínó, Mercado, Colón y Pérez (1998) que validan una versión similar del Stroop en un grupo de niños y niñas de 9 años, en cinco regiones de Puerto Rico ( $n=108$ ).

En Chile, Conca Binfá e Ibarra González (2004) realizan el proceso de estandarización de una versión del Stroop, en un grupo de niños de 8 a 12 años escolarizados, de distintos niveles socioeconómicos ( $n=600$ ). Y, más adelante, lo hacen Maureira, Aravena, Gálvez y Flores (2014), en estudiantes universitarios ( $n=151$ ). Mientras que Fernández y Bendersky (2004), en Argentina, analizan los resultados de una muestra adultos sanos, de ambos sexos y distintos niveles educativos ( $n=120$ ).

Más recientemente, Rivera, et al. (2015), en un amplio estudio en 11 países (Argentina, Bolivia, Chile, Cuba, El Salvador, Guatemala, Honduras, México, Paraguay, Perú y Puerto Rico), con una muestra grande de adultos sanos ( $n=3977$ ), utilizan el protocolo de Golden (2007) para la interpretación de resultados. Luego, Rivera et al. (2017) presentan las propiedades psicométricas del Stroop, también en una muestra grande de niños y niñas escolarizados ( $n=4373$ ) de España y 10 países hispanoamericanos (esta vez sin incluir Argentina).

Y, en Colombia, Rodríguez, Pulido y Pineda (2017) destacan las características de validez y confiabilidad de una versión del Stroop de Golden (1994) que fue aplicado a una muestra de personas sanas, entre 7 y 80 años ( $n=1332$ ).

### **Variantes en Otros Idiomas**

Diferentes estudios de adaptación y validación de la prueba se han hecho en otros idiomas. Tal es el caso de Vanier (1991) que presenta una traducción al francés de la versión estandarizada de Golden (1978) e incluyen los resultados psicométricos en una pequeña muestra de sujetos canadienses sanos ( $n=59$ ) y de pacientes que sufrieron traumatismo craneoencefálico ( $n=32$ ), en un seguimiento longitudinal de los mismos.

Más adelante, Tremblay et al. (2016) realizan la validación del *Victoria Stroop Test* en adultos y ancianos sanos de Quebec ( $n=646$ ), sugiriendo que esta adaptación favorece la discriminación entre los estados cognitivos normales y patológicos.

En italiano pueden encontrarse igualmente varios trabajos. Caffarra, et al. (2002) presentan los resultados de una muestra de adultos sanos y con demencia ( $n=248$ ) y analizan el Efecto Stroop estimando el tiempo de ejecución y el número de errores. Valgimigli, et al. (2010), luego de analizar una muestra de adultos sanos ( $n=160$ ), proponen una medida única de interferencia (aciertos en unidad de tiempo). Y Brugnolo, et al. (2016), mediante análisis de regresión, evalúan factores como la edad, el nivel educativo y el género, en un grupo de sujetos normales ( $n=192$ ).

En Grecia, Zalonis et al. (2009) considera la prueba Stroop como instrumento de detección neuropsicológico (*screening*) y presenta los datos normativos en una muestra de adultos entre 18 y 84 años ( $n=605$ ).

En Países Bajos, Valentijn et al (2005) relacionan el funcionamiento sensorial y cognitivo en un estudio longitudinal con adultos de 55 años o más ( $n=418$ ). Seguidamente Van der Elst, Van Boxtel, Van Breukelen y Jolles (2006, 2008) presentan los datos normativos de una versión del Stroop, en un estudio con adultos sanos entre 24 y 81 años ( $n=1856$ ), señalando que el método basado en la regresión es apropiado para derivar normas para las puntuaciones dependientes de la velocidad (tiempos de respuesta), pero menos significativo al analizar las medidas de precisión en las tres tareas del Stroop (puntuación de aciertos).

Por su parte, Seo, et al. (2008) validan una versión coreana del Stroop y exploran los efectos de variables sociodemográficas en su desempeño, tomando como muestra un grupo de adultos entre 60 y 90 años ( $n=564$ ). También Kang, Lee, Yi, McPherson, et al. (2013), utilizando una versión abreviada del Stroop, proporcionan los datos normativos para la interpretación de los resultados en una muestra de adultos sanos entre 50 y 89 años ( $n=153$ ).

Asimismo, se utilizan adaptaciones del Stroop en dos sistemas de escritura de la ortografía japonesa que corresponde a procesamientos diferenciables: *kana* (de símbolos silábicos) y *kanji* (de logogramas). Las activaciones corticales que se identifican ante uno u otro tipo de escritura fueron observadas por Coderre, Filipi, Newhouse y Dumas (2008), al utilizar resonancia magnética funcional en una muestra pequeña de mujeres ( $n=7$ ), así como por Watanabe, Sumitani, Hosokawa y Ohmori (2015), en un estudio con espectroscopía, en un grupo de mujeres y hombres sanos ( $n=100$ ).

Morrow (2013) presenta los datos normativos de una adaptación de la prueba Stroop de Golden (1978) para uso de hablantes en inglés de Canadá y Estados Unidos ( $n=146$ ), hallando consistencia con los datos normativos de la prueba original. Mientras que Erdodi, et al. (2018) analizan los indicadores psicométricos de una versión del Stroop en una muestra de pacientes con síntomas somáticos, ansiedad y depresión ( $n=132$ ).

Por su parte, Roselli et al. (2002) realizan la comparación de muestras de adultos ( $n=122$ ) monolingües en inglés, monolingües en español y bilingües, encontrando que éstos últimos tienden a demorar aún más su respuesta ante los estímulos conflictivos.

En portugués, Zimmermann, Cardoso, Trentini, Grassi-Oliveira y Fonseca (2015) presentan un estudio preliminar de normatización para una adaptación de la prueba de

Stroop (1935) en una muestra de jóvenes y adultos, entre 19 y 75 años ( $n=158$ ). Luego, Oliveira, Mograbi, Gabrig y Charchat-Fichman (2016) contrastan los resultados de tres pruebas de evaluación cognitiva de memoria y atención en un grupo de niños y niñas brasileños ( $n=390$ ), con edades entre 7 y 14 años.

### **Observaciones sobre la Proliferación de Versiones del Stroop**

En consecutivas publicaciones se han modificado sistemáticamente las características de los ítems presentados en las láminas del Stroop, tales como el número de estímulos, la selección de los colores y tintas, el tamaño, el fondo o marco del estímulo, la dirección de las palabras, el orden de las tareas, etc. Ninguna de esas variaciones elimina el Efecto Stroop.

Las distintas versiones del Test comparten el mismo paradigma, que implica una tarea de lectura de palabras presentadas en letras negras, una tarea de control que requiere nombrar los colores de los elementos que no son palabras (estímulo neutro) y una tarea de interferencia que requiere inhibir la lectura y más bien decir el color tinta (estímulo de conflicto palabra-color), en la que se puede observar como efecto generalizado el aumento del tiempo de respuesta.

Anticipándose al gran número de versiones y variantes del test, Jensen y Rohwer (1966) habían observado que ninguna forma de la prueba de Stroop ha sido adoptada en forma exclusiva por los investigadores, y que ninguna de las formas existentes parece ser considerada “ideal” en todos los aspectos.

Tal como reseña Fernández (2014), las versiones del Stroop pueden agruparse según la cantidad de ensayos, el formato de aplicación y el número de ítems mostrados (entre 17 y 176), la disposición o presentación del estímulo (de forma vertical u horizontal, manual u oral, en versión computarizada o en papel, etc.), el número de colores utilizados (3 a 5 opciones generalmente) y el sistema de puntuación (tiempo de respuesta, número de respuestas correctas, número de errores).

En términos generales, las variaciones en cuanto a la forma de aplicación de las tareas presentan resultados similares en distintos grupos muestrales, bien sea que se trate de un listado, estímulos presentados individualmente o en bloques, estímulos sorteados al

azar, etc. (Jensen, 1965; Dyer, 1973; Dunbar & McLeod, 1984; McLeod, 1991a), que aparecen sobre papel o en una pantalla.

Salo, Henik y Robertson (2001) comparan el desempeño en tres formatos de presentación en una pantalla de los estímulos en bloques y uno a uno, tomando la respuesta verbal de los participantes. Si bien los autores observan consistencia intraindividual, señalan que las presentaciones de la tarea tipo Stroop pueden producir diferentes niveles de interferencia, por lo que es importante considerar un análisis más detallado de las formas alternativas del Stroop y, especialmente, de los métodos de cuantificación de la interferencia.

Distintas versiones del test proponen algunas variantes respecto del número total de ítems presentados por lámina o del número de colores usados, sin que esto afecte de modo significativo la observación del Efecto Stroop. No obstante, Bugg, DeLosh, Davalos y Davis (2007) afirman que las versiones abreviadas del test pueden ser más sensibles a las diferencias por edad.

### **Stroop Emocional, Pictórico, Numérico y Otras Adaptaciones**

Existen variantes Stroop que no están centradas en la interferencia semántica palabra-color, aunque mantienen la idea de analizar la respuesta dada ante elementos incongruentes. Estas pueden agruparse según diferentes categorías.

Por ejemplo, aquellas pruebas que refieren como conflicto una intrusión emocional en la respuesta: Stroop emocional (Compton et al., 2003; McKenna & Sharma, 2004; Larsen, Mercer & Balota, 2006; Koizumi et al., 2007; Gantiva et al., 2012; Toro, Arias & Avella, 2013).

Una versión disponible por internet del Stroop emocional es el llamado Tastiva, acrónimo de Tareas Stroop de tiempo variable, de Scurtu, Manzano-Arrondo y Rodríguez Testal (2016), que incluye palabras de contenido neutro y sexual.

Entre otras variantes de éste pueden citarse también el Stroop facial (Urrego, Restrepo, Pinzón, Acosta, Díaz & Bonilla, 2014), las pruebas de generación semántica (Dos Santos Assef, Seabra Capovilla & Capovilla, 2007) y el Stroop de alcohol (Sánchez-López et al., 2015).

Están también aquellas pruebas que usan significados de la naturaleza, en su mayoría con elementos pictóricos que pueden usarse con lectores poco entrenados, por ejemplo, el Stroop léxico de tarea palabra-dibujo (Wilbourn, Kurtz & Kalia, 2012). O también en personas que no leen, como en la versión del Stroop Día-Noche (Gerstadt, Hong & Diamond, 1994; Cuevas, Calkins & Bell, 2016), Test Luna-Sol (Archibald & Kerns, 1999), Stroop de Frutas (Okuzumi et al., 2015) y Stroop de animales o Stroop de quimeras (Adam & Jarrold, 2009; Ikeda, Okuzumi & Kokubun, 2014), entre otros.

Otras son las adaptaciones con estímulos neutros e incongruentes de cantidad y magnitud, por ejemplo: Test de los cinco dígitos (Sedó, 2004) y el Stroop numérico (Kaufmann et al. 2005; Heine et al., 2010).

Se encuentran también adaptaciones diseñadas para modalidades sensoriales específicas, entre ellas: Prueba de variaciones cromáticas (Kim, Cho, Yamaguchi & Proctor, 2008) y Prueba de estímulos visuales y gustativos (Razumiejczyk, Macbeth & Adrover, 2011).

Deben mencionarse las variantes de la prueba Stroop que son incluidas como parte de baterías de evaluación. Muestra de ello es la Batería Neuropsicológica de Funciones Ejecutivas y Lóbulos Frontales (BANFE) de Flores Lázaro, Ostrosky Shejet y Lozano Gutiérrez (2014), donde los resultados de la tarea de conflicto Stroop se utilizan como indicador de la capacidad de control inhibitorio, relativa al funcionamiento de áreas corticales orbitofrontales y prefrontales medias.

Lo mismo ocurre dentro de la batería *Leiter International Performance Scale* (Leiter-3) se incluye el *Nonverbal Stroop Card Sorting Test* (NSCST), de Koch y Roid (2012), utilizada para evaluar atención, inteligencia y habilidades cognitivas; con una aplicación para Android del NSCST, como instrumento para la evaluación del funcionamiento ejecutivo en personas con problemas de lenguaje.

Y también el sistema computarizado *Mindstreams™* de *NeuroTrax™*, de Schweiger, Doniger, Dwolatzky, Jaffe y Simon (2003), diseñado para detectar deterioro cognitivo leve, que incluye una adaptación del Stroop para evaluar capacidad de la atención, tiempo de reacción y precisión en la ejecución de la tarea.

### **Propiedades Psicométricas del Stroop**

La aplicación de uno o varios métodos para corroborar la estabilidad temporal de una prueba psicométrica, su congruencia o consistencia interna, así como el análisis de las partes que lo componen, su extensión o las formas de aplicación, dan cuenta de la confiabilidad de esta (Hernández Sampieri, Fernández Collado & Baptista, 2014). Es decir que pueden replicarse o reproducirse al menos con un cierto margen de error. En tanto que la veracidad, exactitud, autenticidad y solidez de las mediciones realizadas, dan cuenta de su validez, permitiendo determinar si mide lo que dice medir.

Según Argibay (2006), la confiabilidad es esencialmente una cuestión empírica, tiene que ver con la precisión del procedimiento de medición, mientras que la validez trata más de los elementos teóricos implicados en la prueba y de la utilidad en el registro de los valores medidos, esto es, la autenticidad de la medición respecto de un atributo, hipótesis o concepto.

Por lo general, en el estudio del Stroop se usan métodos basados en la covarianza de los ítems, mediante coeficiente Alfa de Cronbach, mediciones test-retest para estimar su estabilidad en el tiempo, y correlación de Pearson para el análisis de asociación entre tareas y para comparar instrumentos que se consideran paralelos.

De otra parte, dado que la definición de validez por sí misma incluye más elementos teóricos, ya que persigue la explicación teórica, puede instrumentarse de distintas maneras (Argibay, 2006; Prieto & Delgado, 2010; Hernández Sampieri, et al., 2014).

Para evaluar la validez de contenido, por lo general se consulta el juicio de expertos o se realizan aplicaciones piloto. La validez de constructo puede evaluarse mediante el análisis factorial, la diferenciación entre grupos, las correlaciones con otras medidas del mismo constructo, las matrices multimétodo-multirasgo. Y para determinar la validez de criterio es necesario relacionar las puntuaciones con otras variables que puedan corroborarse en un momento dado o predecir su desempeño futuro, utilizando correlaciones, contrastes múltiples, coeficientes de contingencia, entre otras técnicas estadísticas.

Ahora bien, toda vez que la presente investigación está centrada en el Test Stroop, resulta imprescindible revisar los argumentos psicométricos de éste, preguntando por los

procedimientos que se usan para determinar su confiabilidad y los resultados empíricos que así lo confirman, e igualmente, por el tipo de validez que posee y las inferencias conceptuales que permite, en especial respecto de los procesos atencionales o de control de interferencia.

En líneas generales, cada uno de los manuales de las pruebas normatizadas del Stroop, en su formato clásico o de alguna de sus adaptaciones y variantes, describen resultados de confiabilidad y validez estadísticamente significativos (entre moderados y altos).

Distintas revisiones y metaanálisis coinciden al concluir que el Stroop es probablemente la prueba psicométrica con el mayor número de estudios de confiabilidad como de validez. Por tanto, se dice que es la de mayor robustez en términos estadísticos (Jensen & Rohwer, 1966; MacLeod, 1991b; Van Mourik, Oosterlaan & Sergeant, 2005; Lansbergen, Kenemans & Engeland, 2007; Ben-David, Nguyen & Van Lieshout 2011; Scarpina y Tagini, 2017).

### **Hallazgos de Confiabilidad en el Stroop**

Se reconoce que, independientemente de la versión utilizada, las mediciones Stroop son estables en cuanto a las diferencias individuales (Jensen, 1965; Golden, 1978; Koch, Gobell & Roid, 1999; Smolen & Chuderski, 2010), en lo que parecen ser tres aspectos bastante simples y básicos del desempeño cognitivo humano: la lectura de palabras, la nominación de color y la resistencia a la interferencia palabra-color; con algunos efectos de práctica evidentes de manera consistente (Monahan, 1997; Davidson, Zacks & Williams, 2003; Ziaka, Moirou, Vlahou & Protopapas, 2011).

Aunque, según Homack y Riccio (2004) no se ha demostrado la confiabilidad con ninguna de las puntuaciones de interferencia derivadas (índices), señalamiento que coincide con la carencia de un método único de cuantificación del control de interferencia (Jensen & Rohwer, 1966; Van Mourik, Oosterlaan & Sergeant, 2005; Lansbergen, Kenemans & Engeland, 2007; Scarpina & Tagini, 2017).

Como se menciona antes, las estandarizaciones del Stroop se han realizado por lo general para las versiones de aplicación manual y respuesta verbal, en tanto que distintas versiones computarizadas se han usado en trabajos experimentales, por lo que los



referentes de baremación, como las mediciones de estabilidad y congruencia interna, suelen estar en los primeros.

Análisis test-retest en el Stroop muestran coeficientes de confiabilidad altos en las tareas de lectura y nominación de color en estímulos neutros, aunque levemente más bajos en la tarea de conflicto, considerando las mediciones de latencias de respuesta y respuestas correctas (Jensen, 1965; Jensen & Rhower, 1966; Golden, 1978; Gualtieri & Johnson, 2006; Van der Elst et al., 2006, 2009), en tanto que para los puntajes derivados y los cálculos de la interferencia presentan mayor variabilidad (Lansbergen, Kenemans & Van Engeland, 2007).

En menor proporción aparecen estudios con coeficientes Alfa de Cronbach, computados para las tres láminas del Stroop. Por ejemplo, Ludwig, Borella, Tettamanti y Ribaupierre (2010) refieren una confiabilidad satisfactoria del tiempo de respuesta, pero no para las mediciones de errores. Mientras que Gutiérrez-Martínez, Ramos-Ortega y Vila-Chaves (2018) refieren coeficientes excelentes para las mediciones de aciertos y tiempos de respuesta.

Es importante mencionar que en todas las versiones se corrobora la observación del efecto de interferencia, aunque en diferente grado. Asimismo, los estudios coinciden en mostrar una mayor velocidad de respuesta o número de aciertos en el formato tradicional de respuesta verbal, no obstante, también tienden a presentar puntuaciones más bajas en la tarea de conflicto y en las cuantificaciones del control de interferencia.

### **Hallazgos de Validez en el Stroop**

Como se describe anteriormente, el Stroop tiene un diseño sencillo y una aplicación muy breve, centrándose en las tres tareas. Tal especificidad puede reñir con la extensión exigida en otras pruebas, en los formatos de tipo cuestionario, por ejemplo, en las que la validez de contenido está dada por una muestra representativa y suficiente del universo de conductas para evaluar el atributo en cuestión.

No obstante, se ha observado que variaciones en el tamaño y duración de cualquiera de las versiones Stroop no presenta resultados significativamente diferentes (Jensen, 1965; Dyer, 1973; Dunbar & McLeod, 1984).

En los manuales de aplicación se refieren a aspectos específicos de la validez de constructo del Stroop, por cuanto se busca probar que las mediciones registradas puedan ser consideradas indicadores válidos de la interferencia cognitiva (Golden; 1978, 1999, 2001), según se consideren uno u otros modelos explicativos sobre automaticidad y velocidad de procesamiento, el control atencional, o del neurodesarrollo atencional y ejecutivo.

Ahora bien, al analizar la validez de constructo, para distintas medidas neuropsicológicas, es importante tener en cuenta la evidencia de sensibilidad a la disfunción del sistema nervioso central (Homack & Riccio, 2004).

Algunos estudios, asociados a lesión o disfunción cerebral y patologías psiquiátricas, corroboran un desempeño general disminuido para las tres tareas, no obstante, se mantiene el Efecto Stroop (Golden, 1978, 1999, 2001; López-Villalobos et al., 2010; Rognoni et al., 2012; Rubiales, Bakker & Urquijo, 2013).

Así también, hallazgos en neuropsicología confirman la utilidad del Stroop y su validez de criterio complementaria para el diagnóstico clínico. La regresión logística de López-Villalobos, et al. (2010) sugiere que el mejor predictor para el déficit atencional está formado por la variable edad y los resultados de la tarea de conflicto palabra-color, con una sensibilidad del 81% y una especificidad del 72%, así como validez de criterio complementaria para el diagnóstico de este trastorno (índice de 0.76), tomando como referencia los criterios del DSM-IV.

Por su parte, en el estudio de Sánchez López et al. (2015) se analiza el sesgo atencional en una variante del Stroop adaptado para en palabras asociadas al consumo de alcohol, que muestra una sensibilidad del 71% y una especificidad del 60%.

En otros estudios con versiones computarizadas se ha informado que los tiempos de reacción en la tarea de conflicto presentan validez de constructo, al informar mayor interferencia en grupos clínicos (Salo, Henik & Robertson, 2001; López-Villalobos et al., 2010; Penner et al., 2012; Sisco et al., 2016). Por lo que éstos pueden considerarse mediciones válidas para evaluar el desempeño neuropsicológico en grupos clínicos (Erdodi et al., 2018).

Asimismo, distintos resultados del Stroop han mostrado correlaciones significativas en el contraste con otros instrumentos, lo que sugiere la validez divergente y validez

convergente (Golden, 1978; Busigó, Martínó, Mercado & Pérez, 1998; Van der Elst et al., 2006, 2009; Wilbourn, Kurtz & Kalia, 2012; Morrow, 2013; Sisco et al., 2016).

Los contrastes se han hecho especialmente con subpruebas de la batería Weschler, tareas de atención sostenida y atención selectiva, escalas fonológicas y pruebas de vocabulario y, en general, pruebas de funciones ejecutivas: Torre de Hanoi, tareas de ejecución *go/no-go*, Trail Making Test, Prueba de cartas de Wisconsin, entre otros.

En otro sentido, según Argibay (2006), un test puede estar asociado a la validez interna en un proceso de investigación, permitiendo establecer las causas de las variaciones en la interacción de variables y rechazando interpretaciones alternativas; o puede dar lugar a la validez externa, en tanto posibilidad de generalización de los resultados o comprobación de que lo mismo ocurre en contextos más amplios.

La validación de los resultados del Stroop se ha apoyado en un soporte psicométrico robusto y amplio en diversas muestras. En el campo de la neurofisiología y neuropsicología su aplicación ha sido útil para corroborar el incremento en la actividad de cortical durante el Stroop y, en general, una activación deficitaria en casos clínicos (Banich et al., 2000; Salgado-Pineda et al., 2003; Compton, Banich et al., 2003; Hanslmayr et al., 2008; Roberts y Hall, 2008; Ramírez-Benítez et al. 2011, entre otros).

### **Estado del Arte**

Actualmente se encuentran disponibles todo tipo de revisiones, discusiones y hallazgos sobre el Test Stroop, el efecto Stroop, el control de interferencia, entre otros temas específicos asociados. A inicios del 2020, pueden restrearse un número aproximado de 184.000 citas en el buscador de Google Académico, que incluyen la palabra “Stroop”. Asimismo, otras búsquedas específicas por los términos “Stroop y funciones ejecutivas” o “Stroop y control inhibitorio”, por ejemplo, muestran numerosos enlaces en Internet o en las redes de revista científicas, como Redalyc, Scielo, Elsevier, entre otras.

La revisión que se presenta a continuación cita trabajos de reciente aparición, señalando el uso cada vez más extendido de formas computarizadas del Stroop, algunos antecedentes de comparación entre versiones y líneas de investigación actuales.

### **Introducción de Formas Computarizadas**

En las últimas décadas se han diseñado versiones computarizadas del Stroop, en especial como herramienta de investigaciones en procesos cognitivos y para contrastar neurodesarrollo normal y patológico, en las que se han adaptado las tareas originales presentadas en papel de las que se obtiene una respuesta verbal, presentando los estímulos en una pantalla y registrando la respuesta verbal y/o manual a las mismas tareas.

La versión de Etchepareborda y Comas (1994, 2001) es un formato similar al de Stroop (1935), usado tanto en España como en Argentina como parte de una batería computarizada de pruebas de evaluación cognitiva, que permite distinguir latencias de respuestas, número de aciertos y errores, porcentaje de desempeño, discriminando errores congruentes e incongruentes en la tercera tarea, como parámetro para identificar la interferencia.

Ludwig, Borella, Tettamanti y Ribaupierre (2010) usan dos formatos de aplicación, uno clásico (blocked version) y su adaptación computarizada (item-by-item), en el estudio de las diferencias según la edad en el Stroop. Por su parte, Penner, et al. (2012) comparan el rendimiento en tres tareas tipo Stroop, adaptadas de Zysset, et al. (2001) y Mead, et al. (2002), para analizar la disminución del efecto de interferencia en dos versiones computarizadas en comparación con la prueba original. Zarghi, et al.

(2012) muestran los resultados de evaluación de atención selectiva en niños y adultos, mediante una versión computarizada del Stroop, encontrando que los tiempos registrados y el número de errores aumentan de manera estable en tareas conflictivas.

Pueden encontrarse también variantes computarizadas del Stroop en investigación aplicada en salud. Tal es el caso de la batería CNS Vital Signs de cribado cognitivo, de Gualtieri y Johnson (2006), que fue desarrollada como una herramienta de aplicación breve, con capacidad discriminante para simuladores y pacientes con trastornos de conversión.

Los estudios de Córdova, Karnikowski, Pandossio y Nóbrega (2008) y de Barbosa, et al. (2010) presentan el software Testinpacs® utilizado para evaluar atención selectiva asociada al estrés psicológico agudo y reactividad cardiovascular, comparando sus resultados una adaptación de la versión tradicional. Mientras que Calleja y Hernández-Pozo (2009) desarrollan una versión computarizada que incluye un listado de palabras asociadas al consumo de tabaco.

Algunos autores resaltan las ventajas de las versiones computarizadas, por el control del tiempo, la reducción del efecto del evaluador y la precisión en las mediciones (Assef, Capovilla & Capovilla, 2007; Etchepareborda, Paiva Barón & Abad, 2009; Paiva Barón, Saona Olivera & Perna Ramos, 2009).

En este sentido, García-Oqueta (2001) resalta que las aplicaciones de las pruebas computarizadas permiten identificar otros parámetros de medición, contabilizar y discriminar el tipo de errores (omisiones, comisiones, respuestas precoces y perseveraciones) y facilitan el análisis de aspectos temporales del desempeño.

Recientemente se han hecho populares un sin número de aplicaciones para equipos móviles con juegos tipo Stroop que no cuentan con ningún estudio sobre su validez o confiabilidad, por ahora. No obstante, el uso de celulares o Tablet as como herramientas en la evaluación neuropsicológica podría convertirse en un campo nuevo para la clínica y la investigación.

### **Antecedentes de Comparación entre Formas**

Se encuentran estudios empíricos que comparan versiones del Stroop, y analizan las diferencias entre respuestas verbales y manuales. Uno de los primeros trabajos

comparativos es el de Redding y Gerjets (1977), que usa un listado en papel y tecleo en un taquistoscopio y, si bien se corrobora el efecto Stroop, la muestra reducida de 12 voluntarios no permite un análisis riguroso. Igual que en el trabajo de Logan y Zbrodoff (1998), que compara distintas formas del Stroop, utilizando un grupo de 36 voluntarios. Y también en el caso de Repovs (2004) que compara el registro de la actividad electrocortical en dos modos de aplicación del Stroop en una muestra de 22 universitarios.

Entre otros ejemplos, puede citarse también el estudio de Van Maanen, Van Rijn y Borst (2009), que analiza dos versiones del Stroop, una semejante a la aplicación tradicional (estímulos de palabras y colores, congruentes e incongruentes) y una prueba pictórica (dibujo de un perro con la palabra gato superpuesta, por ejemplo). En otra experiencia, Fernandes, Prada, Glanner, De Toledo Nóbrega y Córdova (2010) toman una muestra de 20 mujeres jóvenes, para comparar la reactividad cardiovascular inducida por la versión computarizada del Stroop comparada con la tradicional.

Por su parte, el trabajo de Ludwig, Borella, Tettamanti y Ribaupierre (2010), con una muestra de 78 jóvenes (18 a 35 años) y 150 adultos mayores (62 a 84 años), divididos a su vez en dos grupos, realiza la aplicación del Stroop en el formato clásico (blocked versión) y la adaptación computarizada (ítem-by-item). Y, Penner et al. (2012) comparan el rendimiento en tres tareas tipo Stroop, la prueba original y dos versiones computarizadas distintas, en 29 niños y 40 adultos.

En suma, los antecedentes de comparación entre versiones distintas del Stroop evidencian algunas fallas en cuanto al tipo y tamaño de las muestras, así como la necesidad de igualar la tarea entre versiones para minimizar el error, realizar comparaciones intrasujetos y estudiar la equivalencia entre versiones. La Tabla 1, a continuación, presenta una reseña de los hallazgos en distintas adaptaciones del Stroop que utilizan estímulos palabra-color. Específicamente se citan aquellos estudios que tuvieron como propósito comparar entre los formatos de respuesta verbal y respuesta manual.

**Tabla 1***Hallazgos de confiabilidad en estudios comparativos entre distintas versiones del Stroop*

Referencia	Versiones Stroop	Hallazgos
Monahan (1997)	Vocal <sup>a</sup> vs Keypress <sup>b</sup>	Mayor demora de la respuesta manual. Efecto de práctica.
Logan y Zbrodoff (1998)	Vocal <sup>a</sup> vs Arbitrary-keypress <sup>b</sup> vs Typewritten <sup>b</sup>	Mayor efecto de interferencia en Typewritten. Mayor número de aciertos en la respuesta verbal.
Salo, Henik y Robertson (2001)	Computarized clinical <sup>b</sup> vs Random Single-Trial <sup>b</sup> vs Blocked Single-Trial <sup>b</sup>	Diferencias significativas del tiempo de respuesta según la versión. Mayor confiabilidad de la versión clínica.
Roselli et al. (2002)	Spanish Stroop <sup>a</sup> vs English Stroop <sup>a</sup>	Efecto principal de la proficiencia e interacción proficiencia-versión.
Fuentes et al. (2003)	Card Stroop Task <sup>a</sup> vs Computarized version <sup>b</sup>	Mayor número de errores en la versión computarizada. Baja correlación entre ambas pruebas.
Repovs (2004)	Verbal response mode <sup>a</sup> vs Keyboard key press <sup>b</sup>	Mayores latencias en la respuesta manual. Diferencias tanto en la interferencia como en la facilitación.
Lansbergen, Keneman & Van Engeland (2007) <sup>c</sup>	Computer/Card - Vocal/Manual	Consistencia y variabilidad entre índices de interferencia.
Ludwig et al. (2010)	Blocked Color Stroop <sup>a</sup> vs Item-by-item Color Stroop <sup>b</sup>	Significativos efectos principales por edad en ambas versiones, e interacción edad-versión. Fiabilidad entre .69 y .88
Parsons, Courtney, Arismendi y Dawson (2011),	Paper-and-pencil <sup>a</sup> vs ANAM version <sup>b</sup> vs Virtual Reality <sup>a</sup>	Las versiones concuerdan en la nominación del color, pero tienen diferencias significativas en la interferencia. Mayores aciertos en las versiones de respuesta verbal.
Penner et al. (2012)	Victoria Stroop <sup>a</sup> vs Color matching task <sup>a</sup> vs Computarized color-word task <sup>b</sup>	Desempeño variable de los niños en las tres versiones; concordancia en los adultos. Fiabilidad entre .50 y .80
Sánchez-López et al. (2015)	Stroop clásico <sup>a</sup> vs Stroop Neutro <sup>a</sup> vs Stroop Alcohol <sup>a</sup>	Correlación positiva y alta entre las tres pruebas. Interacción con la dependencia alcohólica.
Gutiérrez-Martínez et al. (2018)	Cantidad-número <sup>b</sup> vs Palabra-color <sup>b</sup>	Fiabilidad entre .62 y .78 por aciertos; y .96 por latencias.

<sup>a</sup> Respuesta verbal. <sup>b</sup> Respuesta manual.<sup>c</sup> Metaanálisis con 19 estudios que comparan entre formato clásico de aplicación del Stroop, adaptaciones computarizadas y versión Golden (1978).

## Resultados del Stroop en Grupos Clínicos

La aplicación y el uso extensivo del Stroop, o alguna de sus variantes, ha sido consistente en la identificación de déficit ejecutivos, específicos para la atención selectiva, el control de interferencia y la flexibilidad cognitiva en niños y adolescentes con TDAH, TAPZ y otros trastornos, al compararlos con grupos controles (Langleben et al., 2006; Pliszka et al., 2006; Rubiales, 2012), con un tamaño del efecto de .50 o mayor (Homack & Riccio, 2003; Lansbergen, Kenemans & Van Engeland, 2007; King et al., 2007). La tabla 2 presenta algunos resultados específicos para la tercera tarea del Stroop en estudios con muestras clínicas y controles.

**Tabla 2**

*Antecedentes de comparación del desempeño en la tarea de conflicto palabra-color (PC) entre grupos clínicos y controles*

Referencia	Muestra	Grupo clínico	Controles
Chafetz y Matthews (2004)	Adultos en tratamiento y controles ( $n=105$ )	31.4 rc	46.8 rc
King, Colla, Brass y Von Cramon (2007)	Adultos con TDAH y controles ( $n=44$ )	999 ms	841 ms
Protopapas, Archonti y Skaloumbakas (2007)	Lectores hábiles y con déficit ( $n=88$ )	1334 ms	995 ms
Aran Filippetti y Mías (2009)	Menores de 7 a 12 años, por nivel socioeconómico ( $n=254$ )	-	1750 ms (medio) 2230 ms (bajo)
Christiansen y Oades (2010)	Menores de 8 a 14 años con TDAH, sus hermanos y controles ( $n=96$ )	1289 ms	1187 ms
López-Villalobos et al. (2010)	Menores de 7 a 11 años con TDAH y controles ( $n=200$ )	22.8 rc	31.8 rc
Rubiales, Bakker y Urquijo (2013)	Menores de 8 a 14 años con TDAH y controles ( $n=60$ )	36.7 rc	43.8 rc
Giménez-Donoso et al. (2018)	Pacientes con trauma emocional ( $n=119$ )	42.0 rc (leve) 32.6 rc (intenso)	

*ms: Valores promedio en milisegundos. rc: Número de respuestas correctas.*

Una revisión más detallada de comparaciones en niños y adolescentes puede hallarse en los metaanálisis de Homack y Riccio (2004), Lansbergen, Kenemans y Van Engeland (2007) y Van Mourik, et al. (2009). Los autores examinan como las muestras



clínicas de pacientes con trastornos de neurodesarrollo coinciden en mostrar un peor performance en el Stroop, en comparación con grupos controles.

Entre las compilaciones y metaanálisis se encuentran publicaciones que agrupan resultados de diversas investigaciones y permiten rastrear datos específicos, tanto para el número de respuestas correctas como los tiempos de respuesta (latencias), consideradas las condiciones o tareas del Stroop que fueron usadas en cada experiencia (Van Mourik, Oosterlaan y Sergeant, 2005; Larsen, Mercer y Balota, 2006; Lansbergen, Kenemans y Van Engeland, 2007, Ben-David y Schneider, 2009). Tales registros bien pueden utilizarse como parámetro de comparación, ya que se trata de resultados empíricos en distintos grupos clínicos o en grupos de personas con un desarrollo típico de distintos rangos etarios y procedencias.

### **Líneas de Investigación Actuales**

Dos fenómenos se han dado recientemente en el estudio del Stroop. En primer lugar, el uso de muestras más amplias, como en el caso de Rivera et al. (2015), Rodríguez, Pulido y Pineda (2017) y Rivera et al. (2017), por ejemplo. Así como la presentación de una nueva tipificación española del Test Stroop de TEA Ediciones, adaptada por Ruiz, Luque y Sánchez-Sánchez (2020).

En segundo lugar, la insistencia en el uso de distintas herramientas neurofisiológicas de última tecnología para el estudio del Stroop, buscando un abordaje neurocognitivo integrador. En este sentido, los trabajos de Ferrand et al. (2019), Parris, Augustinova y Ferrand (2020), entre otros, señalan que el análisis discriminado de las modalidades de respuesta en la tarea de Stroop contribuyen a descubrir los componentes de la interferencia y pone en discusión las activaciones de regiones cinguladas asociadas a la resolución del Stroop, al menos para las respuestas vocales.

En el marco de las neurociencias se ha propuesto profundizar en distintos tópicos: a) los métodos utilizados para indexar los diferentes tipos de conflicto, b) los efectos moduladores del modo de respuesta, c) posibles efectos de Stroop en la ejecución de la respuesta, d) la aplicación de formas de análisis de distribución (por ejemplo, la función ex-gaussiana), y d) evidencia del control sobre las diversas formas de conflicto.

Pueden consultarse también otras publicaciones de muy reciente aparición que no alcanzaron a ser revisadas para la presente investigación (entre otros: Banich, 2019; Barnett, Sawyer & Moore, 2020; Hershman & Henik, 2020; Huang, Su & Ma, 2020; Kinoshita & Mills, 2020; Liefoghe, Hughes, Schmidt & De Houwer, 2020; Nicosia & Balota, 2020; Sobel, Puri & York, 2020; Spinelli & Lupker, 2020; Yoshihara et al., 2020). Se trata de estudios empíricos, meta-análisis y revisiones que buscan abordar tales cuestionamientos.

### **Mediciones en el Stroop**

Como se ha mencionado, distintas mediciones pueden obtenerse para las tareas de lectura (P), nominación de colores (C) y resolución de la tarea conflictiva palabra-color (PC). Según Jensen y Rohwer (1966), ninguna otra prueba psicológica arroja tantos puntajes diferentes como el Test Stroop.

Estas diversas puntuaciones caen en una de tres clases: a) los puntajes de tiempo básicos (o de número de ítems logrados en un tiempo determinado) y todos los puntajes derivados a los que dan lugar sus manipulaciones algebraicas, b) los refinamientos para calificar varios tipos de error y c) los patrones temporales de respuestas, no del rendimiento comparado entre las tareas, sino de los cambios en el curso de respuesta a una sola lámina, generalmente en la tarea de conflicto. Se reseñan a continuación hallazgos relativos a la primera categoría de mediciones.

#### **Puntajes Básicos**

Los puntajes básicos más frecuentemente utilizados comprenden: a) latencias de respuesta, dado el tiempo empleado para completar el listado o el promedio de tiempo usado para responder a un ítem, que suele interpretarse como velocidad de procesamiento, b) el registro del número de respuestas correctas, mediante los cuales se informa la precisión en el desempeño individual ante cada tarea o también el porcentaje de logro y, c) el número de fallos cometidos, en algunos casos, discriminando entre el tipo de error, para identificar la eficiencia para respetar la consigna y controlar la respuesta emitida.

#### **Parámetros de Latencias**

En algunas versiones se considera el tiempo empleado para leer o nombrar el color en el total de los ítems dispuestos (Stroop, 1953, Comalli, Wapner & Werner, 1962; Amato et al., 2006; Ludwig, Borella, Tettamanti & Ribaupierre, 2010). Otras versiones contabilizan el número de aciertos en un tiempo delimitado para cada tarea, por lo general de 45 segundos (Golden, 1978, 1999; Vanier, 2001; Chafetz & Matthews, 2004; Morrow, 2013; Rivera et al., 2015). Aunque también hay versiones cuyo límite de tiempo es de 30 segundos (Brugnolo et al., 2015) o de 120 segundos (Zalonis et al., 2009).

Una alternativa usada frecuentemente es la estimación de la latencia media de respuesta por ítem, que se obtiene al dividir el tiempo límite de la tarea sobre el número de respuestas, incluidos aciertos y errores (Lansbergen, Kenemans & Engeland, 2007; Tremblay et al., 2016). En tanto que las versiones computarizadas pueden dar cuenta de los valores exactos de latencias medias de aciertos o de errores (Etchepareborda, Paiva-Barón & Abad, 2009).

La tabla 3 presenta algunos resultados de distintos tiempos de respuesta registrados en las tres tareas del Stroop.

**Tabla 3**

*Reporte de tiempos de respuesta para las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)*

Referencia	Muestra	P	C	PC
<i>Tiempo empleado para responder a 100 ítems presentados. Valor en segundos.</i>				
Stroop (1935)	Estudiantes de pregrado ( <i>n</i> =70 y <i>n</i> =100)	41.0	63.3	110.3
Comalli, Wapner y Werner (1962)	Sujetos sanos de 7 a 80 años ( <i>n</i> =235)	39.4 a 89.9	56.1 a 126.9	103.0 a 264
Jensen (1965)	Estudiantes de pregrado ( <i>n</i> =436)	38.1	58.2	100.3
Houx, Jollwa y Vreeling (1993)	Voluntarios de 20 a 80 años ( <i>n</i> =247)	37.0 a 44.1	50.1 a 61.6	77.3 a 105.5
<i>Media estimada de respuesta por ítem. Valor en milisegundos.</i>				
Dunbar y MacLeod (1984)	Estudiantes de pregrado ( <i>n</i> =32)	510	620	780
MacLeod (1986)	Estudiantes de pregrado ( <i>n</i> =50 y <i>n</i> =40)	412	598	1023
Cohen, Dunbar y McClelland (1990)	Modelo computacional (simulación)	350 a 450	550 a 600	>850
Repovs (2004)	Estudiantes de pregrado ( <i>n</i> =22)	675 verbal 825 manual	625 verbal 690 manual	790 verbal 865 manual

La relación del desempeño en los ítems palabra-color representa casi el doble de tiempo que la tarea de lectura o más y suma aproximadamente un tercio más al desempeño en la nominación de color (Jensen, 1965; Golden, 1978; Dunbar & MacLeod, 1984; MacLeod, 1990; Lovett, 2005; Ríos Lago & Periañez Morales, 2010).

Sin embargo, estas maneras de responder pueden variar según la edad, el nivel de formación escolar o la presencia de ciertas patologías, como déficit atencional, dislexia, afasia, etc.

El análisis detallado de tiempos de reacción se ha utilizado para implicar la atención deteriorada en grupos clínicos, por ejemplo, en pacientes con diagnóstico de TDAH (Hervey, Epstein, Curry, Tonev, Arnold, Conners, et al, 2006; King et al., 2007; López-Villalobos et al., 2010; Sjowall, Roth, Lindqvist & Thorell, 2012; Rubiales, Bakker & Urquijo, 2013). Así como estudios de tiempos de respuesta en pacientes con esquizofrenia y episodios psicóticos (García Valls et al., 2007; Giménez-Donoso et al., 2018), o enfermedad de Parkinson (Sisco, Slonena, Okun, Bowers & Price, 2016), daño cerebral (Ben-David, Nguyen & Van Lieshout, 2011), entre otros.

Distintos estudios han observado que existe una curva de crecimiento en la habilidad lectora que se inicia con la escolarización y se incrementa de manera importante durante los primeros años, se estabiliza en la adolescencia y puede sufrir un enlentecimiento en los adultos mayores (Protopapas, Archonti & Skaloumbakas, 2007, Ziaka, Moirou, Vlahou & Protopapas, 2014; Lorentz, Mckibben, Ekstrand, Gould, Anton & Borowsky, 2016).

En personas que no presentan dificultades en el proceso lector (de cualquier origen), una media estimada para la lectura de palabras se sitúa en 400 ms aproximadamente (Stroop, 1935; Comalli, Wapner & Werner, 1962; Jensen, 1965; MacLeod, 1986; Houx, Jollwa & Vreeling, 1993). Aunque en niños las mediciones pueden ascender a 600 ms o más (Golden, 1999; Martín et al., 2012).

Es necesario tener en cuenta que las variaciones en la lectura de nombres de colores en el Stroop pueden estar asociadas al idioma y a la extensión misma de las palabras utilizadas en el test (Roselli et al., 2002; Coderre, Filipi, Newhose & Dumas, 2008). Por ejemplo, la palabra “verde” en español tiene una articulación bisilábica, no así en inglés (*green*) o en francés (*vert*).

Al igual que con la lectura, existe un incremento progresivo en la capacidad para identificar y nombrar colores (Kahneman, 1973). Una media estimada en niños es de 700 a 800 ms y en adultos de 600 ms aproximadamente (Stroop, 1935; Comalli, Wapner & Werner, 1962; Jensen, 1965; MacLeod, 1986).

Y ante los estímulos incongruentes se ha reportado una media estimada de 1200 a 1500 ms en niños y de 900 a 1200 ms en adultos, aproximadamente. La resolución misma de esta tercera tarea puede tomarse como un indicador de propensión a la interferencia o de la capacidad para resistirla (Amato et al., 2006; López-Villalobos et al., 2010; Brugnolo et al., 2015).

En términos generales, la interferencia observada en la tarea conflictiva del Stroop se evidencia en la edad escolar (Comalli et al., 1962; Golden, 1978; Houx, Jolles & Vreeling, 1993; Fuentes et al., 2003), se incrementa hasta el segundo o tercer año de la primaria, acorde al desarrollo de la habilidad lectora, declina en la adultez hasta los 60 años aproximadamente y, posteriormente, se registra también un deterioro progresivo en el control de la interferencia, con enlentecimiento de las respuestas e incremento de errores.

Esta curva de deterioro se acentúa en especial en grupos con clínicos con diagnósticos de demencia (Campanholo et al., 2014), esclerosis múltiple (Vitkovitch, Bishop, Dancey & Richards, 2002; Denney & Lynch, 2009), esquizofrenia (Hepp, Maier, Hermle & Spitzer, 1996; García Valls, et al., 2007), enfermedad de Parkinson (Sisco et al., 2016), entre otras patologías.

### **Parámetros de Aciertos**

Según los baremos de la prueba de Golden (1999), en un adulto joven una media de respuesta esperable es de 119 ítems correctos en la tarea de lectura de palabras en 45 segundos, 79 en nominación de color y 50 en el conflicto palabra-color. Mientras que, por ejemplo, en niños de 10 años son esperables 85, 55 y 26 aciertos, respectivamente. La tabla 4 presenta una reseña del número de estímulos completados en 45 segundos para este y otros estudios referentes.

**Tabla 4**

*Reporte de respuestas correctas obtenidas en las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)*

Referencia	Muestra	P	C	PC
Golden (1978)	Jóvenes y adultos (n=235)	93 a 108	69 a 80	27 a 44
Vanier (1991) <sup>a</sup>	Voluntarios de 20 a 40 años (n=20)	101	70	39
Golden (1999, 2001) <sup>a</sup>	Voluntarios de 7 a 80 años (n=254)	119	79	50
Martín et al. (2012)	Estudiantes de 6 a 12 años (n=1032)	48 a 89	40 a 61	22 a 36
Rognoni et al. (2012) <sup>a</sup>	Sujetos sanos de 18 a 49 años (n=179)	111	76	48
Rivera et al. (2015) <sup>b</sup>	Adultos sanos 18 a 77 años (n=3977)	98	69	42
Rodríguez, Pulido y Pineda (2016) <sup>a</sup>	Sujetos sanos de 7 a 80 años (n=1332)	103	69	41

<sup>a</sup> Promedio de respuestas o valor ubicado en el percentil 50.

<sup>b</sup> Se informan los baremos de la muestra argentina en el rango entre 18 a 22 años, para el nivel educativo de 12 años o menos.

En la versión clásica de Golden (1978, 1999), así como en las demás estandarizaciones del mismo formato, la conversión a puntuaciones *T* de los resultados de cada tarea permite una interpretación según los parámetros de normalidad, en donde esperable un valor sobre 40 y son significativas las diferencias entre tareas de más de 10 puntos.

Asimismo, Golden (1978, 1999) expone ciertas condiciones clínicas que pueden estar asociadas al registro de puntuaciones bajas en las tareas del test. En tanto que una buena capacidad para inhibir el conflicto se traduce en puntajes altos de la tercera tarea ( $T \geq 40$ ), junto con una diferencia de puntuaciones brutas de 20% y 30% más ítems logrados en la nominación de color y la lectura, respectivamente.

La tabla 5 resume algunas observaciones adicionales sobre las tres tareas en el estudio de Golden (1978, 1999).

**Tabla 5***Observación de Golden (1978, 1999) sobre los puntajes T en el Stroop*

	P	C	PC
Dislexia	Bajo	Normal	Normal
Incapacidad para reconocer el color	Alto	Bajo	Bajo
Patología prefrontal o agitación emocional aguda	Normal	Normal	Bajo
Lesiones prefrontales y disfunción prefrontal	Normal	Normal	Muy bajo
Lesiones derechas	Normal	Bajo	Bajo
Retardo mental, lesiones en el hemisferio izquierdo y lesiones cerebrales difusas	Muy bajo	Muy bajo	Muy bajo

*Bajo  $T < 40$ ; Alto  $T > 60$ ; Rango normal o esperable en valores T entre 40 a 60.*

### Parámetros de Errores

Según Miller (2010), los errores miden un constructo diferenciable del tiempo requerido para completar las tareas del Stroop. La observación detenida de total de errores, así como la discriminación del tipo y presentación de las comisiones, tiene gran utilidad en la comprensión de los procesos de control atencional y los síntomas de impulsividad.

En este sentido, algunos autores defienden el registro del número de errores como una medida fiable en la evaluación de la función de inhibición en niños con diagnóstico de TDAH (Sørensen, Plessen, Adolfsdottir & Lundervold, 2013), por lo que se alienta a distinguir esta medición de los análisis que pueden hacerse con los tiempos de respuesta o con los aciertos.

La interferencia Stroop puede ser estudiada mediante la identificación del número y tipo de los errores registrados durante la tercera tarea del test (Busigó, Martínó, Mercado & Pérez, 1998; Zimmermann et al., 2015), bien sea el total de comisiones o se discrimine entre los errores congruentes y las imprecisiones (Etchepareborda & Comas, 1994) o también entre errores no corregidos y autocorrecciones (Zalonis et al., 2009; Miller, 2010).

La tabla 6 presenta algunos estudios en los que se incluye el número de errores en tareas tipo Stroop.



**Tabla 6**

*Reporte de errores para las tareas de lectura de palabras (P), nominación del color (C) y conflicto palabra color (PC)*

Referencia	Muestra	P	C	PC
King et al. (2007).	Adultos con TDAH y controles ( $n=44$ )		4.8 y 3.6	9.9 y 5.0
Protopapas, Archonti y Skaloumbakas (2007)	Lectores hábiles y con déficit lector de 10 a 13 años ( $n=88$ )	1.3 y 9.3	0.1 y 0.4	0.6 y 1.3
Zaloni et al. (2009)	Adultos sanos ( $n=605$ )			Err: 0.1 a 1.3 ACrr.: 0.7 a 1.7
Miller (2010)	Adultos sanos ( $n=172$ )			Err: 0.93 ACrr.: 1.39
Zimmerman et al. (2015)	Adultos de 19 a 75 años ( $n=$ )	0.17	0.26	0.64
Tremblay et al. (2016)	Adultos sanos ( $n=646$ )	0.34	0.21	1.98
Ramos-Galarza & Pérez-Salas (2017)	Niños con TDAH y controles ( $n=68$ )			3.35 y 1.65

*Err: Errores, ACrr: autocorrecciones.*

### **Métodos de Cuantificación de la Interferencia**

El cálculo de la interferencia en el Test Stroop es un poco más complejo que el simple conteo de aciertos y errores o la verificación de los tiempos de respuesta; se trata de los cálculos resultantes de todas o algunas de las mediciones básicas. Se presenta a continuación brevemente cómo se aplican algunos de los métodos de cuantificación.

#### **Índice de Diferencia ( $I_D$ )**

El cálculo de las diferencias entre las tareas de lectura y nominación de color fue anticipado Catell (1886) y luego retomado por Stroop (1935) en su estudio. En éste se propone evaluar cuánto más demora la nominación del color en un estímulo incongruente que en otro neutro, tomando para ello las latencias de respuesta.

$$I_D = PC - C$$

Asimismo, Jensen (1965) señala que pueden realizarse otras estimaciones utilizando la latencia de respuesta, por ejemplo, C-P, para conocer cuánto tiempo más se tarda en nombrar el color que en leer. En la revisión de Jensen y Rohwer (1966) se

sugiere además calcular la diferencia de puntuación entre la nominación del color y la tarea de conflicto, pero considerando el número de ítems logrados en estas tareas.

$$I_D = C - PC$$

Bien sea que se calcule por latencias o por aciertos, un resultado menor significa menos interferencia. Según Salo, Henik y Robertson (2001), la diferencia hallada entre las condiciones neutral e incongruente de los estímulos es una medida típicamente usada para evaluar la habilidad para inhibir o ignorar la información irrelevante.

En tanto que para Ben-David, Nguyen y Van Lieshout (2011), las latencias de respuesta en la tarea de conflicto reflejan una influencia conjunta del procesamiento del código de color e inhiben el código léxico rival, mientras que las latencias para la tarea neutra simplemente reflejan el procesamiento del código de color. Por tanto, calcular la diferencia entre tareas constituye una medida robusta de atención selectiva, particularmente de la interferencia.

En muestras con sujetos sanos, distintos autores reportan una diferencia de 300 a 450 ms aproximadamente entre las tareas PC y C, que indicativa de un adecuado control de interferencia; entendiéndose que cuanto más elevada sea la diferencia entre éstas, menor es la capacidad para resistir la interferencia (Stroop, 1935; Comalli, Wapner & Werner, 1962; Jensen, 1965; Houx, Jollwa & Vreeling, 1993; Golden, 1999, 2001, entre otros).

### **Índice de Razón ( $I_R$ )**

Este índice tiene antecedente en las fórmulas propuestas por Thurstone y Mellinger (1953, citado por Jensen & Rohwer, 1966), con las que propone cocientes entre las tareas de nominación de color y lectura, como en el caso del cálculo:  $C / P$ , que permiten conocer la proporción en que la habilidad para leer el listado de palabras excede a la nominación del color y viceversa si se realiza:  $P / C$ .

Así también se propone realizar el cálculo de:  $(C - P) / (C + P)$ , manipulación algebraica que muestra una función lineal entre las tareas.

Para el análisis de la relación nominación del color entre estímulos neutros y conflictivos, Lansbergen, Kenemans y Van Engeland (2007) sugieren calcular la interferencia mediante el promedio de latencias:

$$I_R = C / PC$$

O también mediante el número de aciertos:

$$I_R = PC / C$$

El resultado es igual o muy aproximado entre ambas maneras de cuantificación. Cocientes altos (>0.5) indicarían un adecuado control de la interferencia para estímulos incongruentes; coeficientes inferiores a este (que tiendan a 0.0) pueden reflejar un enlentecimiento cognitivo general, en tanto que valores muy altos (>1.0) pueden estar asociados a una discrepancia en el desempeño para nominar el color y leer, propia de patologías como la dislexia.

Como se advierte, este cálculo de  $I_R$  no tiene en cuenta las mediciones en P, sin embargo, Lansbergen, Kenemans y Van Engeland (2007) sugieren que la lectura de palabras puede usarse como covariable en el análisis de los efectos del grupo.

Distintos estudios sugieren que la estimación de  $I_R$  puede reflejar de modo más fiable la eficiencia general del procesamiento de la información (Rouder & Speckman, 2004; Lansbergen, Van Hell & Kenemans, 2007; Scarpina & Tagini, 2017). Teniendo en cuenta que se trata de una forma de medición intrasujeto, este índice se postula como un indicador de gran utilidad clínica (Lansbergen, Kenemans & Van Engeland, 2007; Sisco et al., 2016).

### **Índice de Interferencia de Golden ( $I_G$ )**

La fórmula de Golden presentada en los manuales de aplicación del test (Golden, 1978, 1999, 2001) ha sido ampliamente difundida, aunque se han realizado diversas críticas a su validez en la interpretación del Efecto Stroop (Chafetz & Matthews, 2004; Lansbergen, Kenemans y van Engeland, 2007; Sisco et al., 2016) dado el manejo artificioso del factor de corrección o valor asignado al total de aciertos en las tres tareas, en vez de usar las mediciones brutas.

El cálculo propuesto en este índice es:

$$I_G = PC - PC'$$

$$\text{donde } PC' = (P \times C) / (P + C)$$

El resultado puede ser un valor positivo o negativo (Golden, 1978, 1999, 2001). Una puntuación esperable estaría en  $I_G$  entre -10 y 10, equivalente a un puntaje  $T$  entre 40 y 60 o desviación estándar entre -1 y +1, en la población general alfabetizada.

Puntajes negativos muy bajos dan cuenta de una alta interferencia. Se señala que cuanto más alto sea el puntaje obtenido mejor es el rendimiento en las tareas de lectura y nominación de color, y menor el nivel de interferencia.

En la interpretación de los resultados de  $I_G$ , se considera que puntajes  $T$  altos son observados en personas con adecuada flexibilidad cognitiva y habilidad para responder a las demandas de las tareas del test.

En tanto que puntajes bajos en  $P$  pueden observarse también en pacientes con problemas motores del habla, o en personas con un pobre desarrollo de las habilidades lectoras atribuidas a trastornos de aprendizaje o a desventajas socioeducativas, aunque también pueden deberse a que la prueba haya sido tomada en una lengua no natural, y puntajes bajos en  $C$  pueden indicar disfunción de áreas temporo-occipitales en el hemisferio derecho (Golden, 1978, 1999).

Una revisión de  $I_G$  plantea el uso de regresión polinomial a los datos brutos obtenidos (Chafetz & Matthews, 2004), considerando los constructos de maduración y lateralización cerebral. Su cálculo utiliza un valor constante de 216 (equivalente a 5 *DE* sobre la media):

$$I_G = PC - [((216 - P) \times C) / ((216 - P) + C)]$$

### **Índices de Tempo Personal ( $I_{\text{READ}}$ e $I_{\text{NAME}}$ )**

Comalli, Wapner y Werner (1962) consideran que el rendimiento en la tarea de conflicto refleja una capacidad general para mantener un curso de acción frente a la intrusión de otros estímulos. De modo que una persona con un estilo altamente verbal puede mostrar un mejor performance en la primera tarea, y por tanto una puntuación relativamente alta al calcular:  $(C-P) / P$ . Por otro lado, “la fuerza del ego” que se refiere al control y la regulación de la respuesta, se refleja en la relación:  $(PC + C) / C$ .

Retomando este antecedente, Ludwig et al. (2010) señalan que es importante tener en cuenta el “tempo personal” o los rasgos individuales en la ejecución de las tareas, por lo que proponen dos métodos de cuantificación de la interferencia basados en observar la predominancia de la capacidad de lectura:

$$I_{\text{READ}} = [(PC - P) / P]$$

O la predominancia de la capacidad de nominación del color:

$$I_{\text{NAME}} = [(PC - C) / C]$$

En ambos casos utiliza mediciones de tiempo. En el primero de estos índices son esperables índices altos ( $>1.0$ ), que puede interpretarse como la proporción que excede el control de interferencia por efecto de la capacidad lectora.

Cuanto mayor sea  $I_{\text{READ}}$ , mayor será la interferencia de la lectura. Así, un índice de 1.40, por ejemplo, indica que es necesario un 140% más de esfuerzo atencional para evitar decir lo que se lee y sólo nombrar el color, según la observación del tiempo personal para estas tareas.

En  $I_{\text{NAME}}$  son esperables índices bajos ( $<1.0$ ), que se entienden como la proporción de esfuerzo adicional requerido para nombrar el color en el estímulo conflictivo respecto del estímulo neutro. Por ejemplo, para un índice de .75, por ejemplo, se dice que el tiempo personal para la nominación del color representa un 75% más.

En este caso, cuanto más bajo sea  $I_{\text{NAME}}$ , menor será la interferencia del color. Sin embargo, los resultados de signo negativo no deben considerarse válidos para la medición de la interferencia.

### Otros Índices de Interferencia

Como alternativa, distintos autores señalan la necesidad de incluir mediciones directas de las tres tareas del test (Caffara et al., 2002; Valentijn et al., 2005; Van der Elst, Van Boxtel, Van Breukelen & Jolles, 2006; Van Mourik et al., 2009).

La propuesta de un índice variable ( $I_v$ ) permite reconocer una relación del rendimiento observado entre mediciones, que bien puede aplicarse al cálculo por las latencias de aciertos o también al número de errores registrado en las tareas, de modo que

cuanto más alto es el resultado del índice, mayor es la interferencia producida por el conflicto.

$$I_V = PC - [(P + C) / 2]$$

Sisco et al. (2016) revisan algunas cuantificaciones que contrastan con resultados de  $I_G$  e  $I_R$ , entre ellas el Índice de Interferencia Relativo ( $I_{REL}$ ), citado en los estudios de Vitkovitch et al. (2002) y Macniven et al. (2008) y el Índice Residual ( $I_{RES}$ ), descrito en Capitani et al. (1999) y Denney et al. (2009).

$I_{REL}$  involucra mediciones de tiempos totales de respuesta en las tareas del test, indistintamente de si son latencias de aciertos o errores. El resultado se interpreta sencillamente como el porcentaje de interferencia, cuanto más alto menor es la capacidad para resolver el estímulo incongruente. Se obtiene mediante la operatoria de las latencias:

$$I_{REL} = 100 \times [(PC - C) / C]$$

También puede calcularse por número de aciertos, así:

$$I_{REL} = 100 \times [(C - PC) / C]$$

Dichos índices pueden ser calculados manualmente, en tanto que  $I_{RES}$  requiere el análisis de regresión de PC sobre C, por lo que su cuantificación suele hacerse mediante el uso de software estadístico, para determinar la varianza del desempeño en PC no asociada a C, como medida “pura” de la interferencia. Por ejemplo, Sørensen et al. (2013) sugieren que el residuo estándar corresponde a la puntuación de interferencia, considerando la condición PC como variable y la condición C como predictor en la regresión lineal.

Recientemente, Scarpina y Tagini (2017) revisan la adecuación teórica de distintos métodos para calcular el efecto de interferencia, en 44 estudios que aportan datos normativos italianos de distintas versiones del Stroop.

Su análisis incluye los cálculos de  $I_D$ ,  $I_R$ ,  $I_G$ ,  $I_V$  e  $I_{REL}$ , discriminando el tipo de medición utilizada en cada caso, al tiempo que propone examinar y retomar el uso de un índice compuesto ( $I_{COMP}$ ), que tiene la ventaja de incluir mediciones de tiempo y número de errores.

En la propuesta de Stroop (1935), la habilidad para inhibir la interferencia cognitiva puede medirse usando los resultados de la tercera tarea, que incluyen la latencia total de respuesta para completar 100 ítems y la media de latencia por ítem, así como el número de errores no corregidos, mediante la fórmula:

$$I_{\text{COMP}} = \text{Tiempo total} + [(2 \times \text{media de tiempo}) \times \text{Errores}]$$

La variante de este cálculo en Gardner, et al. (1959) es:

$$I_{\text{COMP}} = \text{Total del tiempo} + [(\text{Total de tiempo} / 100) \times \text{Errores}]$$

Finalmente, para Scarpina y Tagini (2017), la heterogeneidad en los métodos de cuantificación de la interferencia obliga un análisis minucioso sobre cuál ha de ser la mejor alternativa de interpretación de los resultados del Stroop.

## PARTE II: METODOLOGÍA



### Planteamiento del Problema

Como se menciona en los apartados anteriores, la utilidad y vigencia del Test de Colores y Palabras Stroop (Stroop, 1935; Golden, 1978), así como del estudio del fenómeno de la interferencia o Efecto Stroop (Stroop, 1935, 1938; Jensen, 1965; MacLeod, 1991), se sustenta en una extensa producción científica, particularmente en el campo de la investigación neurocognitiva (MacLeod & MacDonald, 2000; Salgado-Pineda et al., 2003; Wühr & Waszak, 2003; Herd, Banich & O'Reilly, 2006; Bugg, Delosh, Davalos & Davis, 2007; Buyukaksoy, Sengor, Gurvit, & Guzelis, 2007; Gao, Chen & Russell, 2007; King, Colla, Brass, Heuser & von Cramon, 2007; Bugg, Jacoby & Toth, 2008; Christiansen & Oades, 2010; Ben-David, Nguyen & van Lieshout, 2011; Meier & Kane, 2013; Levin & Tzelgov, 2014; Gutiérrez-Martínez, Ramos-Ortega & Vila-Chaves, 2018). A pesar de tal vastedad, persisten diferentes interrogantes de orden teórico, psicométrico y aplicado.

Distintas revisiones y metaanálisis concuerdan al señalar que no existe una forma estándar o dominante de la prueba, respecto de los materiales utilizados, la administración o la puntuación y las interpretaciones de sus resultados (Jensen, 1966; MacLeod, 1991, 2005; Homark & Riccio, 2004; Lansbergen, Kenemans & Engeland, 2007; Martín, Hernández, Rodríguez, García, Díaz & Jiménez, 2012; Meier & Kane, 2013; Scarpina & Tagini, 2017).

Aun así, una gran proporción de publicaciones hacen mención del test como herramienta en el estudio de las funciones atencionales (Lazarux, Ludwig & Aberson, 1984; MacLeod, 1992; Bench et al., 1993; Banich et al., 2000; MacLeod & MacDonald, 2000; Mead et al., 2002; Wühr, 2003; Van Mourik et al., 2009; Zarghi et al., 2012; Morrow, 2013) y de las funciones ejecutivas (Miyake, Friedman, Emerson, Witzki & Howerter, 2000; Swick & Jovanovic, 2002; Long & Prat, 2002; Soprano, 2003; Tirapu-Ustárroz, Muñoz-Céspedes, Pelegrín-Valero & Albéniz-Ferreras, 2005; Gao, Chen & Russell, 2007; Ramírez-Benítez & Díaz-Bringas, 2011; Meier & Kane, 2013; Gutiérrez-Martínez, Ramos-Ortega & Vila-Chaves, 2018) que, junto a otras pruebas, aporta información en la evaluación de la memoria de trabajo y la flexibilidad cognitiva.

Se usan adaptaciones del Stroop en el estudio de los procesos de discriminación, selección, focalización y mantenimiento activo de la atención en la resolución de

estímulos incongruentes presentados simultáneamente (Cohen, Dunbar & McClelland, 1988; MacLeod, 1991a; Besner, Stolz & Boutiuer, 1997; Durgin, 2000; Stafford & Gurney, 2007; Levin & Tzelgov, 2014; Scarpina & Tagini, 2017), la evaluación de la velocidad de respuesta (West & Alain, 2000; Repovs, 2004; Hervey et al., 2006; Bugg, DeLosch, Davalos & Davis, 2007; Rognoni et al., 2013; Leao et al., 2014), la interferencia verbal generada por los automatismos de la lectura (Cohen, Dunbar & McClelland, 1990; Besner, Stolz & Boutiuer, 1997; Ziaka, Moirou, Vlahou & Protopapas, 2014) y los modelos de procesamiento para respuestas competitivas (entre otras, Stafford & Gurney, 2007; Kaplan, Sengor, Gurvit & Guzelis, 2007; Van Maanen, Van Rijn & Borst, 2009; Lamers, Roelofs & Rabeling-Keus, 2010).

Adicional a ello, su estudio ha sido útil en la definición de los mecanismos de control inhibitorio, que inciden en aspectos básicos de la cognición (Stafford & Gurney, 2007; Smoleń & Chuderski, 2010; Meier & Kane, 2013; Gutiérrez- Martínez, Ramos-Ortega y Vila-Chaves, 2018).

Hallazgos empíricos señalan que el control de interferencia no está asociado, ni depende, de la raza, el idioma o el contexto cultural (MacLeod, 1991; Mead et al., 2002; Lovett, 2005; Ziaka, Moirou, Vlahou & Protopapas, 2014; Rivera et al., 2015), aunque si se observan cambios significativos en relación con la edad y el nivel de escolarización (Schiller, 1966; Davidson, Zacks & Williams, 2003; León-Carrión, García-Orza & Pérez-Santamaría, 2004; Van der Elst, Van Boxtel, Van Breukelen & Jolles, 2006; Christiansen & Oades, 2010; Ludwig, Borella, Tettamanti & Ribaupierre, 2010; Heine, et al., 2010; Tremblay et al., 2016; Oliveira, Mograbi, Gabrig & Charchat-Fichman, 2016).

Asimismo se ha encontrado un efecto variable en relación con el sexo (MacLeod, 1991a; Van der Elst et al., 2006; Zarghi et al., 2012) y una significativa afectación en el control de la interferencia en casos de lesión cerebral y en ciertas psicopatologías (Golden, 1976; Salgado Pineda et al., 2003; Henik & Salo, 2004; Van Mourik, Oosterlaan & Sergeant, 2005; García Valls, et al., 2007; Lansberben, Kenemans & van Engelad, 2007; López-Villalobos et al., 2010; Ben-David, Nguyen & van Lieshout, 2011; Morrow, 2013; Max et al., 2013; Sánchez-López et al., 2015).

Múltiples investigaciones corroboran el Efecto Stroop o demora en la respuesta ante una tarea conflictiva, de allí que se señale la fiabilidad del fenómeno, esto es, como

hecho neuropsicológico observable. Sin embargo, existen importantes diferencias en el tratamiento de los datos y mediciones que arroja el Stroop: bien si se trata de puntuaciones dependientes de la velocidad o los tiempos de respuesta, o bien si se refieren al análisis de las medidas de precisión en la prueba, aciertos y errores, especialmente.

Debe entenderse además que las cuestiones respecto de la psicometría del test Stroop se originan no en el Efecto Stroop en sí mismo, sino en la multiplicidad de versiones utilizadas, los parámetros de aplicación y la interpretación de sus mediciones.

Para la interpretación de los datos arrojados por el test Stroop es posible seguir los criterios de administración y de puntuación, considerando los baremos por edad, que aparecen tanto en los manuales de mayor divulgación en América Latina del Test Colores y Palabras Stroop (Golden, 1999, 2001, 2007), como en las distintas adaptaciones (Vanier, 1991; Busigó, Martinó, Mercado, Colón & Pérez, 1998; Fernández & Bendersky, 2004; Rognoni et al., 2012; Martín et al., 2012; Kang et al., 2013; Maureira, Aravena, Gálvez & Flores, 2014; Rivera et al., 2015; Sánchez-López et al., 2015; Tremblay et al., 2016; Brugnolo et al., 2016; Rodríguez Barreto, Pulido & Pineda Roa, 2016; Rivera et al., 2017; Gutiérrez-Martínez, Ramos-Ortega & Vila-Chaves, 2018).

Sin embargo, debe insistirse en la necesidad de actualización de los aspectos psicométricos del Stroop, particularmente al analizar la validez de su adaptación a formas computarizadas ya que, como señalan Salo, Henik y Robertson (2001), distintas presentaciones de la tarea tipo Stroop pueden producir diferentes niveles de interferencia.

Y, como concluyen Lansberben, Kenemans y van Engeland (2007), adicional a esto, debe tenerse en cuenta que métodos existentes para cuantificar la interferencia pueden afectar la interpretación de los datos reales.

La comprensión de los resultados, la implementación de alguno de los métodos de cuantificación de la interferencia, así como el problema de la equivalencia de este tipo de estimaciones entre formatos de aplicación del test, siguen siendo cuestiones empíricas en debate (Dunbar & MacLeod, 1984; Van Mourik, Oosterlaan & Sergeant, 2005; Larsen, Mercer y Balota, 2006; Van der Elst, et al, 2006; Ben-David & Schneider, 2009; Ludwig, Borella, Tettamanti & Ribaupierre, 2010; Penner, Kobel, Stöcklin, Weber, Opwi, & Calabrese, 2012; Levin & Tzelgov, 2014; Scarpina & Tagini, 2017), al considerar que los

grados de fiabilidad y validez de un test conciernen a propiedades de las interpretaciones que puedan surgir del mismo, como a los usos específicos de sus medidas (Prieto & Delgado, 2010).

Vistos estos desarrollos y los retos que plantea la existencia de múltiples versiones y adaptaciones del Stroop, la presente investigación se plantea el análisis comparativo de dos formas de aplicación de la prueba: una versión en español de respuesta verbal, de presentación similar al formato tradicional de Stroop (1935) y una versión computarizada, adaptada por Etchepareborda y Comas (1994), de respuesta manual o ejecución motora ante estímulos presentados en una pantalla, para las tareas de lectura, nominación de color y conflicto palabra-color.

Son escasos los estudios que abordan los efectos de la presentación de los ítems en bloque en una lámina de papel *versus* los ítems presentados en forma serial en una pantalla, así como de la respuesta, sea verbal o motora, teniendo en cuenta la existencia de un desarrollo específico de los mecanismos neurocognitivos asociados a cada uno (véanse revisiones de MacLeod, 1991a; Homack & Riccio, 2004; Penner et al., 2012; Zarghi et al., 2012; Augustinova, Parris & Ferrand, 2019).

Por otro lado, es importante tener en cuenta factores como la edad o el nivel de escolaridad, que igualmente pueden marcar diferencias en cuanto a la habilidad lectora y la velocidad de respuesta ante estímulos visuales en cada versión (Posner & Rothbart, 2000; Davidson, Zacks & Williams, 2003; Davidson, Amso, Anderson & Diamond, 2006; Best, Miller & Jones, 2009; Heine, Tamm, De Smedt, Schneider, et al., 2010; Penner et al., 2012; Rognoni et al., 2012), así como contrastes entre grupos clínicos y sujetos de desarrollo normotípico (Van Mourik, Oosterlaan & Sergeant, 2005; Christiansen & Oades, 2010; Duchek et al., 2013; Rodríguez Barreto, Pulido & Pineda Roa, 2016).

Resulta imprescindible una rigurosa discusión sobre los métodos de cuantificación de la interferencia para determinar cuál de ellos resulta más apropiado o pertinente en el uso clínico y la investigación, o si pueden ser utilizados indistintamente (Jensen, 1966; Golden, 1978; MacLeod, 1992; Salo, Henik & Robertson, 2001; Lansbergen, Kenemans & Van Engeland, 2007; Scarpina & Tagini, 2017).

### **Preguntas de Investigación**

En primer lugar, debe atenderse a interrogantes sobre el desempeño, tales como: ¿cuáles son los resultados en cada una de las versiones a utilizar?, ¿se trata verdaderamente de formas equivalentes del Stroop? y, por tanto, ¿pueden interpretarse los resultados de ambas versiones del mismo modo?

Asimismo, ¿existen discrepancias intrasujeto e intragrupo para las mediciones de aciertos, de errores o de latencias de respuesta?, ¿cuál es el efecto de factores tales como el sexo, la edad o la condición clínica de los participantes en la resolución de ambas pruebas?

De hallarse interacciones significativas en relación a los factores ¿dónde se encuentra mayor variabilidad? e igualmente, respecto de los resultados en cada versión, ¿existe asociación o correlación significativa entre las distintas mediciones básicas consideradas?

Finalmente, son importantes también las preguntas sobre la cuantificación del Efecto Stroop ¿qué método puede medir mejor la capacidad de control de interferencia?, esto es, ¿cuál de ellos presenta mayor sensibilidad y especificidad? y ¿existe concordancia entre las versiones tradicional y computarizada para un mismo método de cuantificación?

### Objetivos

El objetivo general de la investigación es comparar dos protocolos de aplicación del Test de Colores y Palabras Stroop, versiones tradicional y computarizada, en niños y adolescentes con desarrollo típico, con déficits atencionales y trastornos de aprendizaje.

Los objetivos específicos son:

- Describir el desempeño de los participantes en cada una de las versiones del Stroop para las mediciones de latencias de respuesta.
- Describir el desempeño de los participantes en cada una de las versiones del Stroop para las mediciones de número de aciertos.
- Describir el desempeño de los participantes en cada una de las versiones del Stroop para las mediciones del número de errores.
- Identificar el efecto de los factores sexo, rango etario y condición clínica en el desempeño de cada versión
- Analizar los métodos de cuantificación de la interferencia en cada versión.
- Examinar la equivalencia entre ambas versiones.

### Hipótesis

- H<sub>1</sub>: Existe una mayor demora en la respuesta en la versión computarizada que en la versión tradicional del Test Stroop.
- H<sub>2</sub>: El número de respuestas correctas es mayor en la versión tradicional que en la versión computarizada del Test Stroop.
- H<sub>3</sub>: El número de errores es mayor en la versión computarizada que en la versión tradicional del Test Stroop.
- H<sub>4</sub>: Cada una de las tareas se correlaciona positivamente entre ambas versiones del test.
- H<sub>5</sub>: El desempeño en la versión tradicional es equivalente al desempeño en la versión computarizada.
- H<sub>6</sub>: El rendimiento en ambas versiones del Test Stroop no varía en función del sexo de los participantes.
- H<sub>7</sub>: El rendimiento en ambas versiones del Test Stroop cambia en función del rango etario de los participantes.
- H<sub>8</sub>: Los grupos clínicos presentan un rendimiento disminuido en ambas versiones del Test Stroop, en comparación con el grupo de desarrollo normotípico.
- H<sub>9</sub>: Cada método de cuantificación de la interferencia Stroop se correlaciona positivamente entre ambas versiones del test.
- H<sub>10</sub>: Cada método de cuantificación de la interferencia Stroop muestra diferencias significativas por condición clínica.

### **Tipo y Diseño de Investigación**

Se propone un estudio de tipo descriptivo-comparativo (Montero & León, 2007; Hernández Sampieri, et al., 2014), por medio del cual se efectúa la descripción para cada una de las tareas del Stroop en ambas versiones, comparadas y correlacionadas sus mediciones (aciertos, errores, latencias de respuesta y cuantificaciones de la interferencia), frente a determinados factores (sexo, rango etario y condición clínica).

El enfoque y diseño de la investigación corresponde a un estudio cuantitativo, no experimental y transversal. El enfoque es eminentemente cuantitativo por cuanto se toman estas distintas mediciones para cada versión, en su mayoría variables cuantitativas continuas. La muestra fue organizada de manera equitativa por grupos según la edad y la condición clínica, se aplican ambas versiones del Stroop a todos los participantes, aunque propiamente no hay manipulación experimental de las variables. Se dice que es un estudio transversal en tanto que se tiene en cuenta el desempeño de los participantes en un periodo específico.

### **Determinación de la Muestra**

El tipo de muestreo es no probabilístico, por conveniencia (Montero & León, 1993; Hernández Sampieri, et al., 2014), considerando los criterios de inclusión y exclusión como se especifican adelante.

Todos los participantes fueron atendidos en el Laboratorio para el Estudio de las Funciones Cerebrales Superiores (LAFUN), institución privada que realiza evaluaciones neurocognitivas en sus consultorios de la Ciudad Autónoma de Buenos Aires, y quien otorga la autorización correspondiente para el uso y resguardo de los datos consignados las historias clínicas, hojas registro de las pruebas y demás documentos que forman parte de los protocolos de evaluación neurocognitiva (ver anexo 2).

El proceso diagnóstico es riguroso y tiene en cuenta la valoración interdisciplinar (neurología pediátrica, fonoaudiología, neuropsicología, entre otros), así como el reporte de los síntomas en distintos contextos mediante escalas y listas de chequeo respondidas por padres (Conners, CBCL, Achenbach) e informes escolares de rendimiento y conducta.



Se realiza inicialmente una entrevista familiar para identificar antecedentes clínicos relevantes y se revisan los resultados de instrumentos de exploración neurofisiológica (resonancia magnética y electroencefalograma cuantificado)

Según sea requerido, se realiza la aplicación de una batería de pruebas psicométricas para evaluar Coeficiente Intelectual o CI (Prüfsystem für Schul-und Bildungsberatung, Prueba K-ABC y/o versión abreviada de las escalas de inteligencia Wechsler), lenguaje y memoria (Test de vocabulario Peabody-Dunn, Tareas de fluidez verbal, Span de memoria, Figura compleja de Rey), capacidad lectora (Prueba Prolec), estado de ánimo (Escala Birlson), funciones atencionales (Test de tiempo de reacción, Test de Atención Selectiva, Test de Atención continua y Test Stroop) y funciones ejecutivas (Tarea de Planificación, Test MFF-20, Test Stroop y Test de Cartas de Wisconsin), entre otras.

Es necesario mencionar que para el diagnóstico diferencial de TDAH fueron considerados los criterios del DSM-IVR (APA, 2002), que propone tres subtipos clínicos: de predominio hiperactivo-impulsivo (-HI), déficit de atención dominante (-DA) y de síntomas combinados (-C).

La elección de la muestra clínica de los TDAH se hizo sólo para los subtipos TDAH-DA y TDAH-C, debido a la baja prevalencia del subtipo TDAH-HI, definido como precursor del subtipo combinado (Barkley, 2012; APA, 2002, 2014). En tanto que la selección de los participantes con TAPZ incluyó indistintamente pacientes con Dislexia, Trastorno de escritura y Trastorno de aprendizaje no especificado (APA, 2002).

La muestra control se obtiene mediante convocatoria personal y por Internet. Inicialmente la familia tiene una entrevista de selección en la que se chequean antecedentes de desarrollo, se verifican criterios de inclusión y exclusión, se explican los propósitos de la investigación, se solicita su consentimiento y se acuerda la entrega posterior de un informe individual de resultados.

### **Criterios de Inclusión y Exclusión**

En los grupos clínicos se incluyeron niños y adolescentes, de ambos sexos, con edades entre 7 y 22 años de edad, con CI en el rango promedio, con diagnóstico de TDAH y TAPZ.

Fue verificado que los participantes no estuvieran bajo ningún tratamiento con psicofármacos en el momento de la evaluación neuropsicológica y que no presentaran comorbilidad con otros psiquiátricos, lesión cerebral o algún trastorno orgánico (Epilepsia, por ejemplo), ni otras alteraciones sensoriales que puedan justificar el déficit atencional.

En el grupo control se incluyeron sujetos en el mismo rango de edades, con un desarrollo típico, es decir, de CI promedio, con un desarrollo escolar adecuado, no repetentes, que no presentaran déficits motores, auditivos, visuales o de lenguaje, ni trastornos neurológicos o psiquiátricos o que estuvieran bajo tratamiento con psicofármacos.

Los niños menores de edad participaron con la autorización escrita de los padres a quienes se les explicaron los propósitos y procedimientos de la investigación, en lo que se les aseguró la confidencialidad de la información y el anonimato de la participación (Anexo 2). En el caso de los mayores de 18 años fueron ellos quienes firmaron el consentimiento informado.

## **Participantes**

La muestra estuvo conformada por un total de 240 participantes (180 casos clínicos y 60 controles), de ambos sexos (86 mujeres -35.83%- y 154 varones -64.16%-), con una edad media de 13.48 años ( $DE=4.347$ ) y CI dentro del rango promedio ( $M=93.92$ ,  $DE=9.386$ ). El 98.33% de los participantes asistían a distintas instituciones educativas de CABA.

Los participantes fueron agrupados en cuatro grupos del mismo tamaño según la condición clínica ( $n=60$ ):

- Grupo TDAH-DA: pacientes con diagnóstico de Trastorno por déficit de atención con hiperactividad, predominio déficit atencional, que presentan comorbilidad con TAPZ, de ambos sexos (19 mujeres -31.66%- y 41 varones -68.33%-), con una edad media de 13.48 años ( $DE=4.359$ ).
- Grupo TDAH-C: pacientes con diagnóstico de TDAH, subtipo combinado, y comorbilidad con TAPZ, de ambos sexos (11 mujeres -18.33%- y 49 varones -81.66%-), con media de 13.72 años ( $DE=4.709$ ).

- Grupo TAPZ: pacientes con TAPZ, que no presentan diagnóstico de déficit atencional secundario, de ambos sexos (28 mujeres -46.66%- y 32 varones -53.33%-), con media de 13.57 años ( $DE=4.155$ ).
- Grupo CONTROL: niños y adolescentes con desarrollo normotípico, de ambos sexos (28 mujeres -46.66%- y 32 varones -53.33%-), con media de 13.17 años ( $DE=4.235$ ).

Así también, se igualó el número de participantes según la edad al momento de la evaluación, en tres rangos etarios ( $n=80$  cada grupo):

- 7 a 10 años (29 mujeres -39.25%- y 51 varones -63.75%-).
- 11 a 15 años (26 mujeres -32.50%- y 54 varones -67.50%-).
- 16 a 22 años (31 mujeres -39.75%- y 49 varones -61.25%-).

En la Tabla 7 se presenta conformación de la muestra, discriminado el sexo de los participantes, según el rango etario y grupo clínico.

**Tabla 7**

*Número y porcentaje parcial de participantes en cada grupo*

Grupo	7 a 10 años		11 a 15 años		16 a 22 años	
	Mujeres	Varones	Mujeres	Varones	Mujeres	Varones
TDAH-DA	4 1.66 %	16 6.66 %	6 2.50 %	14 5.83 %	9 3.75 %	11 4.58 %
TDAH-C	6 2.50 %	14 5.83 %	2 0.83 %	18 7.50 %	3 1.25 %	17 7.08 %
TAPZ	10 4.16 %	10 4.16 %	8 3.33 %	12 5.00 %	10 4.16 %	10 4.16 %
Controles	9 3.75 %	11 4.58 %	10 4.16 %	10 4.16 %	9 3.75 %	11 4.58 %

*Porcentaje respecto del total de la muestra ( $n=240$ ).*

Los grupos no fueron apareados por sexo, ya que se reconoce un mayor número de casos clínicos en los varones, de acuerdo con los estudios de prevalencia para trastornos de neurodesarrollo (APA, 2002; Jiménez, 2012).

## **Instrumentos**

Se describen a continuación los instrumentos utilizados en el presente estudio, así como las puntuaciones que pueden obtenerse en cada uno. Una ilustración de cada formato utilizado aparece en el anexo 1.

### **Stroop Versión Tradicional**

Se trata de una adaptación de la prueba ideada por Stroop (1935) y luego normatizada por Golden (1978). Permite evaluar la capacidad para separar estímulos neutros y estímulos incongruentes (Stroop, 1935; Golden, 1978) o para examinar el efecto de la interferencia del color sobre la habilidad lectora (Cohen, Dunbar y McClelland, 1990; Besner, Stolz & Boutilier, 1997), poniendo en juego la capacidad de selectividad (Würh & Waszak, 2003) y el control atencional (Bugg, Jacoby & Toth, 2008).

La versión utilizada contiene tres láminas y en cada una se presentan 72 ítems dispuestos en tres columnas de modo pseudoaleatorio (Ver Anexo 1.1). Se pide una respuesta verbal, en voz alta en cada lámina, en un tiempo límite de 45 segundos para cada una.

En la primera lámina (P) se presentan los nombres de cuatro colores impresos en tinta negra (rojo, verde, azul, negro) y se pide al participante leer en voz alta este listado de palabras lo más rápido que pueda.

La segunda lámina (C) presenta 72 rectángulos pintados alternadamente con los mismos cuatro colores y se pide nombrar en voz alta el color que ve en cada ítem lo más rápido posible.

En la tercera lámina (PC) se presenta un listado de 72 palabras, que nombran uno de estos mismos cuatro colores, pero impresas en tinta de un color que no se corresponde con el significado. Se pide nombrar en voz alta el color con que están pintadas las letras y no leer la palabra escrita.

Con la primera lámina se explora la atención focalizada frente a una tarea sencilla: leer, en tanto que la segunda requiere una atención más selectiva mediante la tarea de denominar colores y, la tercera, requiere de un control de la atención más continuo y elevado (Golden, 1999; Conca Binfa & Ibarra González, 2004).

Para su aplicación se utilizan las consignas y registros descritos en el manual en español de la prueba de Golden (1999). Agregando la indicación de que si llega al último elemento del listado debe retomar desde el primero y continuar hasta que se le indique que ha terminado el tiempo.

En cada tarea se da un tiempo límite de 45 segundos. A medida que avanza se marcan los errores y al terminar se cuenta el total de respuestas dadas, restando los errores para obtener los aciertos. La corrección de la puntuación se realizó únicamente para poder calcular  $I_G$  (Golden, 1999).

### ***Operacionalización de variables***

Las puntuaciones brutas tomadas en cuenta corresponden a los registros de aciertos y errores, con un límite de 45 segundos otorgados para cada una de las tareas (P, C y PC).

A partir del total de respuestas dadas en cada tarea (Aciertos + Errores) pudo obtenerse una estimación de los tiempos promedio de respuesta en cada tarea (expresada en milisegundos).

Las mediciones del desempeño en la versión tradicional del Stroop son:

- *Latencia estimada de respuesta.*  $45000 / (\text{Aciertos} + \text{Errores})$
- *Aciertos.* Número total de respuestas correctas dadas.
- *Errores.* Número total de errores cometidos.

Los cálculos de interferencia consideran distintos índices según las anteriores puntuaciones. Estos son:

- $I_D = C - PC$ . Utiliza la puntuación bruta de aciertos.
- $I_R = PC / C$ . Utiliza la puntuación bruta de aciertos.
- $I_G = PC - ((P \times C) / (P + C))$ . A la puntuación bruta de aciertos se suma el factor de corrección por edad sugerido en Golden (1999).
- $I_{REL} = ((C - PC) / C) \times 100$ . Utiliza la puntuación bruta de aciertos.
- $I_V = PC - ((P + C) / 2)$ . Utiliza la puntuación bruta de errores.

### **Stroop Versión Computarizada**

La versión computarizada del Stroop usada en este estudio hace parte de la batería Monitoreo Cognitivo Computarizado (MCC) diseñada por Etchepareborda y Comas (1994). El diseño de la prueba cuenta con los mismos ítems que la versión tradicional antes descrita, aunque dispuestos uno tras otro en el centro de una pantalla (Ver anexo 1.2). Se pide una respuesta motora al sujeto en la que debe presionar una de cuatro teclas para cada color: F5 (negro), F6 (rojo), F7 (verde) y F8 (azul).

En la primera parte (P) se presentan el nombre de uno de los cuatro colores en letras pintadas de negro. Se pide al participante leer y presionar la tecla correspondiente lo más rápido posible. Una vez que marca su respuesta aparece la siguiente palabra, y así sucesivamente hasta completar el total de las 72 palabras.

Seguidamente aparece una barra pintada con uno de los cuatro colores y se pide presionar la tecla correspondiente al color que se ve en pantalla. se continúa hasta completar los 72 ítems (C).

En la tercera parte (PC) aparece una palabra que nombra un color, pero con letras pintadas de otro color distinto del significado. El participante debe presionar la tecla correspondiente al color del que están pintadas las letras (nominación del color de las letras editadas) y no lo que se lee en la palabra.

En cada tarea se obtiene el número de respuestas correctas y el número de comisiones (discriminados entre errores congruentes e imprecisiones), el porcentaje de logro (de los 72 ítems mostrados), el tiempo mínimo y máximo usado, así como la latencia media por ítem. Estos valores son arrojados directamente por el software del MCC.

### ***Operacionalización de variables***

Las puntuaciones brutas tenidas en cuenta para este estudio corresponden a las latencias de aciertos, número de aciertos y número de errores, en cada una de las tareas. No fueron usados los valores de porcentaje de logro, ni valores mínimos y máximos de latencia.

Las mediciones del desempeño en la versión computarizada son:

- *Latencia media de aciertos.* Valor promedio por ítem en milisegundos.

- *Aciertos*. Número total de respuestas correctas de 72 ítems presentados.
- *Errores*. Número total de errores cometidos de 72 ítems presentados. Se describe de modo diferenciado el tipo de error en: número de imprecisiones y número de errores congruentes.

Los cálculos de interferencia consideran distintos índices según las anteriores puntuaciones. Estos son:

- $I_D = PC - C$ . Utiliza el valor de latencias medias de aciertos.
- $I_R = C / PC$ . Utiliza el valor de latencias medias de aciertos.
- $I_{READ} = (PC - P) / P$ . Utiliza el valor de latencias medias de aciertos.
- $I_{NAME} = (PC - C) / C$ . Utiliza el valor de latencias medias de aciertos.
- $I_V = PC - ((P + C) / 2)$ . Utiliza la puntuación bruta de errores.

### **Consideraciones éticas**

Dado que el desarrollo de esta investigación involucra personas, es importante mencionar que la propuesta fue revisada y aprobada por el Comité de Ética de LAFUN (ver anexo 2), teniendo en cuenta las normativas nacionales e internacionales para salvaguardar el bienestar, autonomía y anonimato de los participantes, no se presentan imágenes o datos de identificación personal.

## PARTE III: RESULTADOS



### Análisis de los Datos

La presentación de los resultados incluye las mediciones básicas de latencias de respuesta, número de aciertos y número de errores, registrados en cada una de las tareas del Stroop (P, C y PC) y en cada versión (tradicional y computarizada), seguidas de las cuantificaciones de la interferencia correspondientes ( $I_D$ ,  $I_R$ ,  $I_G$ ,  $I_{READ}$ ,  $I_{NAME}$ ,  $I_{REL}$  e  $I_V$ ).

Todos los análisis se realizan mediante el software IBM SPSS Statistics (versión 24.0), seleccionando pruebas no paramétricas, en razón a que no se cumplen cabalmente los supuestos de normalidad y de homogeneidad de las varianzas ( $p < .05$ ), salvo algunas excepciones.

Inicialmente se presenta una descripción general de las mediciones básicas en cada versión del Stroop, mencionando los resultados de la media ( $M$ ) y la desviación estándar ( $DE$ ), seguido de los contrastes generales por cada factor, es decir, comparaciones intersujeto o de muestras independientes, mediante la prueba de Mann-Whitney ( $U$ ) por sexo y mediante la prueba Kruskal-Wallis ( $X^2$ ) por rango etario y grupo diagnóstico. Se consideran los criterios de significancia estadística  $\alpha$  ( $p < .01$  y  $p < .05$ ) e intervalo de confianza de 95%.

Seguidamente, se examinan los efectos principales sobre cada medición y la interacción entre los factores, mediante el Modelo Lineal General o análisis univariante, incluyendo el análisis Post-Hoc con prueba de Tukey. Se presentan los estadísticos y su nivel de significancia ( $p < .05$ ), utilizando el tamaño de la muestra de la media armónica y, según sea relevante, se menciona la magnitud del efecto Eta cuadrado parcial para valores medios ( $\eta^2 > .06$ ) o grandes ( $\eta^2 > .14$ ).

Respecto de las cuantificaciones de la interferencia en cada tipo de medición, se describen los resultados generales de medias ( $M$ ) y desviaciones estándar ( $DE$ ), así como los valores de sensibilidad y especificidad en el análisis de las Curvas ROC, para explorar su capacidad discriminante, en el total de la muestra y diferenciando por factor. Se resaltan aquellos valores aceptables del área bajo la curva (por su sigla en inglés *Area Under Curve*:  $AUC > .60$ ).

Luego de describir el desempeño en cada versión de forma individual, se presentan los análisis comparativos entre versiones. Inicialmente se estudian las correlaciones mediante el Cociente de Pearson, ya que las mediciones básicas, así como

los cálculos de la interferencia, son presentadas como variables cuantitativas continuas que no fueron transformadas en categorías. Se mencionan únicamente resultados para grados de asociación entre las mediciones de nivel moderado ( $r > .40$ ) o superiores.

Para terminar, se realizan las comparaciones intrasujeto o para muestras relacionadas, utilizando la prueba de rangos con signos de Wilcoxon, tanto de las mediciones básicas como de los índices de interferencia, para evaluar la equivalencia en el desempeño general de los participantes acorde al formato de aplicación. Se menciona el estadístico de contraste estandarizado ( $Z$ ) y su nivel de significancia asintótica ( $p < .01$ ).

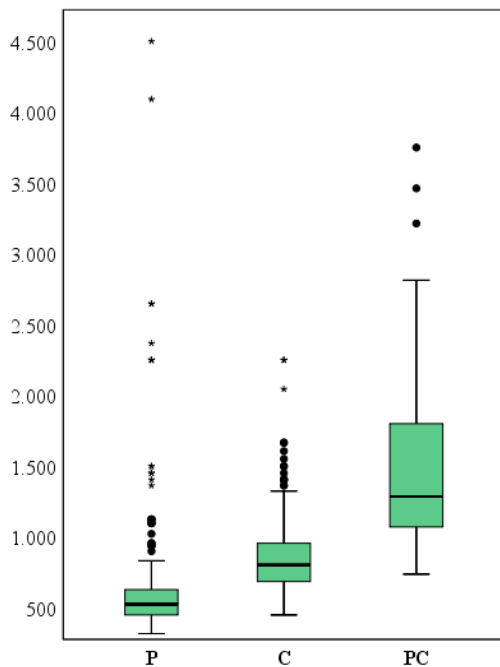
## Desempeño en el Stroop Tradicional

### Latencias de Respuesta Estimadas en el Stroop Tradicional

Como se menciona en la operacionalización de las variables, para esta versión se estima la latencia media de respuesta dividiendo el tiempo total de la tarea (45 segundos) sobre el total de las respuestas dadas en cada tarea (Aciertos + Errores). La figura 6 presenta la distribución general de estos resultados, con valores en milisegundos (ms).

**Figura 6**

*Latencias de respuesta estimadas en el Stroop tradicional (n=240)*



El total de la muestra registra valores promedio de 649.36 ms ( $DE=488.07$ ) en la tarea de lectura (P), de 863.29 ms ( $DE=277.65$ ) en la nominación de color (C) y de 1454.88 ms ( $DE=518.71$ ) en la resolución del conflicto palabra-color (PC).

El amplio rango de latencias estimadas en las tres tareas, así como los valores elevados de las desviaciones estándar y la presencia de valores atípicos y extremos, permiten corroborar la dispersión y variabilidad de las mediciones. La tendencia general indica un aumento del tiempo requerido para responder a la secuencia de tareas en el Stroop tradicional, siendo la tercera la que conlleva una mayor demora.

### Contrastes por Factor

La tabla 8 presenta los resultados de la estimación de latencias en la versión tradicional, descritas por los factores sexo, rango etario y condición clínica. En el anexo 3 pueden consultarse otros descriptivos de estos factores, así como las pruebas de ajuste correspondientes.

**Tabla 8**

*Latencias de respuesta estimadas en el Stroop tradicional y comparación por factores*

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	683.26 (607.32)	<i>ns</i>	880.29 (340.58)	<i>ns</i>	1378.19 (470.92)	<i>ns</i>
Varones	630.44 (407.66)		853.10 (232.06)		1498.40 (541.23)	
7-10 años	977.25 (736.81)		1107.60 (314.62)		1909.61 (512.32)	
11-15 años	512.00 (85.73)	**	797.61 (137.99)	**	1353.26 (382.22)	**
16-22 años	458.84 (75.39)		683.31 (118.21)		1103.09 (249.51)	
TDAH-DA	629.18 (344.16)		893.92 (315.19)		1524.73 (542.91)	
TDAH-C	621.32 (368.56)	<i>ns</i>	855.47 (225.32)	<i>ns</i>	1586.80 (615.37)	**
TAPZ	776.05 (785.99)		914.68 (307.83)		1451.93 (451.51)	
CONTROL	570.90 (262.85)		787.30 (230.22)		1257.82 (391.35)	

*Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )*

*Diferencia no significativa: ns ( $p > .05$ )*

En primer lugar, se observa que las latencias más altas aparecen en la tercera tarea y son levemente superiores en los varones de la muestra. No obstante, la comparación no arroja diferencias significativas en cuanto al tiempo estimado en P ( $U(1)=6747.50$ ,  $p=.808$ ), ni en C ( $U(1)=6938.50$ ,  $p=.539$ ), ni en PC ( $U(1)=7574.00$ ,  $p=.065$ ). Esto indica que los varones y las mujeres registran mediciones que tienden a ser similares en la versión tradicional del Stroop.

Por el contrario, en la comparación por el factor rango etario se hallan diferencias estadísticamente significativas, en la tarea P ( $X^2(2)=116.88$ ,  $p=.000$ ), en C ( $X^2(2)=123.13$ ,  $p=.000$ ) y en PC ( $X^2(2)=112.51$ ,  $p=.000$ ); registrándose una mayor demora en la respuesta en el grupo de 7 a 10 años para las tres tareas.

Mientras que en las comparaciones por grupo diagnóstico se observa que el grupo control presenta mayor velocidad media de respuesta en las tres mediciones del Stroop

tradicional, aunque las diferencias alcanzan el nivel de significancia estadística únicamente en la tercera tarea ( $X^2(3)=14.91, p=.002$ ), pero no en P ( $X^2(3)=1.38, p=.710$ ), ni en C ( $X^2(3)=6.66, p=.084$ ).

Teniendo en cuenta que en estos contrastes iniciales se hallan diferencias significativas asociadas a los factores de edad y grupo diagnóstico, pero no al sexo, de aquí en más se presentan los análisis conjuntos de éstos, mencionando el factor sexo sólo cuando sus efectos o interacciones resulten relevantes.

### **Efectos e Interacción de los Factores y Análisis Post-Hoc**

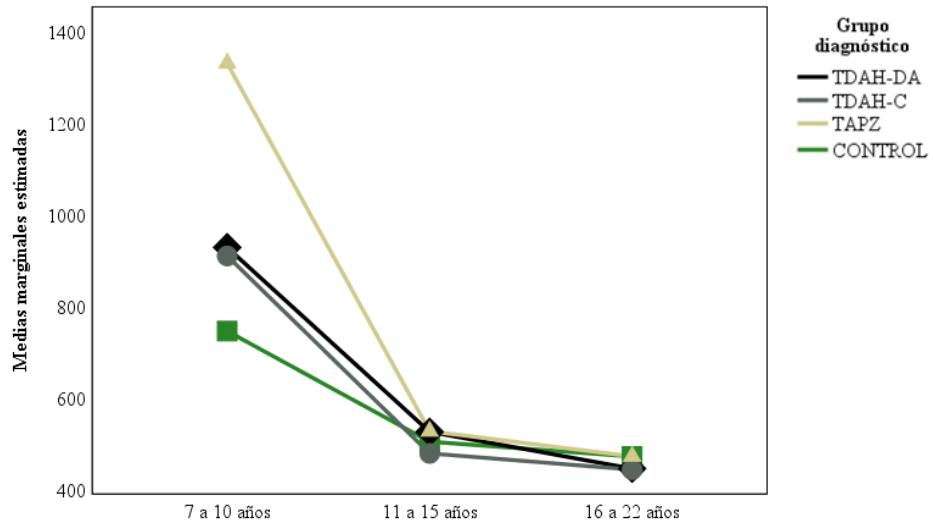
En el examen conjunto de los factores edad y grupo diagnóstico se halla el mejor desempeño en el grupo de 16 a 22 años, siendo más rápidos en promedio los participantes con diagnóstico de TDAH-C (441 ms) en la primera tarea, mientras que el grupo control lo fue en las tareas C y PC (642 y 945 ms, respectivamente).

En tanto que el peor desempeño fue registrado por los participantes de 7 a 10 años del grupo TDAH-C durante PC (2047 ms), seguidos por los registros del grupo TAPZ en P y C (1329 y 1241 ms).

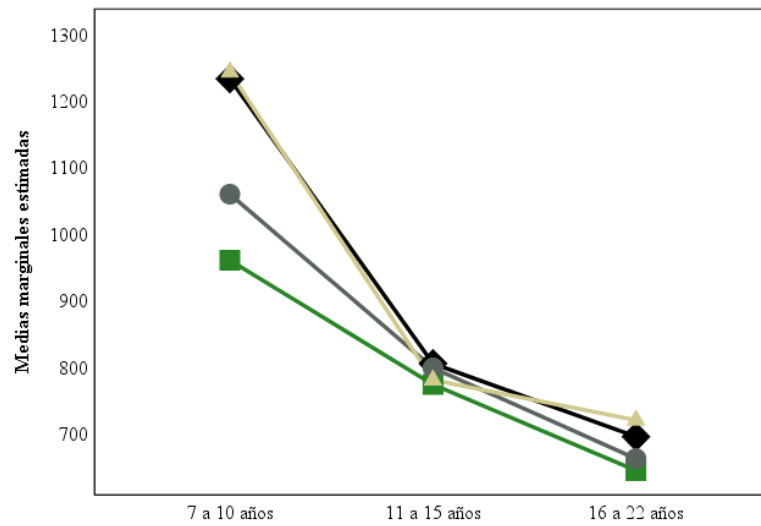
En la página siguiente, la Figura 7 presenta las medias marginales estimadas en las tres tareas del Stroop tradicional. Resultados adicionales de los análisis de varianza y las pruebas intersujeto para estas mediciones pueden consultarse en el anexo 4.

**Figura 7**

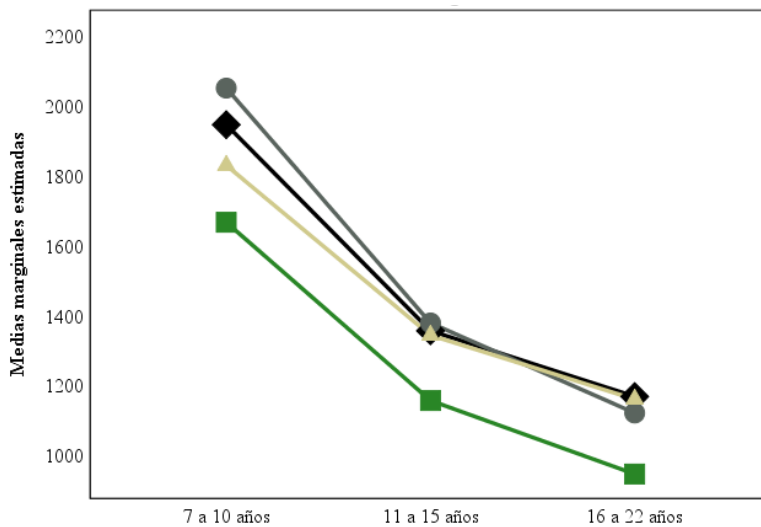
*Discriminación de las latencias de respuesta estimadas en la versión tradicional, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:

Las pruebas de efectos intersujetos para P en el Stroop tradicional muestran un efecto principal del rango etario ( $F(2)=29.280, p=.000, \eta^2=.213$ ), así como una interacción leve entre el rango etario y el grupo diagnóstico, aunque no alcanza el nivel de significación ( $F(3)=2.098, p=.055$ ).

Las pruebas Tukey evidencian la conformación de subconjuntos homogéneos para los rangos de edad 11 a 15 años y 16 a 22 años ( $p=.711$ ), que son más rápidos en general, con 458 ms y 512 ms en promedio, y difieren de los más pequeños que son más lentos, con 977 ms.

En los contrastes de la tarea C se hallan efectos principales significativos para los factores de rango etario ( $F(2)=83.491, p=.000, \eta^2=.436$ ), y de grupo diagnóstico ( $F(3)=4.596, p=.004, \eta^2=.060$ ), así como efectos de interacción entre ambos factores ( $F(6)=2.279, p=.037, \eta^2=.060$ ).

Se corroboran diferencias en el desempeño asociadas a la edad, así como la formación del subconjunto de los grupos clínicos TDAH-DA, TDAH-C y TAPZ ( $p=.389$ ), con latencias superiores a 855 ms, que distan del control con 789 ms en la segunda tarea.

Los resultados para la tarea PC muestran efectos intersujetos para los factores de rango etario ( $F(2)=72.975, p=.000, \eta^2=.403$ ) y grupo diagnóstico ( $F(3)=4.926, p=.002, \eta^2=.064$ ), sin interacción significativa entre ambos factores ( $p>.05$ ).

Las pruebas post-hoc evidencian importante disparidad por edad. De otro lado, se halla homogeneidad entre los grupos clínicos ( $p=.210$ ) con latencias superiores a 1451 ms, mientras el grupo control presenta 1256 ms.

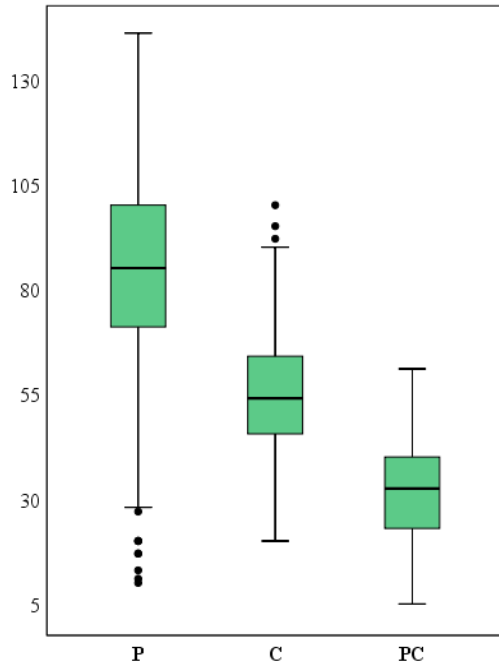
En términos generales, puede advertirse que las latencias de respuesta disminuyen progresivamente con la edad, mientras se mantiene el mismo patrón de demora, cercana al doble del tiempo requerido para la ejecución de PC, respecto de C.

### **Número de Aciertos en el Stroop Tradicional**

Como ilustra la figura 8, la distribución general de respuestas correctas obtenidas en 45 segundos es inversa a la medición de latencias antes descrita. Esto es, se aprecia una disminución progresiva del número de aciertos conforme las tareas de esta versión.

**Figura 8**

Número de aciertos en el Stroop tradicional ( $n=240$ )



En el total de los participantes se registra una media aciertos de 83.13 ( $DE=25.13$ ) en la tarea P, de 54.99 ( $DE=15.01$ ) en C y de 31.79 ( $DE=11.32$ ) en PC. La distribución general de estos resultados, especialmente en la primera y segunda tarea, indica una amplia dispersión, con presencia de valores atípicos.

Al mismo tiempo, el menor número de aciertos logrados en promedio en la tercera tarea ratifica la dificultad que representa la resolución del conflicto palabra-color en la versión tradicional del Stroop.

### Contrastes por Factor

Como sigue, la tabla 9 presenta los resultados del número de aciertos discriminados por factor: sexo, edad y condición clínica. Otros descriptivos y pruebas de ajuste para estas mediciones pueden consultarse en el anexo 5.



**Tabla 9***Número de aciertos en las tareas del Stroop tradicional y comparación por factores*

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	83.75 (26.97)	ns	54.99 (15.78)	ns	33.43 (11.19)	ns
Varones	82.79 (24.12)		54.99 (15.59)		30.86 (11.32)	
7-10 años	59.34 (22.70)	**	41.66 (10.46)	**	21.79 (7.01)	*
11-15 años	89.65 (14.37)		56.71 (10.65)		32.70 (9.03)	
16-22 años	100.40 (16.43)		66.60 (11.77)		40.86 (8.46)	
TDAH-DA	82.55 (25.52)	ns	53.90 (15.75)	*	30.67 (11.08)	*
TDAH-C	84.52 (25.33)		54.53 (15.49)		29.27 (11.51)	
TAPZ	78.67 (26.94)		51.78 (14.56)		29.57 (10.41)	
CONTROL	86.78 (22.40)		59.75 (13.28)		37.63 (10.39)	

*Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )**Diferencia no significativa: ns ( $p > .05$ )*

Vistos individualmente los factores, se encuentra que varones y mujeres registran mediciones bastante similares de aciertos en P ( $U(1)=6445.50, p=.732$ ), en C ( $U(1)=6302.50, p=.535$ ) y en PC ( $U(1)=5722.00, p=.081$ ). Entonces, no existen diferencias significativas asociadas al sexo de los participantes en las tres tareas del Stroop tradicional.

En razón a ello, al igual que se hiciera en la presentación de las latencias de respuesta estimadas, se omite mencionar resultados que involucren el factor sexo, por cuanto muestra poca asociación, interacción o efecto sobre las demás mediciones.

Por el contrario, en la comparación por el factor rango etario se hallan diferencias estadísticamente significativas tanto en la tarea P ( $X^2(2)=116.29, p=.000$ ), como en C ( $X^2(2)=117.74, p=.000$ ) y en PC ( $X^2(2)=117.68, p=.000$ ).

Puede advertirse que el número de aciertos aumenta con la edad, al tiempo que la media de aciertos en PC es baja en todos grupos de edad, en algunos casos se aproxima a la mitad de los aciertos logrados en C o incluso a una tercera parte en relación con P.

El contraste entre los grupos no alcanza un nivel de significancia estadística en P ( $X^2(3)=1.56, p=.667$ ), pero se confirman diferencias significativas en C ( $X^2(3)=8.32, p=.040$ ) y en PC ( $X^2(3)=22.55, p=.000$ ). En las comparaciones por grupo diagnóstico se

observa que el grupo con TAPZ presenta un desempeño levemente inferior en las dos primeras tareas, en cuanto al número de aciertos.

### **Efectos e Interacción de los Factores y Análisis Post-hoc**

En términos generales, el mejor desempeño en la primera tarea fue observado por el grupo con diagnóstico TDAH-DA de 16 a 22 años con una media de aciertos de 104.05, mientras que en el mismo rango etario los participantes del grupo control logran un promedio más alto de respuestas correctas en C (70.75) y en PC (47.35).

Por el contrario, los participantes de 7 a 10 años del grupo TAPZ logran un número menor de aciertos en P (51.90) y C (36.10), y los grupos TDAH-DA y TDAH-C tienen un desempeño bajo bastante similar en la tercera tarea (19.90 y 19.60 aciertos, respectivamente).

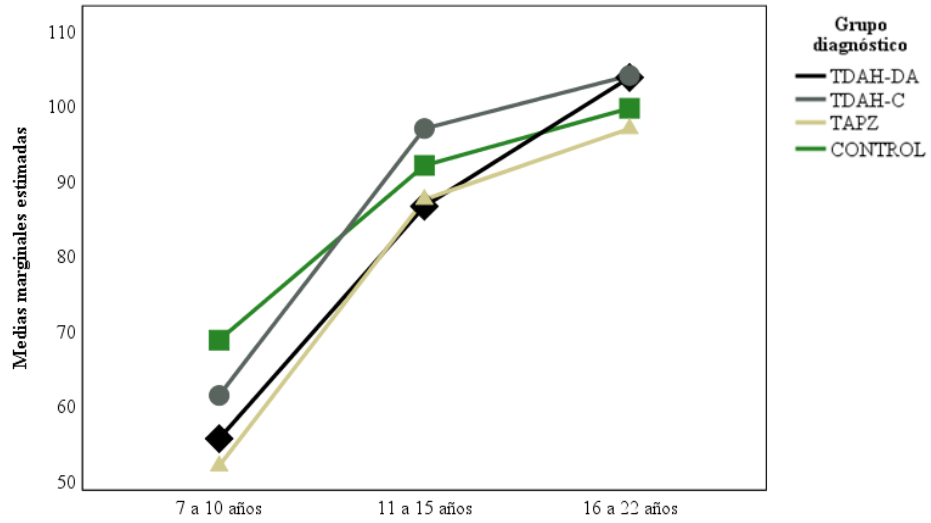
En la página siguiente, la figura 9 ilustra el examen conjunto de los factores de edad y grupo diagnóstico, considerando el registro de respuestas correctas en la versión tradicional. Otros valores obtenidos en los análisis de efectos intersujetos pueden consultarse en el anexo 6.

En el Stroop tradicional, los análisis comparativos por factor muestran un efecto principal del rango etario en la tarea P ( $F(2)=92.935, p=.000, \eta^2=.463$ ) así como una interacción leve con el grupo diagnóstico, aunque no alcanza el nivel de significación ( $F(3)=2.553, p=.057$ ).

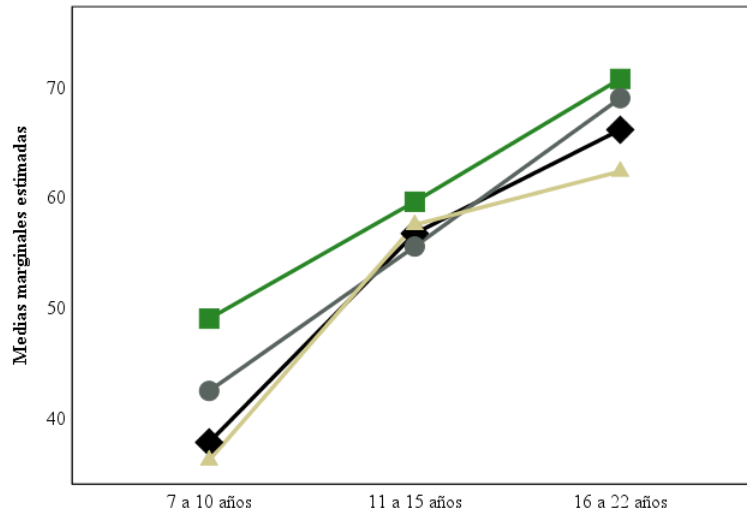
Los análisis post-hoc evidencian que cada rango etario tiene un desempeño diferenciado, con un promedio de 59.34, 89.65 y 100.41 aciertos en cada tarea, mientras que los contrastes muestran cierta homogeneidad entre los grupos clínicos y el grupo control ( $p=.072$ ), con medias entre 78.67 y 86.79 aciertos.

**Figura 9.**

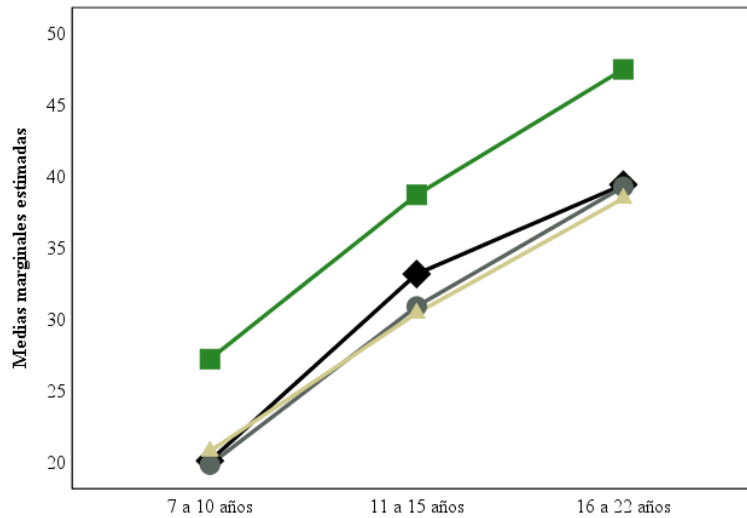
*Discriminación del número de aciertos en la versión tradicional, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:

En la tarea C se encuentran efectos principales significativos del factor rango etario ( $F(2)=95.633$ ,  $p=.000$ ,  $\eta^2=.470$ ), y del factor grupo diagnóstico ( $F(3)=5.721$ ,  $p=.001$ ,  $\eta^2=.074$ ), aunque la interacción entre ambos factores no alcanza un nivel de significancia ( $p>.05$ ).

Se corroboran marcadas diferencias asociadas a la edad de los participantes, así como la conformación del subconjunto TAPZ, TDAH-C y TDAH-DA ( $p=.580$ ), respectivamente con de 51.78, 53.90 y 54.53 aciertos, frente al grupo control con 59.73 aciertos en promedio.

En la tarea PC se hallan efectos intersujetos para los factores de rango etario ( $F(2)=103.923$ ,  $p=.000$ ,  $\eta^2=.490$ ) y de grupo diagnóstico ( $F(3)=13.569$ ,  $p=.000$ ,  $\eta^2=.159$ ), no obstante no existe interacción entre estos ( $p>.05$ ).

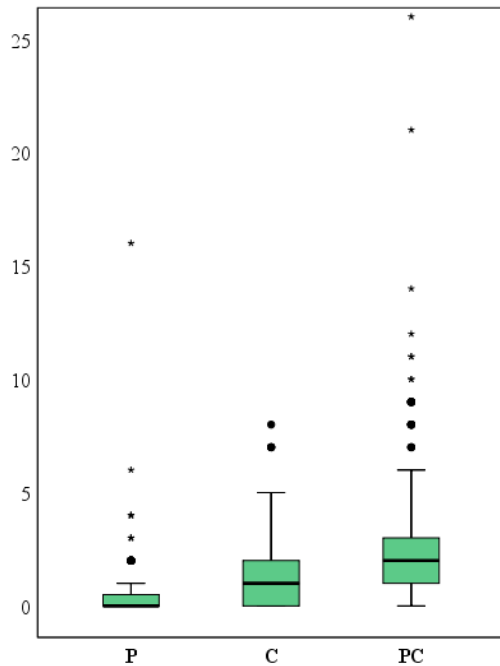
No se demuestra la conformación de subconjuntos homogéneos por el factor de rango etario, pero si por el grupo diagnóstico, donde los grupos clínicos TDAH-C, TAPZ y TDAH-DA conforman un subconjunto homogéneo ( $p=.801$ ), que registran un número menor de aciertos, inferior a 31, en comparación con el grupo control que presenta una media de 37.65 aciertos.

### **Número de Errores en el Stroop Tradicional**

En este apartado de los resultados se presenta la tendencia general en la comisión de errores en la versión tradicional, así como los contrastes por los factores, en el mismo orden que se hiciera con las mediciones anteriores.

En la página siguiente, la figura 10 ilustra la distribución general de errores registrados en 45 segundos, que se incrementan progresivamente conforme las tareas de la versión tradicional.

La mayor dispersión puede apreciarse en la medición de errores de la tercera tarea, con una media de 2.63 ( $DE=3.29$ ), que supera ampliamente las comisiones registradas en las dos primeras, registrando un promedio de 0.46 ( $DE=1.32$ ) errores en P y de 1.39 ( $DE=1.57$ ) en C. Se advierte una alta presencia de valores atípicos que puede atribuirse a una notable presencia de errores en el rango de 7 a 10 años y, particularmente, en los grupos clínicos, para las tres tareas del test.

**igura 10***Número de errores en el Stroop tradicional (n=240)*

### Contrastes por Factor

En términos generales, puede advertirse que la distribución de errores en la versión tradicional del Stroop tiende a ser similar entre varones y mujeres, aunque, como se menciona antes, varía en los contrastes por rango etario y por condición clínica.

En la página siguiente, la tabla 10 presenta dicha tendencia de distribución por cada factor. En el anexo 7 pueden consultarse otros estadísticos descriptivos de las mediciones de errores en las tres tareas del Stroop tradicional, así como las correspondientes pruebas de normalidad y homocedasticidad.

Para la muestra total de participantes ( $n=240$ ), el contraste en la primera tarea del Stroop tradicional arroja resultados similares en cuanto al número de errores entre varones y mujeres ( $U(1)=7140.50$ ,  $p=.185$ ), lo mismo ocurre en la tarea C ( $U(1)=6554.50$ ,  $p=.892$ ), así como en PC ( $U(1)=7081.00$ ,  $p=.365$ ).

**Tabla 10***Número de errores en las tareas del Stroop tradicional y comparación por factores*

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	0.36 (0.94)	ns	1.47 (1.76)	ns	2.60 (3.79)	ns
Varones	0.51 (1.48)		1.35 (1.45)		2.64 (2.98)	
7-10 años	0.54 (1.11)	ns	1.61 (1.89)	ns	3.43 (4.04)	**
11-15 años	0.55 (1.88)		1.43 (1.43)		2.86 (3.29)	
16-22 años	0.29 (0.64)		1.14 (1.29)		1.60 (1.95)	
TDAH-DA	0.73 (2.26)	ns	1.33 (1.52)	**	2.30 (2.46)	**
TDAH-C	0.35 (0.75)		1.72 (1.82)		2.82 (2.70)	
TAPZ	0.57 (0.98)		1.82 (1.66)		4.23 (4.92)	
CONTROL	0.18 (0.43)		0.70 (0.88)		1.17 (1.16)	

*Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )**Diferencia no significativa: ns ( $p > .05$ )*

En el análisis por el factor rango etario no se hallan diferencias para la tarea P ( $X^2(2)=1.74, p=.418$ ), ni para C ( $X^2(2)=2.36, p=.307$ ). En tanto que en PC si se hallan diferencias significativas ( $X^2(2)=15.42, p=.000$ ), con el mayor número de errores promedio en el grupo de 7 a 10 años.

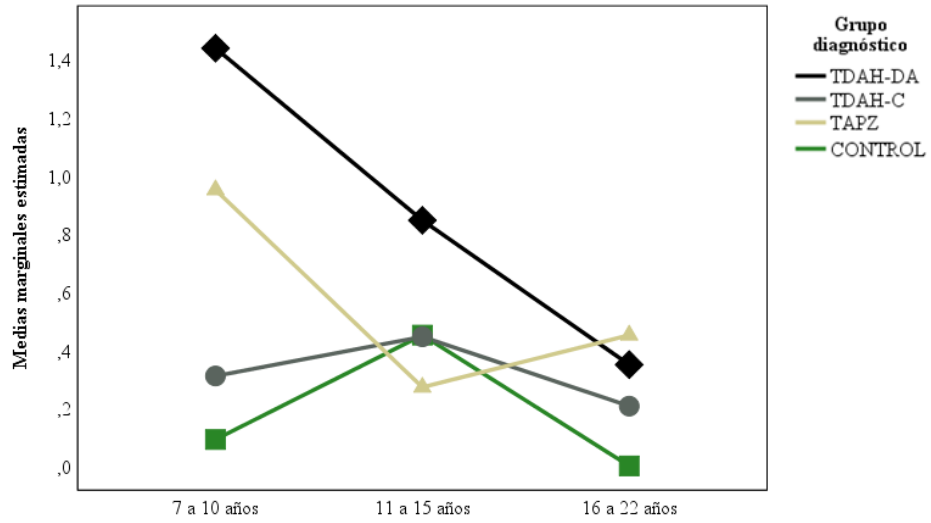
En las comparaciones por grupo diagnóstico se observa que los controles presentan un menor número de errores en las tres mediciones del Stroop tradicional. Aunque las diferencias halladas alcanzan el nivel de significancia estadística en las tareas de nominación del color, C ( $X^2(3)=18.76, p=.000$ ) y PC ( $X^2(3)=26.20, p=.000$ ), pero no en la lectura de palabras ( $X^2(3)=5.63, p=.131$ ).

### **Efectos e Interacción de los Factores y Análisis Post-hoc**

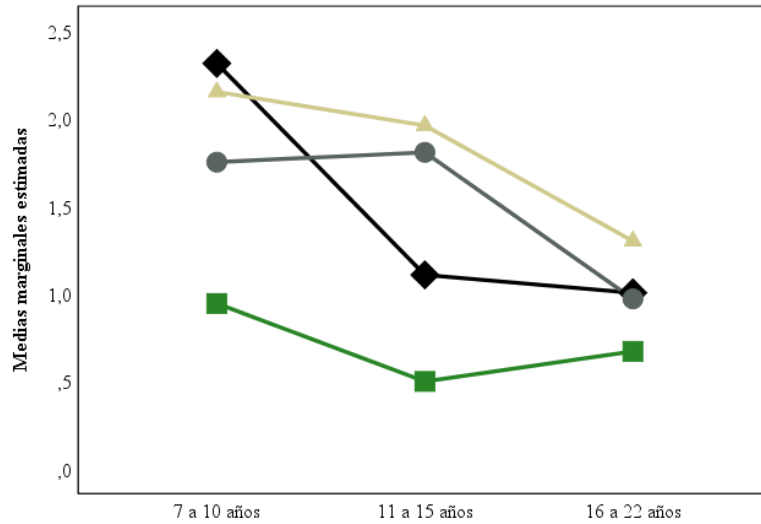
A continuación, la figura 11 presenta las medias marginales estimadas para las tareas P, C y PC en el Stroop tradicional, teniendo en cuenta la edad y la condición clínica. Información adicional puede consultarse en el anexo 8.

**Figura 11.**

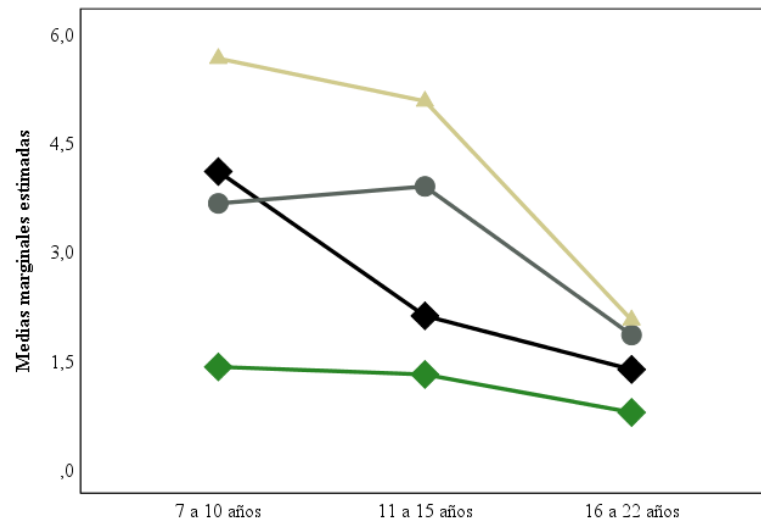
*Discriminación del número de errores en la versión tradicional, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:

Las pruebas de efectos intersujetos para P en el Stroop tradicional muestran un efecto principal del grupo diagnóstico ( $F(3)=2.741, p=.044$ ).

No se corrobora efecto de los factores sexo y rango etario, ni interacciones entre los factores. En esta primera tarea, las pruebas post-hoc evidencian la homogeneidad en los grupos de edad ( $p=.413$ ) y los grupos clínicos ( $p=.100$ ).

En los contrastes de C se hallan efectos principales significativos para los factores de rango etario ( $F(2)=4.815, p=.009$ ) y grupo diagnóstico ( $F(3)=5.800, p=.001, \eta^2=.075$ ). Nuevamente, sobre las mediciones de errores el factor sexo no tiene efecto, aunque se hallan efectos de interacción entre éste y el factor de rango etario ( $F(2)=3.629, p=.028$ ).

Se corroboran homogeneidad en el desempeño durante C en los tres rangos etarios ( $p=.109$ ), así como la conformación de un subconjunto para los grupos clínicos, TDAH-DA, TDAH-C y TAPZ ( $p=.285$ ), con una media de 1.33, 1.72 y 1.82 errores, mientras que el grupo control presenta un menor número de errores, 0.70 en promedio.

Los resultados para la tarea PC muestran efectos intersujetos para los factores de rango etario ( $F(2)=8.861, p=.000, \eta^2=.076$ ) y grupo diagnóstico ( $F(3)=10.384, p=.000, \eta^2=.126$ ), pero no se corrobora interacción significativa entre ambos factores ( $p>.05$ ).

En esta tercera tarea se evidencia la conformación de subconjuntos homogéneos por rango etario ( $p=.476$ ), para los grupos 7 a 10 años y 11 a 15 años, con un mayor número de errores, respectivamente con 3.43 y 2.86 en promedio, frente al grupo de 16 a 22 años, con 1.60 errores. También TDAH-C y TDAH-DA se aproximan ( $p=.791$ ), con 2.82 y 2.30 errores, en comparación con el alto número de errores del grupo TAPZ (4.23) y la menor proporción en los controles (1.17).

Considerando en el examen conjunto de los factores de rango etario y grupo diagnóstico, se halla que el grupo control entre 16 a 22 años muestra el mejor desempeño, con ningún error en P, 0.65 errores en C y 0.80 en PC. Asimismo, el mayor número de errores se presenta en el grupo TAPZ de 7 a 10 años, que registra el mayor promedio con 0.95 en P, mientras que el grupo TDAH-C, en este mismo rango de edad, en C y PC registra respectivamente 2.15 y 5.65 errores.



### Índices de Interferencia calculados en el Stroop Tradicional

Se describen a continuación los resultados de las cuantificaciones de la interferencia en la versión tradicional, que presentan diferencias significativas por los factores de edad y grupo diagnóstico (ver anexo 9).

Se analizan los índices  $I_D$ ,  $I_R$  e  $I_{REL}$ , calculados por las mediciones brutas de aciertos, así como el índice  $I_V$  que utiliza el número de errores registrados. Se incluye también el cálculo de  $I_G$ , haciendo la salvedad que este se obtiene con una puntuación transformada del número de aciertos, según el factor de corrección sugerido por Golden (1999). No se incluyen índices para las mediciones de latencias de respuesta, en tanto que el tiempo estimado de respuesta es general y no distingue entre aciertos o errores, como se indica anteriormente.

### Sensibilidad y Especificidad

La tabla 11 presenta los descriptivos generales ( $n=240$ ), así como el examen de su capacidad de clasificación entre condición clínica (TDAH-DA, TDAH-C y TAPZ) *versus* grupo control, considerado el criterio de presencia o ausencia de condición clínica.

**Tabla 11**

*Resultados generales y análisis de curvas ROC de los índices de interferencia calculados en el Stroop tradicional*

	<i>M (DE)</i>	<i>Corte</i>	<i>Sensibilidad</i>	<i>Especificidad</i>	<i>AUC</i>
$I_D=C-PC$	23.21 (9.08)	15.50	.19	.86	.505
$I_R=PC/C$	0.57 (0.13)	0.57	.56	.66	.598
$I_G=PC-[(P \times C)/(P+C)]$	0.29 (6.97)	0.06	.58	.71	.641*
$I_{REL}=[(C-PC)/C] \times 100$	42.27 (13.86)	42.46	.60	.75	.669*
$I_V=PC-[(P+C)/2]$	1.70 (3.00)	1.25	.48	.76	.618*

*La interpretación de área bajo la curva (AUC) considera valor: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

El análisis las curvas ROC para los índices de interferencia  $I_D$  e  $I_R$  muestra una baja capacidad de clasificación, con valores inadecuados de sensibilidad y especificidad.

En el caso de  $I_D$ , por ejemplo, al hacer la transformación de los datos considerando el punto de corte en 15.50 (diferencia de aciertos), se halla una buena

capacidad para detectar correctamente altos niveles de interferencia en los participantes que presentan alguna de las condiciones clínicas (86%), aunque cuenta con muy poca capacidad para discriminar los niveles bajos de interferencia en los controles (19%).

Por su parte, los resultados de las pruebas en  $I_G$ ,  $I_{REL}$  e  $I_V$  muestran una moderada capacidad de clasificación. En dichos índices los valores de especificidad tienden a ser adecuados, aunque su sensibilidad es modesta.

El punto de corte de 0.06 identificado para la medición de  $I_G$  coincide con la media informada de la prueba clásica del test (WCST de Golden, 1999). No obstante, los resultados de este índice sitúan a la mayoría de los participantes (94%) dentro del rango normal para un adecuado control de interferencia. Y, excepto en 14 de ellos, correspondientes a los grupos de TDAH-C, TDAH-DA y TAPZ (con 10, 2 y 2 casos), se obtienen resultados de  $I_G$  por debajo de  $-10$  (inferior a un puntaje  $T$  de 40), que efectivamente se consideran como “anormales” o indicativos de un bajo nivel de control de interferencia.

### *Análisis por Edad y Condición Clínica*

Vistos los resultados de las mediciones básicas, se encuentra que los factores de edad y condición clínica tienen efectos significativos (individuales y/o conjuntos) sobre el desempeño en las tareas del Stroop y, por consiguiente, sobre los cálculos de los índices de interferencia. Así, se plantea que el análisis de curvas ROC tenga en cuenta ambos factores. Tal como puede apreciarse en la tabla 12, la capacidad de clasificación de estos índices cambia en cada grupo, incrementándose los puntos de corte identificados conforme el rango etario, siendo en consecuencia más estrictos en los grupos de mayor edad (ver anexo 9.4).

**Tabla 12**

*Análisis de curvas ROC para los índices de interferencia calculados en cada rango etario*

	$I_D$	$I_R$	$I_G$	$I_{REL}$	$I_V$
7 a 10 años	.637*	.498	.521	.571	.658*
11 a 15 años	.577	.692*	.726*	.779**	.640*
16 a 22 años	.539	.644*	.699*	.700*	.582

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Se halla que el cálculo de  $I_D$  resulta eficaz al discriminar los niveles de interferencia únicamente en el grupo de 7 a 10 años, pero no así en los más grandes. Similar a lo que ocurre con  $I_V$ . Por el contrario, en  $I_R$ ,  $I_G$  e  $I_{REL}$  se encuentran valores de discriminación adecuados en los grupos de más edad, pero no en los más pequeños.

### Capacidad discriminante de la tarea PC

Se suma a la anterior selección de índices, un análisis específico de los puntajes obtenidos en la tarea PC (latencias estimadas, aciertos y errores). La tabla 13 presenta los resultados de las curvas ROC para estas mediciones básicas.

**Tabla 13**

*Capacidad de clasificación de las mediciones de la tarea PC en el Stroop tradicional*

	<i>M (DE)</i>	<i>Corte</i>	<i>Sensibilidad</i>	<i>Especificidad</i>	<i>AUC</i>
Latencias estimadas	1455 (519.36)	1139.50	.75	.55	.664*
Aciertos	31.78 (11.32)	38.50	.78	.55	.701*
Errores	2.63 (3.29)	1.54	.59	.70	.688*

*AUC: malo <0.60, \*moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Es importante mencionar que los resultados de la estimación de latencias de respuesta en la tarea PC muestra niveles aceptables de sensibilidad (75%) y especificidad (55%). Al observar el desempeño de los participantes en esta tarea del Stroop tradicional se halla que la totalidad de los participantes de los grupos TDAH-DA, TDAH-C y TAPZ presentan latencias superiores al punto de corte en 1139.50 ms, es decir que presentan una mayor demora para responder al estímulo incongruente. En tanto que sólo los controles entre 16 y 22 años presentan latencias inferiores a este umbral.

La medición directa de aciertos en PC, asimismo, muestra niveles aceptables de sensibilidad (78%) y especificidad (55%). Se halla que 142 participantes de los grupos clínicos y 26 del grupo control logran un desempeño por debajo del umbral de 38.50 aciertos.

E, igualmente, la medición de errores en PC muestra niveles aceptables de sensibilidad (59%) y especificidad (70%). Al observar el desempeño de los participantes

se halla que la totalidad del grupo control, así como el 85% de los TDAH-C entre 16 y 22 años, presentan un promedio por debajo del punto de corte de 1.54 errores. En tanto que todos los participantes de los grupos TDAH-DA y TAPZ, junto con los TDAH-C de 7 a 10 años y de 11 a 15 años, se ubican por encima, con un mayor número de errores.

La tabla 14 presenta la capacidad de clasificación en estas tres mediciones básicas en la tercera tarea, discriminando por edad.

**Tabla 14**

*Análisis de curvas ROC para las mediciones básicas en la tarea PC en cada rango etario*

	Latencia estimada	Aciertos	Errores
7 a 10 años	.674*	.778**	.707*
11 a 15 años	.727*	.805**	.717*
16 a 22 años	.803**	.823**	.670*

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Se halla que estas tres mediciones presentan una capacidad de clasificación de nivel bueno a moderado. Se identifican puntos de corte que disminuyen conforme los grupos de edad, haciéndose progresivamente más exigentes (ver anexo 9.6). Y, en general, se hallan mejores resultados de sensibilidad y especificidad, comparados con los distintos índices antes analizados.

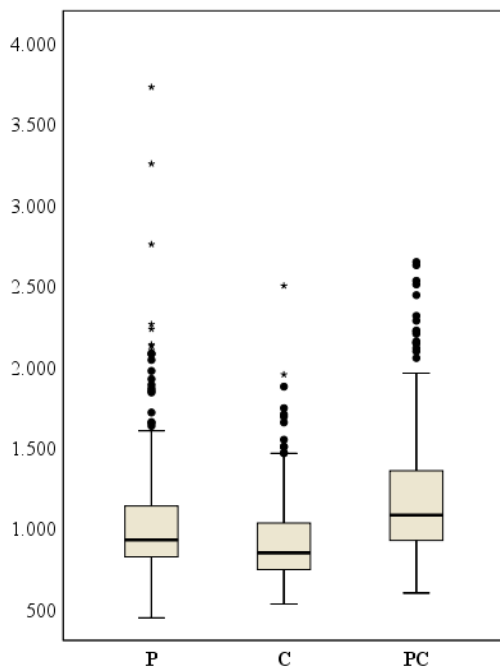
## Desempeño en el Stroop Computarizado

### Latencias Media de Aciertos en el Stroop Computarizado

La tarea PC conlleva mayor demora al responder, con un tiempo promedio de aciertos de 1201 ms ( $DE=402.66$ ), seguido de P con 1058 ms ( $DE=419.45$ ) y, levemente más rápido, de C con 922 ms ( $DE=277.15$ ). La figura 12 ilustra la distribución general de estas mediciones en la versión computarizada del Stroop.

**Figura 12**

*Latencias medias de aciertos en el Stroop computarizado (n=240)*



Si bien los resultados de las tres tareas presentan una tendencia central próxima, dado que las latencias medias de aciertos son cercanas a un segundo por ítem, se advierte una amplia dispersión de las mediciones y presencia de valores atípicos y extremos. Como se expone adelante, tales características están asociadas a los factores de sexo, edad o condición clínica de los participantes.

### Contrastes por Factor

La tabla 15 presenta las mediciones de latencias medias por ítem logrado en cada tarea, comparadas por cada factor. En el anexo 10 pueden consultarse otros estadísticos descriptivos, así como sus respectivas pruebas de ajuste.

**Tabla 15**

*Latencias de respuesta en las tareas del Stroop computarizado y comparación por factores*

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	1011 (378.33)	<i>ns</i>	889 (266.15)	<i>ns</i>	1140 (373.51)	*
Varones	1084 (439.74)		940 (282.31)		1235 (415.29)	
7-10 años	1408 (541.74)	**	1187 (300.23)	**	1576 (433.95)	**
11-15 años	932 (182.51)		827 (149.79)		1071 (230.13)	
16-22 años	834 (121.25)		751 (99.50)		957 (171.90)	
TDAH-DA	1129 (440.07)	**	971 (324.56)	<i>ns</i>	1263 (458.08)	<i>ns</i>
TDAH-C	1089 (488.03)		937 (269.65)		1239 (446.33)	
TAPZ	1094 (434.64)		944 (293.89)		1215 (382.56)	
CONTROL	919 (254.36)		836 (189.73)		1088 (287.71)	

*Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )*

*Diferencia no significativa: ns ( $p > .05$ )*

En el contraste por el factor sexo se observa que las latencias de respuesta son levemente más altas en los varones que las mujeres, sin embargo tal diferencia no alcanza niveles de significancia estadística en la tarea P ( $U(1)=7514.00$ ,  $p=.084$ ), ni en C ( $U(1)=7524.00$ ,  $p=.080$ ) y sólo resulta significativa en PC ( $U(1)=7687.50$ ,  $p=.039$ ).

Mientras que en la comparación por el rango etario se observa una demora generalizada en el grupo de 7 a 10 años, así como una dispersión mayor de sus latencias de respuesta. Los contrastes intersujeto por edad muestran diferencias estadísticas en la tarea P ( $X^2(2)=109.533$ ,  $p=.000$ ), en C ( $X^2(2)=124.055$ ,  $p=.000$ ) y en PC ( $X^2(2)=114.604$ ,  $p=.000$ )

Por su parte, el contraste por el grupo diagnóstico muestra diferencias únicamente para la tarea P ( $X^2(3)=13.035$ ,  $p=.005$ ), mientras que en la tarea C las diferencias son leves y no alcanzan un nivel de significación ( $X^2(3)=7.368$ ,  $p=.061$ ) y en PC las latencias

de respuesta tienden a ser bastante similares entre los grupos clínicos y el grupo control ( $X^2(3)=4.462, p=.216$ ).

### **Efectos e Interacción de los Factores y Análisis Post-hoc**

Como parte de los análisis conjuntos de los factores de edad y diagnóstico, en primer lugar, puede advertirse que las latencias medias por ítem disminuyen en los grupos de más edad, tanto en los grupos clínicos como en el grupo control.

En el rango de 7 a 10 años se encuentran claras discrepancias entre los grupos clínicos y el grupo control, que presenta mayor velocidad en las tres tareas (1132 ms en P, 1002 ms en C y 1290 ms en PC), en tanto que el desempeño tiende a ser homogéneo en los grupos TDAH-DA, TDAH-C y TAPZ de la misma edad con latencias superiores a 1470 ms en P, a 1220 ms en C y a 1620 en PC.

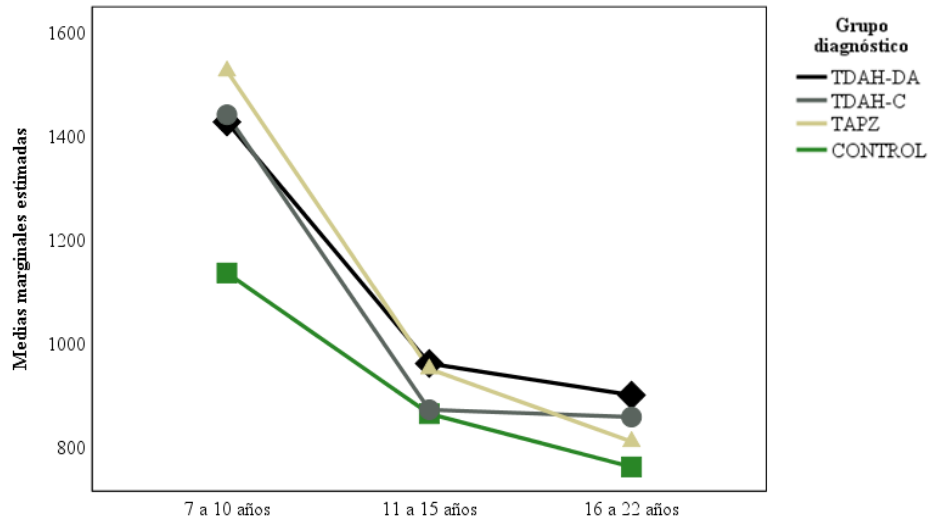
En el rango de 11 a 15 años se observan medias de desempeño similares entre el grupo con diagnóstico TDAH-C y el grupo control en las tres tareas, mientras los grupos TDAH-DA y TAPZ tienden a mostrar más demora.

Y, en el rango de 16 a 22 años, es el grupo con TAPZ el que muestra mejor desempeño en PC (906 ms).

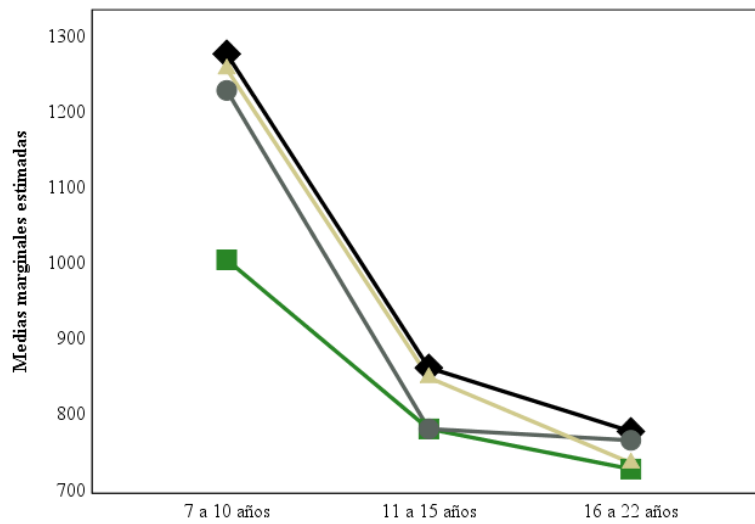
En la página siguiente, la figura 13 ilustra tales discrepancias, mediante la representación de las medias marginales estimadas para las tres tareas de la versión computarizada. Otros valores correspondientes a los análisis de varianza y las pruebas de efectos intersujetos para estas mediciones se incluyen en el anexo 11.

**Figura 13**

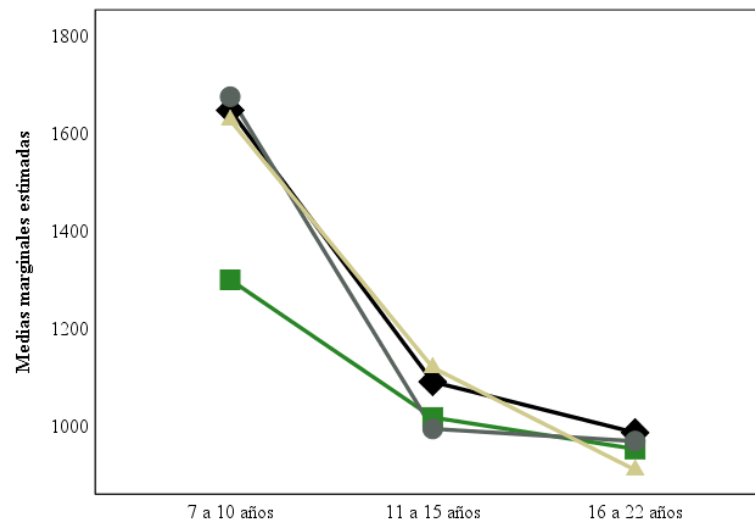
*Discriminación de las latencias de respuesta en la versión computarizada, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:



En la primera tarea, el análisis indica un efecto principal de la edad ( $F(2)=51.441$ ,  $p=.000$ ,  $\eta^2=.323$ ), así como del grupo diagnóstico ( $F(3)=3.559$ ,  $p=.015$ ), aunque no se hallan interacciones entre estos factores ( $p>.05$ ). Tampoco se halla efecto alguno del factor sexo.

Las pruebas Post-hoc para los valores analizados en P demuestran la conformación de subconjuntos homogéneos para los grupos de 11 a 15 años y 16 a 22 años ( $p=.155$ ), que presentan tiempos de respuesta de 834 y 932 ms en promedio, mientras que los más pequeños demoran su respuesta en la lectura, con 1408 ms.

Se observa cierta homogeneidad en el desempeño de los grupos clínicos TDAH-C, TDAH-DA y TAPZ ( $p=.912$ ), que tienen latencias correspondientes a 1089, 1094 y 1129 ms en promedio, siendo el grupo control más rápido, con una media de lectura de 919 ms.

En la tarea C se hallan efectos principales significativos para los factores de rango etario ( $F(2)=93.063$ ,  $p=.000$ ,  $\eta^2=.463$ ) y de grupo diagnóstico ( $F(3)=4.881$ ,  $p=.003$ ,  $\eta^2=.063$ ), no obstante, la interacción entre ambos factores no alcanza nivel de significancia estadística ( $p=.088$ ).

Para esta tarea se confirman las diferencias entre los tres rangos de edad. Mientras que conforman un subconjunto los grupos TDAH-DA, TAPZ y TDAH-C ( $p=.794$ ), con latencias de 971, 944 y 937 ms en promedio, distinguiéndose del grupo control que presenta un tiempo medio de respuesta de 836 ms.

Finalmente, en la tarea PC se encuentran efectos intersujetos para los tres factores: por sexo ( $F(1)=3.921$ ,  $p=.049$ ), por rango etario ( $F(2)=82.113$ ,  $p=.000$ ,  $\eta^2=.432$ ), y por grupo diagnóstico ( $F(3)=3.101$ ,  $p=.028$ ), así como una interacción significativa entre ambos factores ( $F(6)=2.215$ ,  $p=.043$ ).

Así, se evidencia un desempeño diferenciado en PC por de rango etario, sin conformación de subconjuntos homogéneos, aunque por el factor de grupo diagnóstico, al igual que en la segunda tarea, se observa homogeneidad de los grupos TDAH-DA, TDAH-C y TAPZ ( $p=.800$ ), con latencias de respuesta de 1263, 1239 y 1215 ms en promedio, en contraste con la media de desempeño del grupo control de 1088 ms.

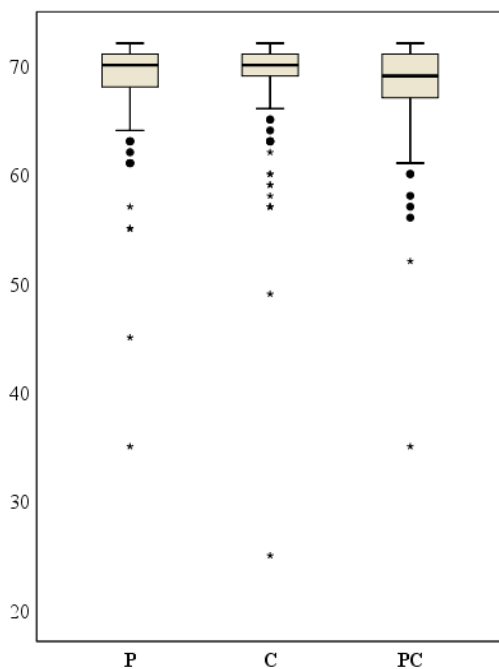
### Número de Aciertos en el Stroop Computarizado

La versión computarizada registra un máximo de respuestas, dado que se muestran 72 ítems en cada tarea. De modo que el total de aciertos logrados se complementa con las comisiones registradas, dando lugar a interpretaciones sobre la precisión y eficiencia en la ejecución de las tareas del test. Teniendo en cuenta que no hay límite de tiempo para responder, es esperable un alto número de aciertos y pocos errores.

En la figura 14 puede observarse como la distribución del número de aciertos tiene una asimetría negativa, con una tendencia a responder adecuadamente a las tres tareas. No obstante, al igual que en las demás mediciones, se registran valores atípicos que pueden explicarse en el análisis de los factores de edad y condición clínica.

**Figura 14**

*Número de aciertos en el Stroop computarizado (n=240)*



El promedio de respuestas correctas es bastante próximo para las tres tareas, en P 69.06 aciertos ( $DE=4.072$ ), en C 69.27 aciertos ( $DE=4.278$ ) y en PC 68.24 aciertos ( $DE=3.977$ ). Se evidencia un mejor rendimiento general en la nominación de los estímulos neutros, superando incluso la lectura.

### Contrastes por Factor

La tabla 16 presenta las mediciones de aciertos en las tres tareas del Stroop computarizado, las cuales muestran diferencias estadísticamente significativas por sexo, edad y condición clínica. Otros valores descriptivos aparecen en el anexo 12.

**Tabla 16**

*Aciertos en las tareas del Stroop computarizado y comparación por factores*

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	69.85 (2.44)	*	69.76 (3.16)	*	69.22 (2.73)	**
Varones	68.62 (4.69)		68.99 (4.78)		67.69 (4.44)	
7-10 años	66.79 (5.955)		67.38 (6.405)		66.61 (5.053)	
11-15 años	66.51 (2.093)	**	69.50 (2.561)	**	68.31 (3.271)	**
16-22 años	70.89 (1.222)		70.93 (1.156)		69.80 (2.543)	
TDAH-DA	68.08 (6.127)		69.60 (2.492)		68.47 (3.362)	
TDAH-C	68.70 (3.752)	**	67.35 (7.253)	**	66.93 (5.757)	**
TAPZ	68.90 (3.261)		69.32 (2.765)		67.88 (3.390)	
CONTROL	70.57 (1.671)		70.80 (1.205)		69.68 (2.021)	

*Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )*

*Diferencia no significativa: ns ( $p > .05$ )*

Inicialmente en el contraste por sexo, se observa que las mujeres obtienen un número levemente más alto de aciertos. Esta diferencia es significativa tanto en la tarea P ( $U(1)=5436.50$ ,  $p=.020$ ), como en C ( $U(1)=5626.50$ ,  $p=.049$ ) y en PC ( $U(1)=5037.50$ ,  $p=.002$ ).

Asimismo, el comparativo por edad muestra significativas variaciones en el desempeño de los grupos, tanto en P ( $X^2(2)=52.78$ ,  $p=.000$ ), como en C ( $X^2(2)=45.52$ ,  $p=.000$ ) y en PC ( $X^2(2)=34.88$ ,  $p=.000$ ).

Lo mismo ocurre en el contraste por condición clínica, con marcadas diferencias en P ( $X^2(3)=16.80$ ,  $p=.001$ ), en C ( $X^2(3)=21.68$ ,  $p=.000$ ) y en PC ( $X^2(3)=13.95$ ,  $p=.000$ ).

### Efectos e Interacción de los Factores y Análisis Post-hoc

En el examen conjunto de los factores se halla el mejor desempeño en las mujeres de 16 a 22 años, siendo más precisas en promedio las participantes del grupo control

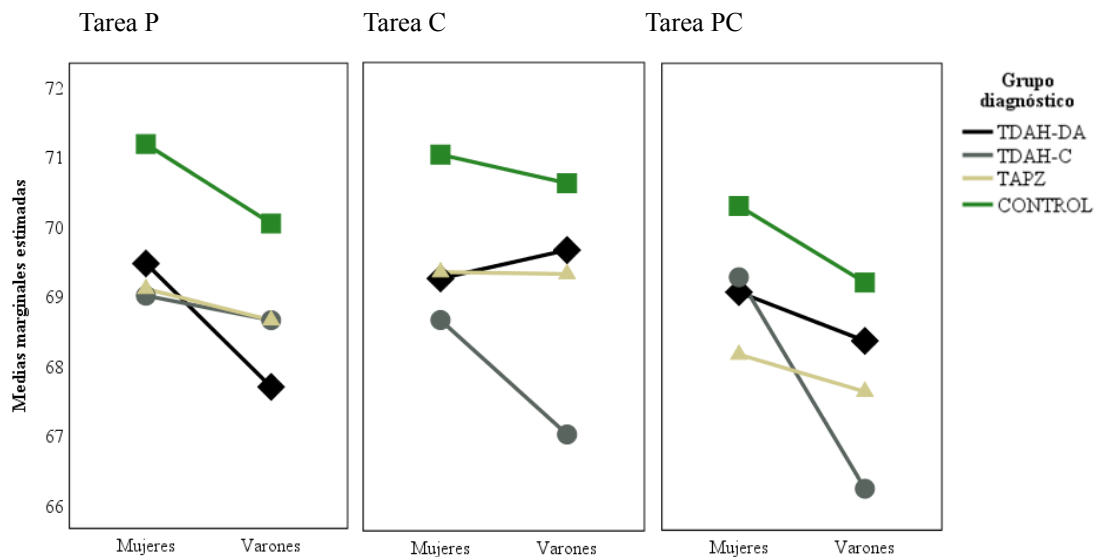
(71.44 aciertos) en la primera tarea, mientras que el grupo con diagnóstico TDAH-C iguala los resultados del grupo control en C (71.67 aciertos) e incluso los supera en PC (71.33 aciertos).

En tanto que el peor desempeño fue registrado por los varones de 7 a 10 años del grupo TDAH-C durante las tareas C y P (64.83 y 66.83 aciertos), mientras que en PC es el grupo TAPZ el que muestra una menor precisión al responder (66.00 aciertos).

Las figuras 15 y 16 presentan los resultados de aciertos para los contrastes Sexo×Grupo y Edad×Grupo, respectivamente. Otros valores obtenidos en los análisis de efectos intersujetos pueden consultarse en el anexo 13.

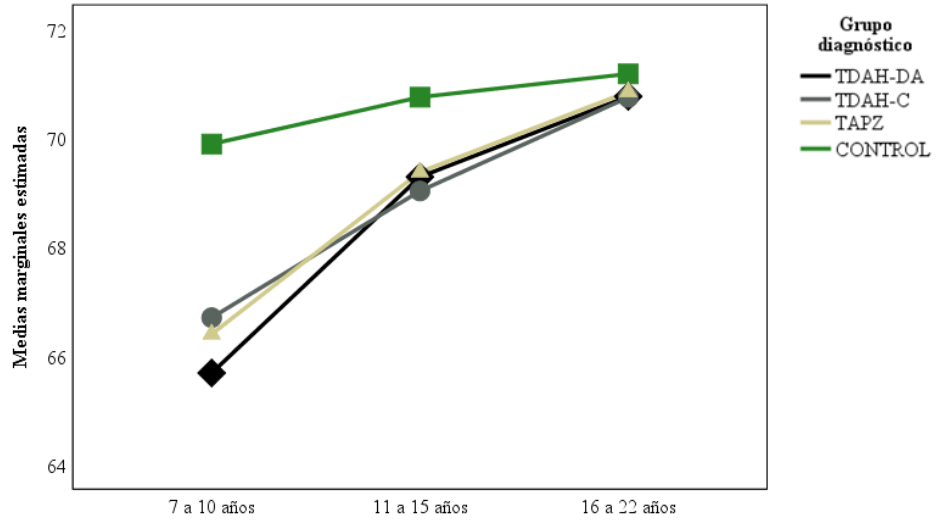
**Figura 15**

*Discriminación de los aciertos en la versión computarizada, por sexo y grupo diagnóstico*

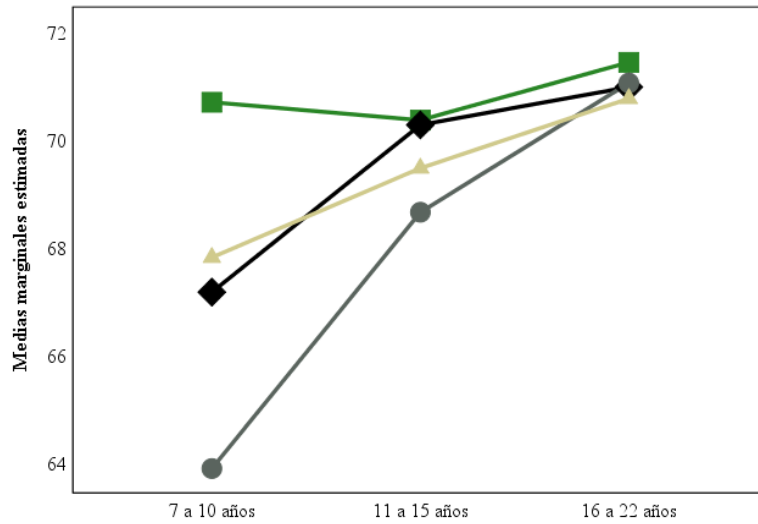


**Figura 16**

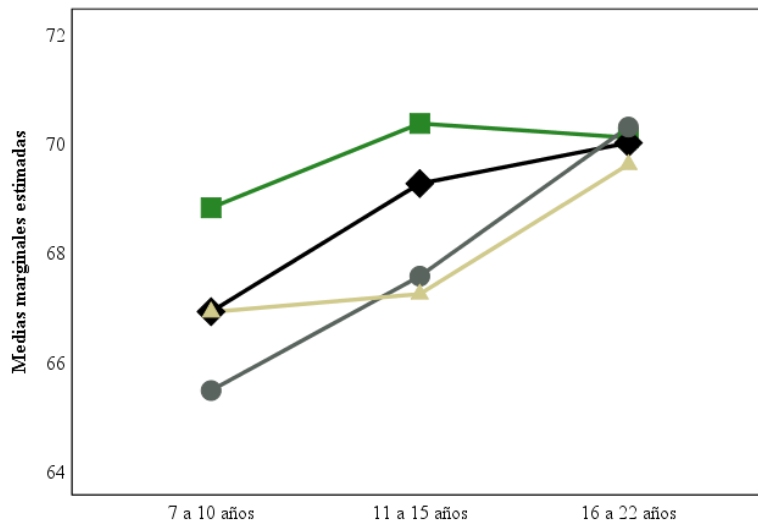
*Discriminación de los aciertos en la versión computarizada, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:

Las pruebas de efectos intersujetos para P en el Stroop computarizado muestran un efecto principal del factor rango etario ( $F(2)=18.049, p=.000, \eta^2=.143$ ), así como del factor grupo diagnóstico ( $F(3)=3.702, p=.013$ ), aunque no se corrobora interacciones entre éstos ( $p>.05$ ).

El análisis post-hoc indica que todos los grupos son diferentes en su desempeño, observándose un incremento en el número de las respuestas correctas con la edad. Por otra parte, los grupos clínicos son homogéneos ( $p=.600$ ), con una media cercana a 68 aciertos, mientras el grupo control logra 70.57 aciertos.

Así también, en la tarea C se hallan efectos principales significativos del rango etario ( $F(2)=15.116, p=.000, \eta^2=.123$ ) y del grupo diagnóstico ( $F(3)=4.162, p=.007$ ).

Los participantes en los grupos de 11 a 15 años y de 16 a 22 años tienen un desempeño similar en la segunda tarea ( $p=.053$ ), con resultados superiores a 69.50 aciertos, por encima del grupo de 7 a 10 años que tiene 67.38 aciertos. Los grupos TAPZ, TDAH-DA y controles conforman un conjunto homogéneo ( $p=.155$ ), con 69 aciertos o superior, en tanto que el grupo TDAH-C alcanza un promedio de 67.35 aciertos.

Y, por su parte en la tarea PC, los efectos intersujetos resultan significativos por sexo ( $F(1)=6.334, p=.013$ ), por rango etario ( $F(2)=11.312, p=.000, \eta^2=.095$ ) y por grupo diagnóstico ( $F(3)=3.327, p=.021$ ).

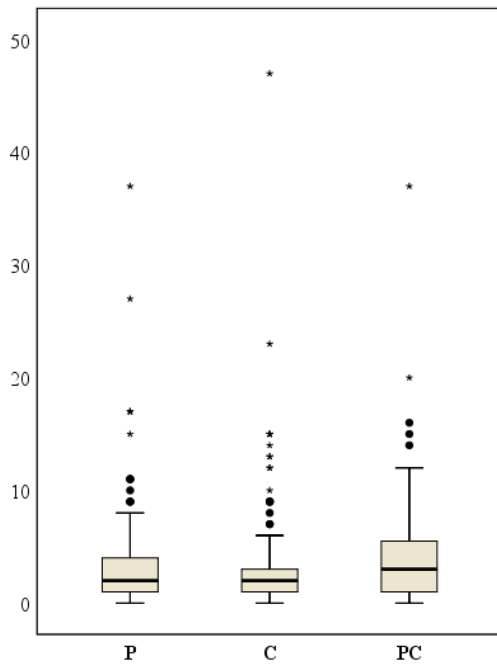
El desempeño es heterogéneo por rango etario en esta tercera tarea, aunque por condición clínica se confirma que los grupos TAPZ, TDAH-DA y controles tienen un desempeño similar ( $p=.092$ ), entre 66 y 68 aciertos, mientras que el grupo control logra 69.68 aciertos en promedio

### **Número de Errores en el Stroop Computarizado**

Curiosamente, la versión computarizada presenta un menor número de errores en la segunda tarea, con una media de 2.73 ( $DE=4.28$ ), seguido de 2.93 ( $DE=4.07$ ) en P y de 3.81 ( $DE=4.03$ ) en PC. En la figura 17 se muestra la asimetría positiva en cada tarea, esto es, la distribución tiende a mostrar pocos errores, aunque se registran valores atípicos. Otros descriptivos de estas mediciones de la eficiencia en la respuesta aparecen en el anexo 14.

**Figura 17**

Número de errores en el Stroop computarizado (n=240)



### Contrastes por Factor

Como presenta la tabla 17, a continuación, el análisis del número de errores muestra diferencias significativas en los tres factores.

**Tabla 17**

Número de errores en las tareas del Stroop computarizado y comparación por factores

	P		C		PC	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	2.14 (2.45)	*	2.24 (3.15)	*	2.84 (2.82)	**
Varones	3.38 (4.69)		3.01 (4.77)		4.36 (4.48)	
7-10 años	5.21 (5.95)		4.63 (6.40)		5.48 (5.10)	
11-15 años	2.48 (2.10)	**	2.50 (2.56)	**	3.74 (3.34)	**
16-22 años	1.11 (1.22)		1.08 (1.15)		2.23 (2.58)	
TDAH-DA	3.92 (6.12)		2.40 (2.49)		3.58 (3.38)	
TDAH-C	3.30 (3.57)	**	4.65 (7.25)	**	5.10 (5.76)	**
TAPZ	3.10 (3.26)		2.68 (2.76)		4.25 (3.58)	
CONTROL	1.42 (1.68)		1.20 (1.20)		2.23 (2.02)	

Diferencia significativa: \*\* ( $p < .01$ ) y \* ( $p < .05$ )

Diferencia no significativa: ns ( $p > .05$ )

Se observa un mayor número de errores en los varones, de modo que se contrastan diferencias significativas por sexo tanto en P ( $U(1)=7838.00, p=.017$ ), como en C ( $U(1)=7618.00, p=.049$ ) y en PC ( $U(1)=8185.00, p=.002$ ).

Los estadísticos de contraste por rango etario también muestran importantes variaciones en el número de errores registrados durante las tareas de la versión computarizada, significativas en P ( $X^2(2)=53.372, p=.000$ ), en C ( $X^2(2)=45.528, p=.000$ ) y en PC ( $X^2(2)=34.980, p=.000$ ).

Y, del mismo modo, el contraste por el grupo diagnóstico muestra diferencias significativas en las tareas, en P ( $X^2(3)=17.424, p=.001$ ), en C ( $X^2(3)=21.868, p=.000$ ) y en PC ( $X^2(3)=14.113, p=.003$ ).

La versión computarizada del Stroop utilizada en este estudio permite reconocer el tipo de error cometido durante la tercera tarea, registrándose un promedio de 2.38 ( $DE=2.54$ ) errores congruentes, sumadas a 1.44 ( $DE=2.11$ ) de imprecisiones.

A continuación, la tabla 18 indica el detalle de las mediciones de error específicas de la tercera tarea, correspondientes a errores congruentes (si responde leyendo, en vez de cumplir la consigna de nominar el color del estímulo palabra-color) e imprecisiones (o error simple, cuando no se respeta criterio de selección alguno).

**Tabla 18**

*Tipo de error arrojado por el Stroop computarizado en la tarea de conflicto palabra-color*

	Error congruente		Imprecisión	
	<i>M (DE)</i>	<i>p</i>	<i>M (DE)</i>	<i>p</i>
Mujeres	1.79 (1.98)	**	1.05 (1.52)	*
Varones	2.70 (2.75)		1.66 (2.35)	
7-10 años	3.22 (2.91)		2.25 (2.84)	
11-15 años	2.35 (2.26)	**	1.39 (1.82)	**
16-22 años	1.55 (2.12)		0.67 (0.92)	
TDAH-DA	2.12 (2.04)		1.47 (2.28)	
TDAH-C	3.02 (3.30)	*	2.08 (2.83)	*
TAPZ	2.78 (2.70)		1.47 (1.81)	
CONTROL	1.58 (1.56)		0.73 (0.77)	

*Diferencia significativa: \*\* ( $p<.01$ ) y \* ( $p<.05$ )*

*Diferencia no significativa: ns ( $p>.05$ )*



Los resultados de errores congruentes como de imprecisiones presentan diferencias significativas para los tres factores ( $p < .05$ ), aunque no aportan nueva información sobre el desempeño en la versión y los contrastes que se realizan respecto de la versión tradicional, por lo que para análisis posteriores sólo se tiene en cuenta el número total de errores en la tercera tarea.

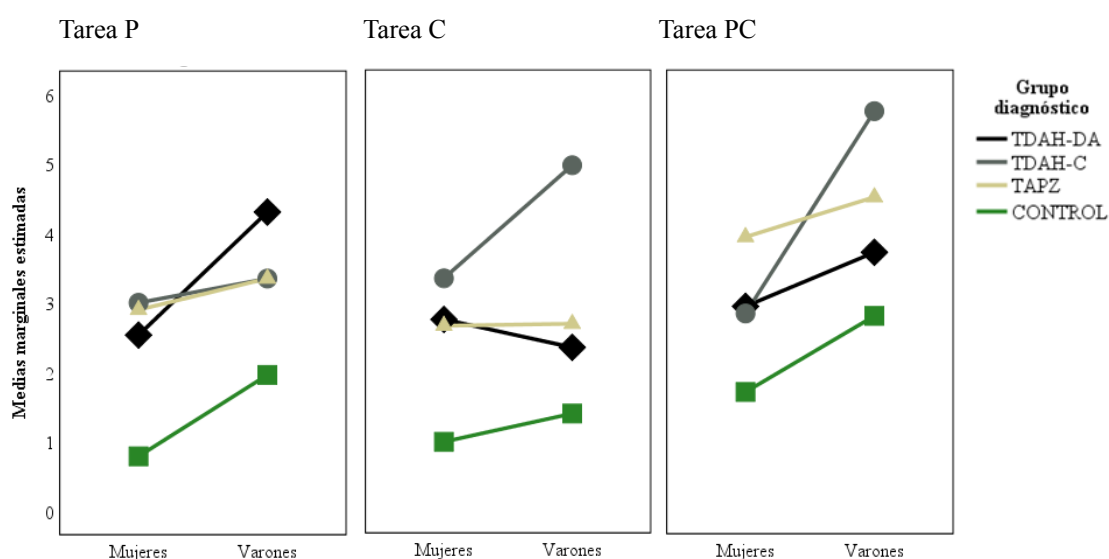
### Efectos e Interacción de los Factores y Análisis Post-hoc

El análisis conjunto de los factores indica que el grupo entre 7 y 10 años presenta el mayor número de errores en promedio, siendo los menos eficientes los varones con diagnóstico TDAH-C, con 9.07 errores en C y en PC, y los varones TDAH-DA en la tarea P, con 8,63 errores en promedio. En tanto que los más grandes tienden a presentar una proporción menor de errores, si bien se mantienen las diferencias por sexo.

La figura 18, a continuación, y la figura 19, en la página siguiente, ilustran las discrepancias mencionadas, considerando los contrastes por los factores (Sexo×Grupo y Edad×Grupo), que dan cuenta de las variaciones en la eficiencia de cada grupo en la versión computarizada. Resultados adicionales de los análisis de varianza y las pruebas intersujeto para estas mediciones se hallan en el anexo 15.

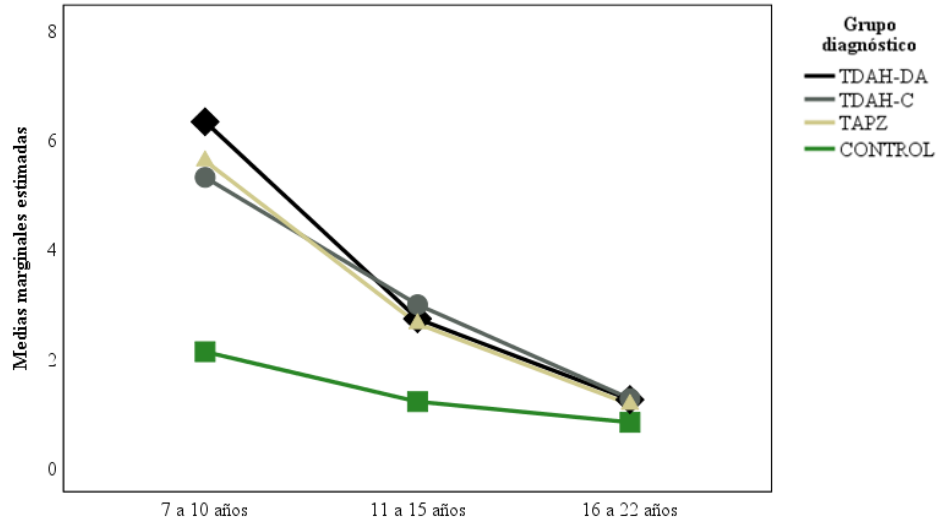
**Figura 18**

*Discriminación de los errores en la versión computarizada, por sexo y grupo diagnóstico*

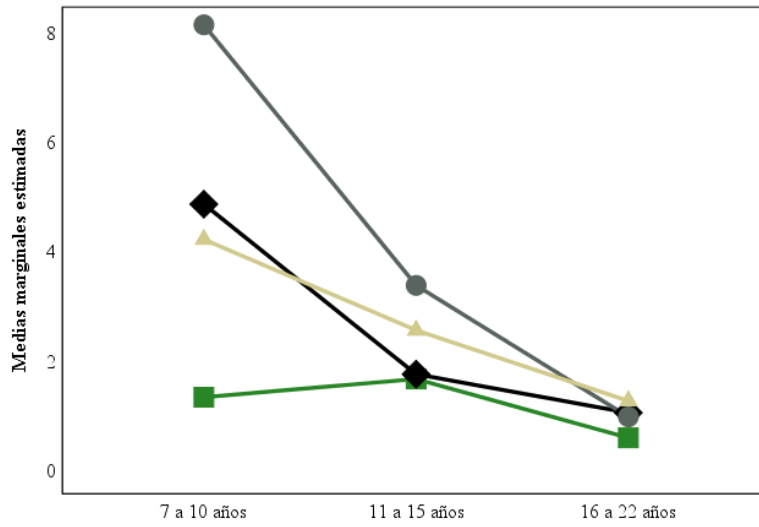


**Figura 19**

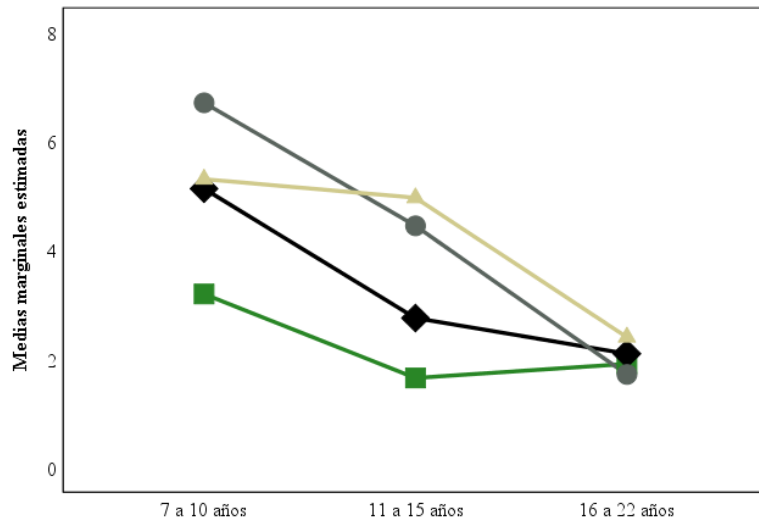
*Discriminación de los errores en la versión computarizada, por rango etario y grupo diagnóstico*



Tarea P:



Tarea C:



Tarea PC:

Las pruebas de efectos intersujetos para P en el Stroop computarizado sugieren un efecto principal del factor rango etario ( $F(2)=10.070, p=.000, \eta^2=.143$ ), así como del factor grupo diagnóstico ( $F(3)=3.767, p=.012$ ). Mientras que las diferencias por sexo observadas no alcanzan el nivel de significancia estadística ( $p>.05$ ).

Las pruebas Tukey demuestran la conformación de subconjuntos homogéneos para los rangos de edad 16 a 22 años y 11 a 15 años ( $p=.059$ ), con un registro de 1.11 y 2.47 errores, respecto del grupo de 7 a 10 años con un promedio más alto de errores (5.21).

Los grupos TAPZ, TDAH-DA y TDAH-C conforman un subconjunto homogéneo ( $p=.673$ ), ya que presentan una media de a 3.10 errores o superior, mientras el grupo control tiene 1.42 errores.

En la tarea C se hallan efectos principales para los factores de rango etario ( $F(2)=15.116, p=.000, \eta^2=.123$ ), y grupo diagnóstico ( $F(3)=4.162, p=.007$ ), así como una interacción significativa entre estos dos factores ( $F(6)=2.171, p=.047$ ).

En esta segunda tarea, los participantes de 11 a 15 años y de 16 a 22 años conforman un subconjunto homogéneo ( $p=.068$ ), con una media de 2.50 y 1.08, en comparación con el registro de 4.63 errores del grupo de 7 a 10 años.

Los grupos TAPZ, TDAH-DA y controles conforman un subconjunto homogéneo ( $p=.222$ ), respectivamente con 2.68, 2.40 y 1.20 errores, mientras que el grupo TDAH-C presenta 4.65 errores en promedio.

Finalmente, en la tarea PC se encuentran efectos intersujetos para los tres factores: por sexo ( $F(1)=6.162, p=.014$ ), por rango etario ( $F(2)=11.535, p=.000, \eta^2=.096$ ) y por grupo diagnóstico ( $F(3)=3.583, p=.015$ ).

Se evidencian un desempeño marcadamente distinto por edad, con una media de errores de 5.47 en el grupo de 7 a 10 años, que disminuye progresivamente en los otros dos grupos (3.74 y 2.23 errores); aunque por el factor de diagnóstico, al igual que en la segunda tarea, los grupos TDAH-C, TAPZ y TDAH-DA muestran homogeneidad ( $p=.161$ ), con 5.10, 4.25 y 3.58 errores respectivamente, frente a 2.32 errores en promedio registrados en el grupo control.

### Índices de Interferencia Calculados en el Stroop Computarizado

Los resultados generales de los índices de interferencia calculados en la versión computarizada no presentan diferencias significativas en las comparaciones de medias por edad y grupo diagnóstico (ver anexo 16).

A continuación, se incluyen los cálculos de  $I_D$ ,  $I_R$ ,  $I_{READ}$  e  $I_{NAME}$ , que utilizan las mediciones de latencias medias de aciertos y el índice  $I_V$  calculado con el número de errores registrados en esta versión. No se incluye resultados para  $I_G$ , que utiliza las mediciones de aciertos, en razón a que este método de cuantificación se aplica como parte de la prueba estandarizada (Golden, 1978, 1999), que considera originalmente un tipo de respuesta verbal, no habiéndose establecido baremos o referentes de interpretación para una versión computarizada.

### Sensibilidad y Especificidad

La tabla 19 presenta los descriptivos generales ( $n=240$ ) y los resultados de los índices de interferencia en la versión computarizada, así como el análisis de la sensibilidad y especificidad.

**Tabla 19**

*Resultados generales y análisis de curvas ROC de los índices de interferencia calculados en el Stroop computarizado*

	<i>M (DE)</i>	<i>Corte</i>	<i>Sensibilidad</i>	<i>Especificidad</i>	<i>AUC</i>
$I_D=PC-C$	205 (126.542)	458.50	.14	.97	.529
$I_R=C/PC$	0.79 (0.09)	0.81	.63	.40	.482
$I_{READ}=[(PC-P)/P]$	0.15 (0.17)	0.57	.04	.98	.419
$I_{NAME}=[(PC-C)/C]$	0.27 (0.67)	0.14	.20	.90	.521
$I_V=PC-[(P+C)/2]$	1.13 (2.27)	2.25	.29	.82	.492

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Ninguno de los valores arrojados es aceptable en los análisis de las curvas ROC, los resultados de *AUC* son malos en todos los índices analizados y no presentan niveles adecuados de sensibilidad. Dicho de otra forma, los índices de interferencia calculados para la versión computarizada del Stroop no resultan eficaces en la discriminación de la condición clínica de los participantes.

Se intenta un segundo análisis de la capacidad discriminante de dichos índices, teniendo en cuenta que los contrastes por factor muestran efectos significativos en la versión computarizada.

Como presenta la tabla 20, la capacidad de clasificación de estos índices varía apenas levemente sus valores respecto del análisis general. Si bien se aprecian cambios en los puntos de corte en los subgrupos, éstos no son concluyentes, mostrando en su mayoría niveles adecuados de especificidad, pero muy baja sensibilidad (ver anexo 16.4).

**Tabla 20**

*Análisis de curvas ROC para los índices de interferencia calculados considerando los factores analizados*

	I <sub>D</sub>	I <sub>R</sub>	I <sub>READ</sub>	I <sub>NAME</sub>	I <sub>V</sub>
Mujeres	.536	.451	.449	.553	.421
Varones	.516	.489	.404	.512	.517
7 a 10 años	.674*	.587	.520	.417	.396
11 a 15 años	.489	.444	.439	.557	.662*
16 a 22 años	.437	.409	.308	.597	.427

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

### Capacidad discriminante de la tarea PC

Al igual que se hiciera en la versión tradicional, se suma a la anterior selección de índices de interferencia, un análisis específico de los puntajes brutos obtenidos en la tarea PC del Stroop computarizado (latencia media de aciertos, número de aciertos y número de errores).

**Tabla 21**

*Capacidad de clasificación de las mediciones de la tarea PC del Stroop computarizado*

	M (DE)	Corte	Sensibilidad	Especificidad	AUC
Latencia media de aciertos	1201 (402.66)	1499.50	.24	.95	.590
Aciertos	68.24 (3.98)	69.50	.59	.67	.641*
Errores	3.81 (4.03)	2.50	.59	.66	.643*

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Anteriormente, la tabla 21 presenta los resultados de las curvas ROC para estas mediciones, donde los análisis de aciertos y errores muestran una mejor capacidad de clasificación de la condición clínica, que las latencias medias de respuesta.

Vistos los puntos de corte general se halla que 46 de los participantes se ubican con una demora superior a 1499.50 ms, de los cuales 43 pertenecen a los grupos clínicos, en su mayoría en el rango entre 7 y 10 años. En otras palabras, el 80.8% del total de la muestra presenta una velocidad de respuesta que se clasifica como adecuada.

La medición de latencias medias de aciertos presenta una buena capacidad para detectar correctamente altos niveles de interferencia en los participantes que presentan alguna de las condiciones clínicas (95%), aunque poca capacidad para discriminar niveles bajos de interferencia en los controles (24%).

De otra parte, los umbrales señalados para las mediciones de aciertos y errores ubican al 52.5% de los participantes en niveles inadecuados de precisión y eficiencia al responder a la tarea de conflicto. Esto es, un total de 20 controles y 106 de los grupos clínicos, predominantemente TDAH-C, y de diferentes edades, logran correctamente menos de 69 aciertos o cometen más de dos errores. Ambas mediciones muestran niveles aceptables de sensibilidad (59%, en ambos casos) y de especificidad (67 y 66%, respectivamente).

La tabla 22 presenta la capacidad de clasificación en estas tres mediciones básicas en la tercera tarea del Stroop computarizado, considerando además los factores de sexo y rango etario.

**Tabla 22**

*Análisis de curvas ROC para las mediciones básicas en la tarea PC del Stroop computarizado, considerando los factores analizados*

	Latencia media de aciertos	Aciertos	Errores
Mujeres	.613*	.600*	.629*
Varones	.552	.628*	.603*
7 a 10 años	.796**	.685*	.689*
11 a 15 años	.577	.776**	.779**
16 a 22 años	.530	.498	.499

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

La capacidad de clasificación de la interferencia varía de una medición a otra y se identifican distintos puntos de corte conforme los grupos de edad y el sexo (ver anexo 16.6 y 16.7).

En líneas generales, el número de aciertos y el número total de errores son las mediciones que arrojan mejores niveles de sensibilidad y especificidad, seguidas por la latencia media de aciertos, que resulta especialmente adecuada en los grupos de menor edad.

### Stroop tradicional versus Stroop Computarizado

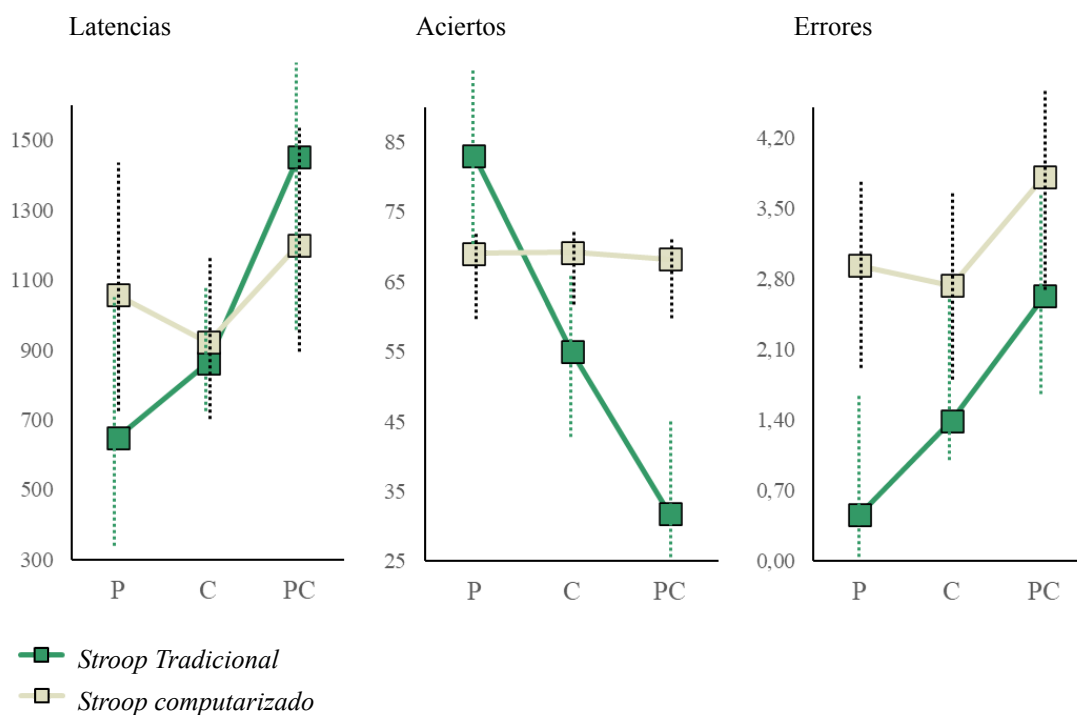
En este apartado presenta la comparación de las mediciones básicas entre las versiones tradicional y computarizada del test, seguida de los índices calculados para identificar el nivel de interferencia, así como la inspección de los umbrales hallados en cada caso y su capacidad de clasificación.

#### Comparación General de las Mediciones Básicas ( $n=240$ )

Las mediciones básicas analizadas permiten reconocer el desempeño de los participantes en cada versión del Stroop, en términos de velocidad, precisión y eficiencia. Como ilustra la figura 20, cada una de las mediciones describe un desempeño diferenciado en las tareas en cada versión.

**Figura 20**

*Comparación del desempeño entre las versiones ( $n=240$ )*



Rápidamente se advierten no sólo discrepancias en cuanto a las magnitudes observadas, sino una relación diferenciada entre las tareas de cada versión:  $P < C < PC$  en el Stroop tradicional y  $C < P < PC$  en el Stroop computarizado, tanto para latencias como para errores, y en sentido inverso para aciertos.



Si bien se reconoce una importante diferencia entre las versiones respecto de las mediciones consideradas, ambas versiones coinciden en mostrar cómo la tarea de conflicto PC requiere más tiempo promedio para su resolución, tiende a implicar mayor número de errores y un menor número de aciertos. Mientras que en las tareas P y C se advierte una relación inversa, esto es, la lectura del listado de palabras es más rápida, precisa y eficiente en la versión tradicional que la nominación del color, pero el desempeño de la segunda tarea supera a la primera en la versión computarizada.

### **Comparación entre Grupos Clínicos ( $n=180$ ) y Controles ( $n=60$ )**

Como se menciona previamente en la descripción de cada versión del Stroop, el análisis de las diferencias por el factor de condición clínica destaca el desempeño general del grupo control, que tiene una respuesta más rápida, precisa y eficiente, en contraste con los grupos clínicos (TDAH-DA, TDAH-C y TAPZ), que tienden a mostrar mayor demora, menor número de aciertos y mayor número de errores en las tres tareas.

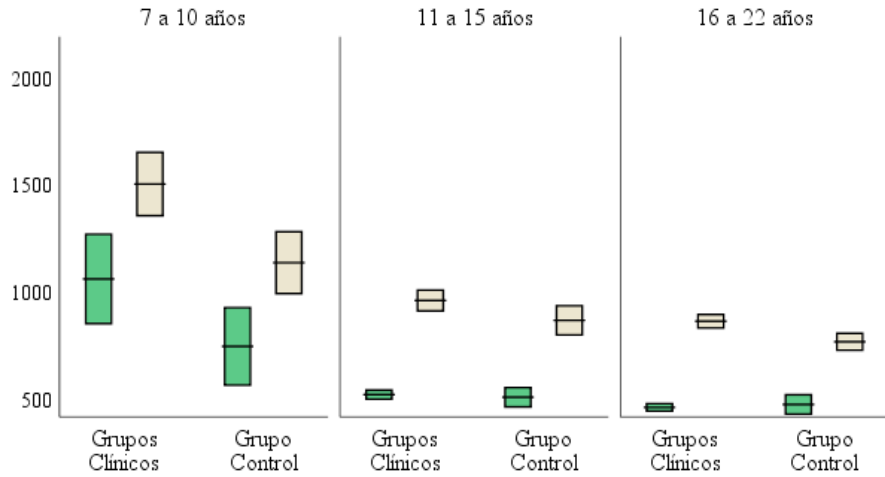
A su vez, cada uno de los grupos clínicos presenta un desempeño diferenciado, según se considere una u otra medición en ambas versiones del test. Los participantes con diagnóstico TDAH-DA son lentos al responder, los TDAH-C son más veloces, pero cometen más errores. Por su parte, los TAPZ muestran bajo rendimiento en las tareas P y C, pero sus registros de las tres mediciones básicas en PC se aproximan al grupo control.

Asimismo, tanto en la versión tradicional como en la computarizada del Stroop, el factor de rango etario muestra diferencias significativas en las tres tareas, así como efectos principales en todos los contrastes realizados y, en algunos casos, efectos de interacción. Por lo que se considera de suma importancia el análisis del carácter evolutivo de los procesos cognitivos que implican las tres tareas. En particular, se observa que la resolución de la tarea de conflicto acentúa las diferencias entre los grupos clínicos y control.

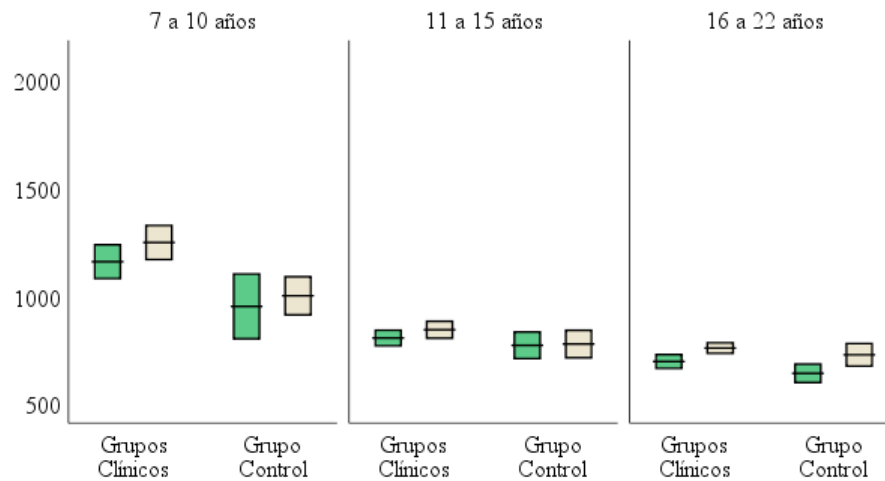
A continuación, las figuras 21, 22 y 23, presentan los cambios asociados a la edad observados en ambas versiones del Stroop. La inspección visual permite advertir una curva decreciente en las mediciones de latencias y errores, e inversamente, el progresivo incremento de los aciertos.

**Figura 21**

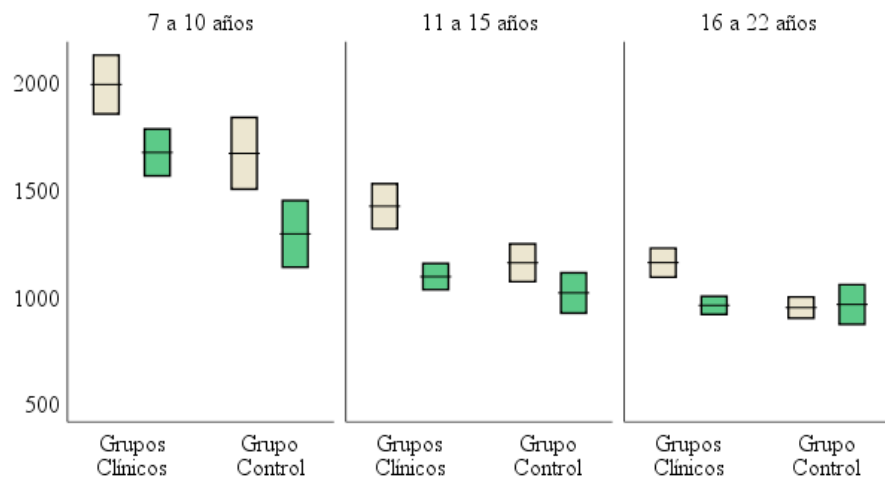
*Comparación de latencias en ambas versiones del Stroop*



Tarea P:



Tarea C:



Tarea PC:

■ Stroop Tradicional (latencias estimadas de respuesta)

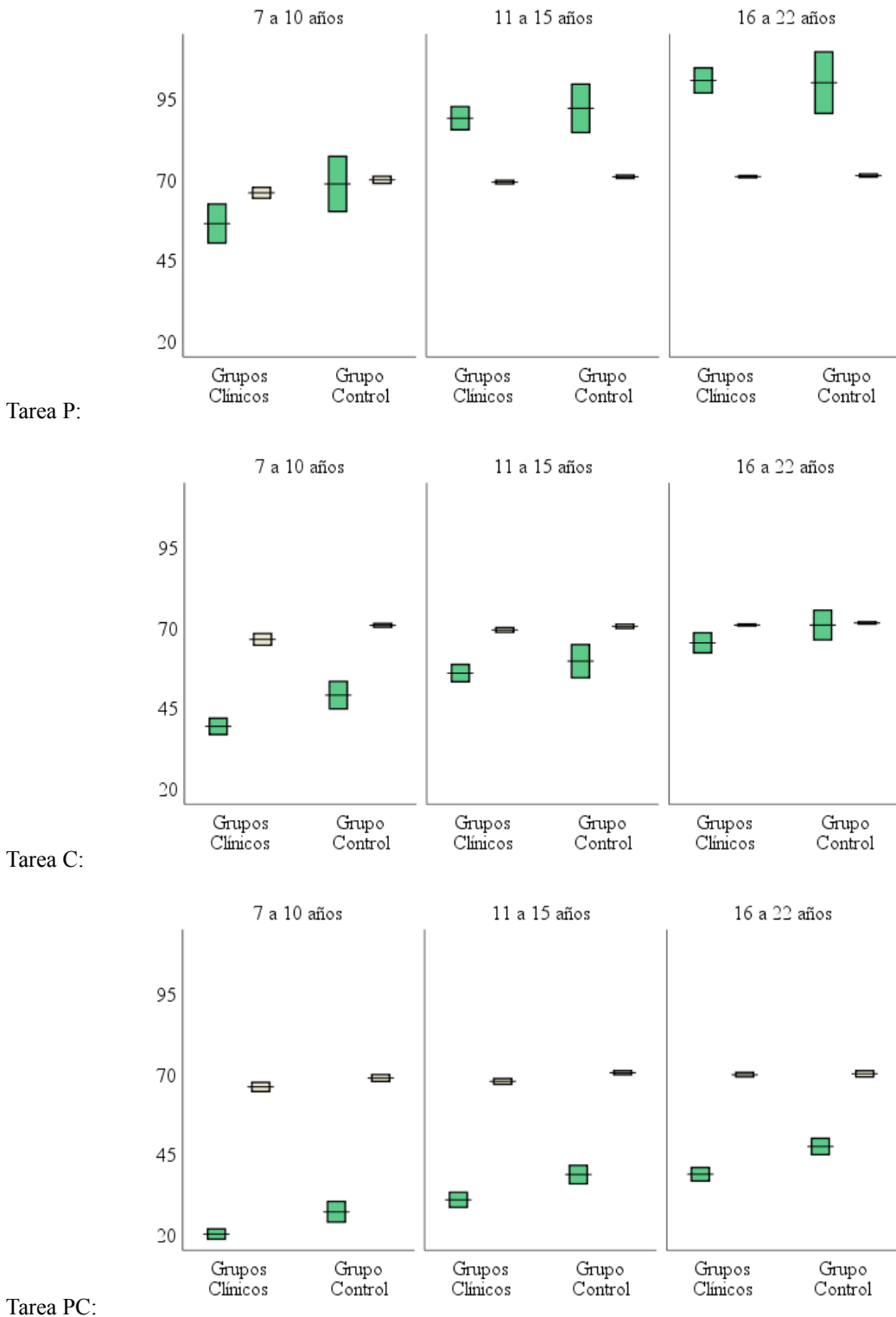
■ Stroop computarizado (latencias medias de aciertos)

Barras de error para latencias registradas en ambas versiones (intervalo de confianza al 95%).

Valores en ms.

**Figura 22**

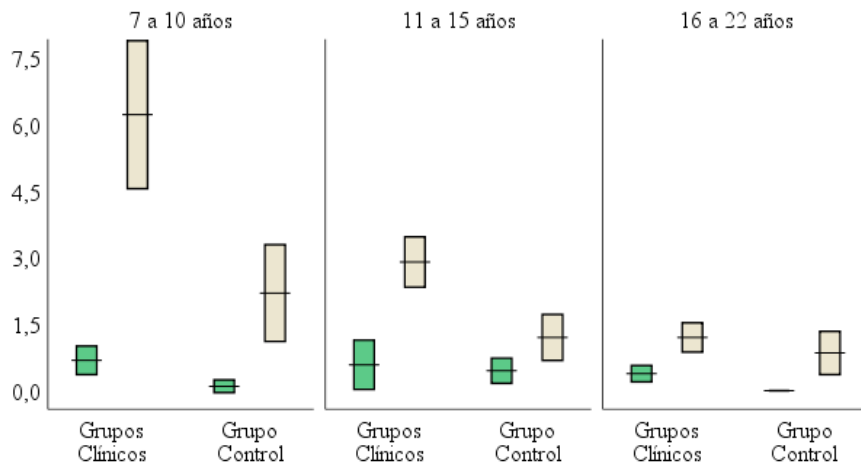
*Comparación de número de aciertos en ambas versiones del Stroop*



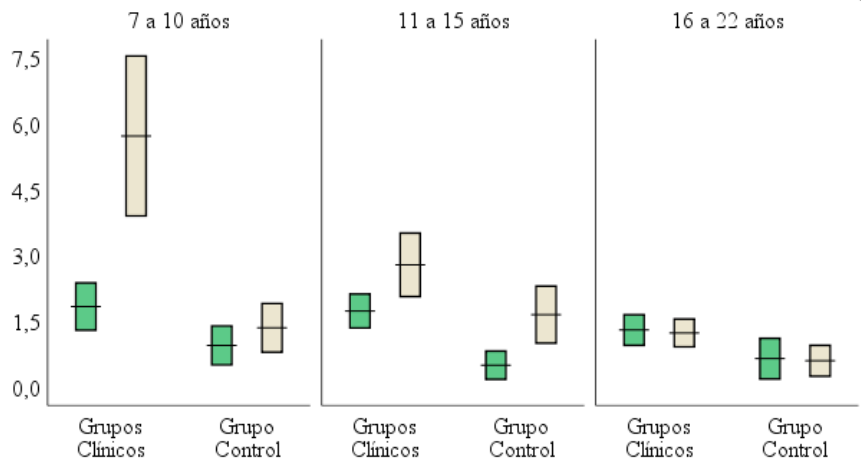
■ Stroop Tradicional (número de respuestas correctas en 45 segundos)  
■ Stroop computarizado (número de respuestas correctas de 72 ítems mostrados)  
 Barras de error para el número de aciertos en ambas versiones (intervalo de confianza al 95%).

**Figura 23**

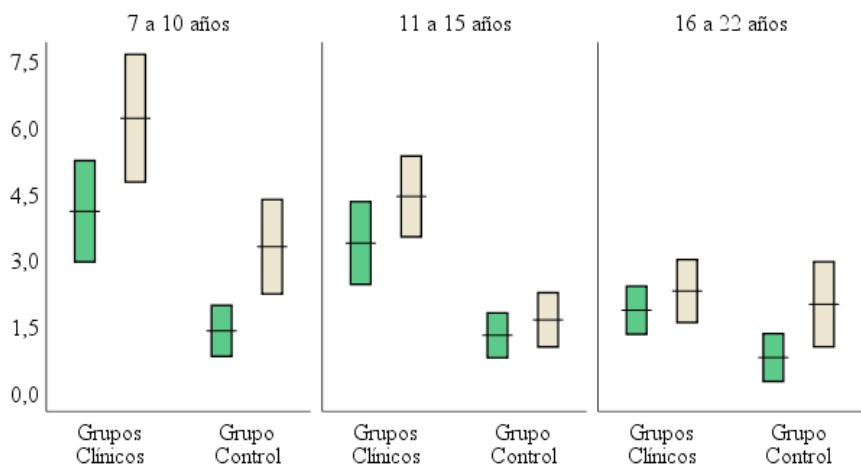
*Comparación de número de errores en ambas versiones del Stroop*



Tarea P:



Tarea C:



Tarea PC:

■ Stroop Tradicional (número de errores cometidos en 45 segundos)

■ Stroop computarizado (número de errores cometidos de 72 ítems mostrados)

Barras de error para el número de errores en ambas versiones (intervalo de confianza al 95%).

Un resumen general del desempeño en las versiones tradicional y computarizada, entre los grupos clínicos y el grupo control, puede consultarse en el anexo 18.

### **Análisis de Correlaciones de las Mediciones Básicas**

Para comenzar, las mediciones de latencias de ambas versiones se halla una magnitud del coeficiente de correlación positiva y alta, tanto en la tarea P ( $r=.740$ ), como en C ( $r=.739$ ) y en PC ( $r=.682$ ), lo cual evidencia una alta consistencia intrasujeto al responder ambas versiones del Stroop. Ello explica cómo aquellos participantes que presentan respuestas lentas en la versión tradicional también tienden a mostrar demoras en la versión computarizada, para las mismas tareas, y viceversa.

Los resultados del análisis de correlación para las mediciones generales de aciertos entre ambas versiones se hallan coeficientes positivos, aunque moderados y bajos, tanto en la tarea P ( $r=.415$ ), como en C ( $r=.258$ ) y en PC ( $r=.311$ ), por lo que puede afirmarse que la cantidad de respuestas correctas registradas en la versión tradicional asimismo tiende a ser consistente con la proporción lograda en la versión computarizada.

Por el contrario, los análisis para el número de errores arrojan coeficientes muy bajos para las correspondientes mediciones de P ( $r=.068$ ), C ( $r=-.011$ ) y PC ( $r=.114$ ). Debe señalarse entonces, de un lado, la marcada diferencia de las mediciones de errores en la comparación entre las versiones, que registra en promedio un mayor número de errores en el Stroop computarizado, en las tres tareas y, de otro lado, la poca consistencia intrasujeto, que en algunos casos presenta incluso una relación inversamente proporcional.

Otros resultados de cocientes entre las mediciones de cada versión, así como sus correspondientes índices de interferencia aparecen en el anexo 17.

### **Análisis de Equivalencia de las Mediciones Básicas**

La prueba de rangos con signo de Wilcoxon revela importantes diferencias al comparar las mediciones básicas de los mismos sujetos en ambas versiones del Stroop. El estadístico de contraste estandarizado indica una diferencia marcada entre los tiempos de

respuesta en P ( $Z=12.749$ ,  $p=.000$ ), como en C ( $Z=5.883$ ,  $p=.000$ ), siendo más altas las latencias de respuesta en la versión computarizada que en la tradicional.

Asimismo, se hallan diferencias significativas para las tareas PC entre ambas versiones ( $Z=9.533$ ,  $p=.000$ ), pero este contraste está basado en rangos positivos, es decir que en la tercera tarea se registra una mayor velocidad en la respuesta de la versión computarizada, al contrario de lo que ocurre en P y C, donde las latencias son más altas en promedio para la versión computarizada.

Seguidamente, el número de respuestas correctas revela diferencias significativas de los mismos sujetos, en ambas versiones del Stroop. El estadístico de contraste estandarizado indica una marcada diferencia en el desempeño en P ( $Z=-8.203$ ,  $p=.000$ ), con un mayor número de aciertos en la versión tradicional que en la computarizada,

Se hallan diferencias significativas entre ambas versiones, para las tareas C ( $Z=11.151$ ,  $p=.000$ ) y PC ( $Z=13.432$ ,  $p=.000$ ), pero este contraste está basado en rangos positivos, es decir que la versión computarizada se destaca en número de aciertos en estas dos tareas.

De la misma manera, respecto del número de errores la comparación intrasujeto corrobora diferencias entre el Stroop tradicional y computarizado. El estadístico de contraste estandarizado indica una diferencia significativa en P ( $Z=10.618$ ,  $p=.000$ ), como en C ( $Z=4.810$ ,  $p=.000$ ) y en PC ( $Z=4.153$ ,  $p=.000$ ), siendo en general más alto el número de errores en la versión computarizada que en la tradicional.

### **Comparación de los Índices de Interferencia**

Tal como se menciona brevemente al describir los índices de cada versión, las comparaciones intersujeto analizadas muestran diferencias significativas por edad y condición clínica en la versión tradicional, pero no así en la versión computarizada que presenta similitudes entre los grupos.

Sumado a ello, dado que en cada caso los cálculos realizados arrojan valores disímiles, que deben interpretarse individualmente acorde a la conceptualización planteada por cada autor proponente, no se considera pertinente comparar directamente los índices de interferencia. Se opta entonces por los hallazgos de las curvas ROC en

ambas versiones, como se presenta en la tabla 23. Otros descriptivos y contrastes por versión pueden consultarse en los anexos 9 y 16.

**Tabla 23**

*Comparación de la capacidad de clasificación de los índices de interferencia*

	Stroop Tradicional			Stroop Computarizado		
	7 a 10	11 a 15	16 a 22	7 a 10	11 a 15	16 a 22
I <sub>D</sub>	.637*	.577	.539	.674*	.489	.437
I <sub>R</sub>	.498	.692*	.644*	.587	.444	.409
I <sub>G</sub>	.521	.726*	.699*	-	-	-
I <sub>REL</sub>	.571	.779**	.700*	-	-	-
I <sub>READ</sub>	-	-	-	.520	.439	.308
I <sub>NAME</sub>	-	-	-	.417	.557	.597
I <sub>V</sub>	.658*	.640*	.582	.396	.662*	.427

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90  
- no calculado*

Sólo en algunos de los casos, los índices de interferencia analizados presentan una adecuada capacidad de clasificación de la condición clínica, sus resultados varían de un grupo de edad a otro y, salvo algunas excepciones, no concuerdan entre versiones.

Tal es el caso del cálculo de I<sub>D</sub> que presenta buenos niveles de sensibilidad y especificidad en el rango de 7 a 10 años, pero no en los grupos de más edad. Un resultado por encima del umbral de 17.50 en la versión tradicional (I<sub>D</sub>=C-PC, calculado por el número de aciertos) y una diferencia superior a 343.50 en la versión computarizada (I<sub>D</sub>=PC-C, calculado por latencias), indican un pobre control de la interferencia bien debido a enlentecimiento en el procesamiento o poca precisión en el cumplimiento de las tareas del test.

Asimismo, el cálculo de I<sub>V</sub> coincide en una adecuada capacidad de clasificación en el rango de 11 a 15 años. En ambas versiones el cálculo se realiza a partir del número de errores registrado, siguiendo el mismo procedimiento (I<sub>V</sub>=PC-[(P+C)/2]), aunque se identifican puntos de corte distintos. Un resultado por encima del umbral de 1.75 en la versión tradicional y de 2.25 en la versión computarizada indican un pobre control de la interferencia, debido a una pobre ejecución de las tareas del test.

### **Análisis de Correlaciones de los Índices de Interferencia**

Los análisis realizados presentan cocientes  $r$  de Pearson muy bajos e incluso negativos (ver anexo 17.4). Los índices de interferencia calculados en una versión parecen no relacionarse con su correspondiente resultado en la otra versión o con cualquier otro cálculo.

Mención al margen, se hallan correlaciones moderadas y altas entre las mediciones básicas de las tareas y sus correspondientes índices. Por ejemplo, el número de errores se correlaciona significativamente con el resultado de  $I_V$ , como es esperable.

### **Análisis de Equivalencia de los Índices de Interferencia**

Las pruebas de rangos con signo de Wilcoxon fueron aplicadas a los índices  $I_R$  e  $I_V$ . No se comparan  $I_G$ ,  $I_{REL}$ ,  $I_{READ}$ ,  $I_{NAME}$ , en razón a que estos índices son calculados para una de las versiones y no la otra, como se menciona anteriormente. Tampoco se incluye  $I_D$ , que comprende magnitudes distintas, aciertos en la versión tradicional y latencias en la computarizada.

El estadístico de contraste estandarizado para  $I_R$  muestra que no existe equivalencia respecto de los resultados de esta cuantificación de interferencia en ambas versiones del Stroop ( $Z=12.094$ ,  $p=.000$ ). Lo mismo ocurre en el análisis de  $I_V$  ( $Z=-2.180$ ,  $p=.029$ ), en la que son más altos los valores arrojados por este índice en la versión computarizada que en la tradicional.

### **Capacidad Discriminante de PC en Ambas Versiones**

Al analizar el puntaje directo de la tercera tarea, como indicador de interferencia, se hallan diferencias en la capacidad de clasificación entre las versiones del Stroop. La tabla 24 presenta la comparación de los valores de AUC para las tres mediciones básicas analizadas en PC.



**Tabla 24***Comparación de la capacidad de clasificación de las mediciones básicas en la tarea PC*

	Stroop Tradicional			Stroop Computarizado		
	7 a 10	11 a 15	16 a 22	7 a 10	11 a 15	16 a 22
Latencias	.674*	.727*	.803**	.796**	.577	.530
Aciertos	.778**	.805**	.823**	.685*	.776**	.498
Errores	.707*	.717*	.670**	.689*	.779**	.499

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

Luego de los análisis de las curvas ROC de las mediciones de latencias, aciertos y errores registrados en la tarea PC, puede decirse que ambas versiones ofrecen aceptables valores de sensibilidad y especificidad en los rangos de edad de 7 a 10 años y 11 a 15 años, mientras que sólo la versión tradicional resulta eficaz en la clasificación de la interferencia para el grupo de 16 a 22 años.

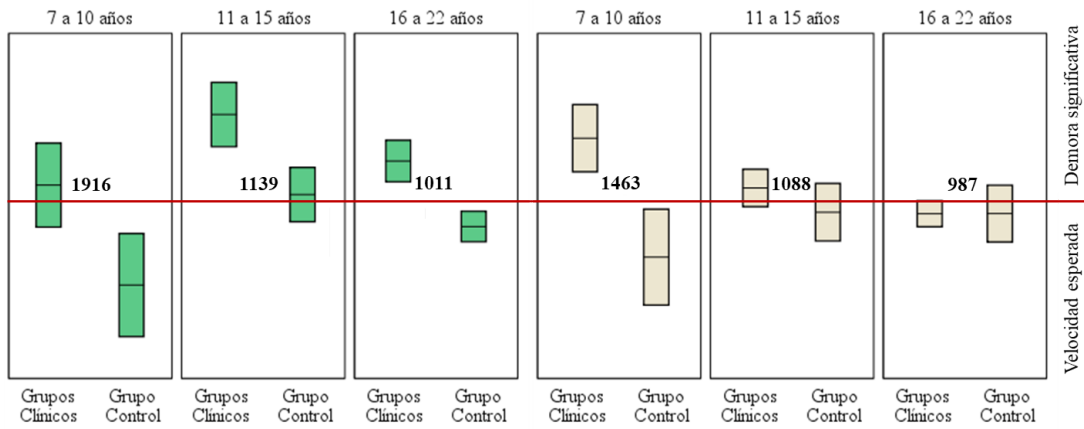
En la página siguiente, la figura 24 ilustra la capacidad de clasificación de las mediciones de latencias, aciertos y errores en la tercera tarea. Puede observarse que tanto los umbrales de las latencias como los de errores disminuyen con la edad y permiten situar los grupos clínicos en un nivel alto de interferencia, vista la demora en la resolución de la tarea de conflicto o la baja eficiencia al resolver el conflicto palabra-color. Particularmente en el rango de 7 a 10 años puede verse cómo el grupo control logra un mejor desempeño respecto de estas dos mediciones.

Los puntos de corte identificados por aciertos se incrementan con la edad y tienden a situar a los grupos clínicos en un nivel más bajo de precisión en la respuesta, en contraste con el grupo control. Para esta medición, la versión tradicional muestra mejores niveles de sensibilidad y especificidad que la computarizada.

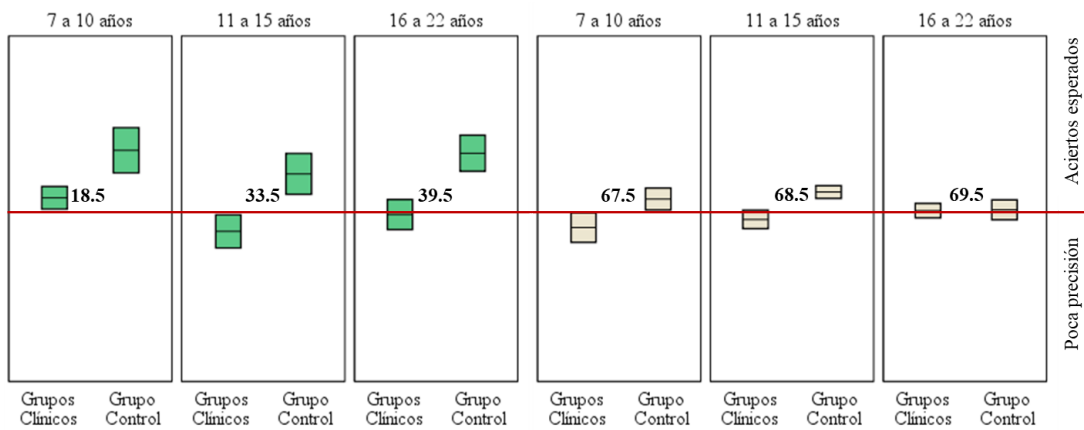
**Figura 24**

*Puntos de corte para los puntajes brutos obtenidos en la tarea PC de ambas versiones*

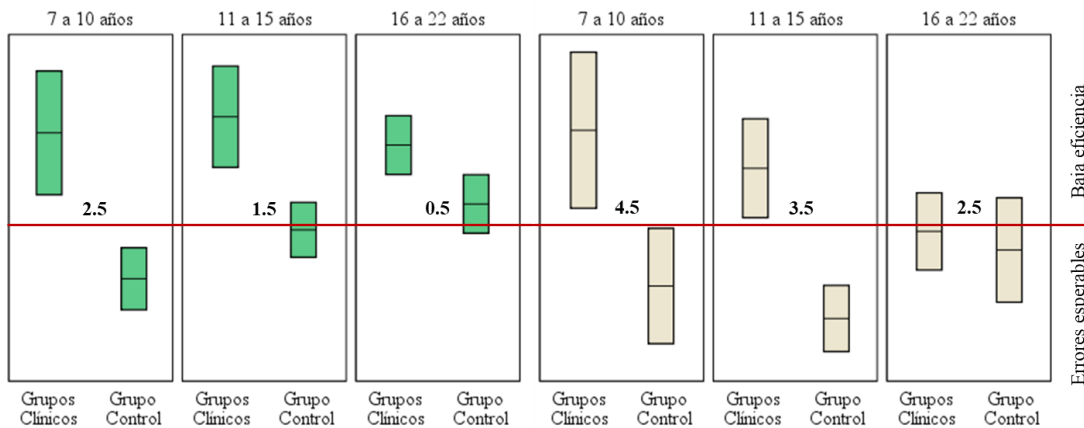
**Latencias**



**Aciertos**



**Errores**



■ *Stroop Tradicional* ■ *Stroop computarizado*

*Barras de error para el número de errores en ambas versiones (intervalo de confianza al 95%).*

## PARTE IV: DISCUSIÓN

### **Diferencias y Semejanzas entre las Versiones del Stroop**

Con el propósito de comparar dos formatos del Test de Colores y Palabras Stroop, se describe el desempeño en una versión tradicional, que registra la respuesta verbal a estímulos enlistados en papel, seguida de una versión computarizada, cuya respuesta es manual sobre el teclado ante los mismos estímulos alternados en la pantalla, para resolver tres tareas: lectura de palabras (P), nominación de color (C) y conflicto palabra-color (PC).

Para esto se examinan tres mediciones básicas (latencias de respuesta, número de aciertos y número de errores), así como una selección de puntajes derivados (índices de interferencia), para el total de la muestra y por factor.

Se incluyen 240 niños y adolescentes, de ambos sexos, organizados por rangos de edad (7 a 10, 11 a 15 y 16 a 22 años) y con diagnóstico de Trastorno por Déficit de Atención con Hiperactividad (subtipos TDAH-DA y TDAH-C) y Trastorno de Aprendizaje (TAPZ), frente a participantes con desarrollo típico (grupo control).

### **Efecto Stroop**

El primer hallazgo a mencionar es la comprobación del Efecto Stroop en ambas versiones aplicadas, fenómeno que se advierte en las tres mediciones. Esto es, todos los participantes presentan una mayor demora en la tarea de conflicto PC, un menor número de aciertos y más errores, en contraste con su desempeño en las tareas P y C, tanto en la versión tradicional como en la computarizada.

Todo ello coincide con la revisión a la literatura científica respecto de la persistencia del Efecto Stroop como uno de los fenómenos cognitivos más fiables y categóricos descubiertos (Stirling, 1979; Dunbar & MacLeod, 1984; Cohen, Dunbar & McClelland, 1990; MacLeod, 1992; Van der Elst, Van Boxtel, Van Breukelen & Jolles, 2006; Meier & Kane, 2013; Duchek et al., 2013; Washburn, 2016).

No obstante, las cuantificaciones de la interferencia mediante los distintos índices dan lugar a la controversia, dada la variabilidad de sus resultados, así como su baja sensibilidad y especificidad.

En segundo lugar, se subrayan diferencias intersujeto en los que existe una mejoría en el desempeño asociada a la edad y una ventaja comparativa del grupo control

sobre los grupos clínicos, con ciertas discrepancias entre éstos, mientras que el desempeño de varones y mujeres tienden a ser similares.

En este sentido, se reconoce que el logro del control de la interferencia ocurre en el curso del neurodesarrollo individual y que el Stroop es un instrumento válido para su estudio, si bien el análisis del efecto de distintos factores sobre el desempeño en el test arroja resultados disímiles (Golden, 1978; MacLeod, 1992; Homack & Riccio, 2004; Hervey Epstein & Curry, 2004; Van Mourik et al., 2005; Bugg, De Losh, Davalos & Davis, 2007, Ludwig et al., 2010; entre otros).

En tercer lugar, las versiones analizadas generan mediciones básicas diferenciables en cada tarea, así como distintos valores para los índices de interferencia calculados.

De ahí que ha debido considerar la afirmación de que diferentes presentaciones de la prueba pueden producir diferentes niveles de interferencia (Logan & Zbrodoff, 1998; Salo, Henik & Robertson 2001; Repovs, 2004; Lansbergen, Kenemans & Engeland, 2007; Borella, Tettamanti & Ribaupierre, 2010; Ludwig et al., 2010), así como la indicación de analizar detenidamente cada uno de los puntajes registrados en el Stroop y las operatorias realizadas a partir de ellos (Jensen & Rohwer, 1966; Cohen, Dunbar & McClelland, 1990; Chafetz & Matthews, 2004).

### **Cuestiones Relativas a las Mediciones Básicas**

Es necesario tener en cuenta que los registros de latencias de respuesta, número de aciertos y número de errores aportan información específica de la ejecución de las tareas del Stroop y del control de interferencia, que pueden interpretarse respecto de los constructos de velocidad en la respuesta, precisión y eficiencia (Jensen & Rohwer, 1966, Golden, 1978; MacLeod, 1992; Chafetz & Matthews, 2004; Miller, 2010; Ludwig et al., 2010).

Los valores atípicos y extremos en el análisis general de las tres mediciones están asociados a la demora en la respuesta (latencias altas), la poca precisión (bajo número de aciertos) y/o el incumplimiento en la ejecución de cada tarea (alto número de errores) que se identifican en los grupos de menor edad y de condición clínica, en ambas versiones.

Latencias y aciertos son las mediciones más comúnmente utilizadas en los reportes científicos. Los trabajos experimentales suelen informar tiempos de reacción para corroborar hipótesis sobre la demora en el procesamiento de la información neutra, congruente e incongruente, que varía según el diseño de la tarea y la modalidad de respuesta utilizada (por ejemplo, Repovs, 2004).

Mientras que los estudios de baremación se hacen a partir de las respuestas correctas, para corroborar la precisión en las tareas de lectura y nominación del color, así como el logro del control de interferencia asociado a la edad (véanse los estudios de Golden, 1978).

En menor medida se hallan trabajos que analizan el tipo y número de los errores, sugiriendo que se trata de puntajes independientes de los dos primeros (Miller, 2010; Sørensen et al., 2013).

En este contexto, los resultados presentados para cada tipo de medición permiten señalar importantes diferencias cuantitativas, tanto en los análisis por los factores como en la comparación entre las versiones del Stroop utilizadas.

### ***Sobre las Latencias***

Debe rechazarse la hipótesis de una mayor demora general en la versión computarizada ( $H_1$ ), ya que esto se cumple para la tarea de lectura, donde todos los participantes se muestran más ágiles en el Stroop tradicional, pero no así para la tarea de nominación de color que tienen resultados variables por edad y condición clínica, ni para la tarea de conflicto palabra-color donde incluso la respuesta manual es más rápida.

Respecto del tiempo medio empleado, puede decirse que la relación entre tareas de la versión tradicional es:  $PC > C > P$ , en el mismo sentido que se presenta en los modelos explicativos del Efecto Stroop (Cohen, Dunbar & McClelland, 1990; Stafford & Gurney, 2007; Van Maanen, Van Rijn & Borst, 2009; Lamers, Roelofs & Rabeling-Keus, 2010).

Dicha relación de desigualdad hace referencia a una menor latencia esperable para la lectura, con un incremento de al menos 200 ms en la nominación del color y hasta la duplicación del tiempo de respuesta frente estímulos incongruentes (Dunbar & MacLeod, 1984; MacLeod, 1991a).

En discordancia con esto, la aplicación computarizada en la presente investigación muestra que la tarea C se realiza incluso más ágilmente que P, en todos los grupos, de modo que se aprecia una relación distinta entre las tareas:  $PC > P > C$ , que ha sido menos documentada.

Así también, Penner et al. (2012) hallan que la modalidad de respuesta verbal es más veloz (614 ms) que la manual (868 ms) en la tarea C, pero que ocurre lo contrario en PC (1279 ms y 1257 ms). En la presente investigación las observaciones coinciden con este mismo patrón para las versiones tradicional y computarizada, vistas las latencias estimadas y las latencias medias de aciertos en C (862 y 922 ms) y en PC (1455 y 1201 ms).

De ahí que se presume un posible efecto de la práctica en la respuesta manual, ya que el tiempo medio de respuesta disminuye una vez iniciada la secuencia de ítems en el Stroop computarizado. En este sentido, Miller (2010) señala que la velocidad de respuesta en el Stroop es variable, un análisis de segmentos de la prueba muestra cambios de ritmo en la respuesta, asociados también a la presentación de errores y autocorrecciones.

Mediante distintos montajes experimentales, los efectos de aprendizaje y práctica pueden observarse tras sucesivas aplicaciones de la prueba (véanse, por ejemplo: Monahan, 2001; Davidson, Zacks & Williams, 2003; Ziaka, Moirou, Vlahou & Protopapas, 2011). Sin embargo, como se advierte aquí, la realización de una única tarea repetitiva puede dar lugar a variaciones en el modo en que se responde.

Adicional a ello, como señala MacLeod (1992), debe tenerse en cuenta que la activación de la tarea manual puede ayudar a eliminar más rápidamente la palabra en el estímulo incongruente. Aunque según Klein (1964), incluso con práctica, la interferencia en la respuesta verbal excede la interferencia en la respuesta manual.

### ***Sobre los Aciertos***

No es posible aceptar rotundamente la hipótesis que el número de respuestas correctas es mayor en la versión tradicional que en la versión computarizada ( $H_2$ ), por cuanto esto se cumple para la tarea P (respectivamente con 83.13 y 69.06 aciertos en promedio), pero ocurre lo contrario en C (54.99 y 69.27), como en PC (31.78 y 68.24), ya que los participantes logran en general más aciertos en la versión computarizada.

Aquí es necesario recordar que la toma de cada versión impone una distinción *a priori*. En el Stroop tradicional se contabilizan las respuestas correctas en un tiempo límite de 45 segundos, mientras que en el Stroop computarizado se registran las respuestas correctas logradas de un total de 72 ítems mostrados, sin restricción de tiempo. En el apartado que revisa el tamaño del conjunto de estímulos presentados, MacLeod (1991) señala que las variantes en los diseños del test pueden afectar el performance, aunque no propiamente la observación del Efecto Stroop.

Como se menciona antes, el uso de mediciones de aciertos se ha privilegiado en los estudios de estandarización de formatos de respuesta verbal (Golden, 1999, 2001, 2007; Martín et al, 2012; Morrow, 2013; Rivera et al., 2015; Rodríguez, Pulido & Pineda, 2017, entre otros), que son aplicados siguiendo el criterio de restringir el tiempo asignado para las tareas (por lo general 30, 45 o 60 segundos). En dichos estudios puede apreciarse la misma relación entre las tareas observadas aquí:  $P > C > PC$ , desde el que se sustenta el Efecto Stroop.

Sin embargo, la interpretación no suele usar la puntuación bruta obtenida, sino que suma un valor de corrección en cada tarea según la edad y luego transforma los datos a puntuaciones  $Z$  o puntuaciones  $T$ , lo que resulta práctico para su interpretación respecto de la curva normal, pero introduce un sesgo a la medición real de desempeño individual.

Esto es, en la prueba tradicional de Golden (1999), las puntuaciones normatizadas se calculan una vez se ha introducido el factor de corrección, alterando la medición original. Para ejemplificar dicho sesgo, la tabla 25 propone el ejercicio de comparar cuatro casos tomados de la muestra de participantes aquí analizada.



**Tabla 25***Ejemplo de puntuaciones de aciertos en las tareas de la versión tradicional del Stroop*

#	Edad	Grupo	P	C	PC
40	9.3	Control	76 + 41	49 + 29	30 + 20
			55 (Normal)	49 (Normal)	55 (Normal)
42	9.4	TAPZ	40 + 41	41 + 29	21 + 20
			37 (Bajo)	43 (Normal)	46 (Normal)
43	9.5	TDAH-DA	61 + 41	46 + 29	17 + 20
			47 (Normal)	47 (Normal)	42 (Normal)
46	9.6	TDAH-C	66 + 41	35 + 29	15 + 20
			49 (Normal)	39 (Bajo)	40 (Normal)

# Código asignado al participante.

Arriba: Puntaje bruto + Factor de corrección propuesto en el manual (Golden, 1999)

Debajo: conversión a Puntuación T (Interpretación)

Puede advertirse que el desempeño se ubica en un rango normal, salvo en dos de las mediciones, pero al comparar los puntajes brutos obtenidos se aprecian amplias diferencias, por ejemplo, en la tarea PC los participantes de los grupos clínicos tienen cerca de la mitad de aciertos. Anticipándose a ello, se ha preferido aquí el registro directo de las distintas mediciones, sin manipulación o transformación de los datos, para hacer una lectura de los valores originales y de las diferencias que podrían arrojar los análisis por factor.

En acuerdo con López Villalobos et al. (2010), el uso de puntajes directos resulta más apropiado para advertir las diferencias individuales. En su análisis, los autores señalan que la medición de respuestas correctas en la tarea PC no sólo es útil, sino que además tiene validez de criterio complementaria para discriminar grupos con diagnóstico de TDAH.

### ***Sobre los Errores***

Con frecuencia se ha señalado que la tarea de conflicto provoca una media más alta de errores, en comparación con las tareas neutras (Peretti, 1973; Van Der Elst et al., 2006; Protopapas et al., 2007; Miller, 2010), tal como se observa en la presente investigación, tanto para la versión tradicional como para la computarizada. Sin embargo,

una vez más la relación de las mediciones es distinta en cada versión: PC>C>P en el Stroop tradicional y PC>P>C en el computarizado.

De modo que se acepta la hipótesis que plantea que el número de errores es mayor en la versión computarizada que en la versión tradicional ( $H_3$ ), lo que se cumple en las tres tareas, aunque en los análisis por factor pueden apreciarse algunas variaciones entre los grupos, particularmente respecto de los factores de edad y condición clínica.

La comisión de errores en el Stroop puede asociarse a la dificultad para inhibir el impulso prepotente, resultado de un control cognitivo reducido, deficitario o inmaduro (Barkley, 1997, 2012; Anderson, 2002; Etchepareborda, 2004). En ambas versiones analizadas aquí, el hallazgo de un mayor número de errores en los grupos clínicos respalda tal observación.

Se ha señalado que realizar con éxito la condición incongruente en el Stroop requiere del papel supervisor de la región frontal medial superior (Stuss et al., 2001; Ravnkilde et al., 2002; Kaplan, Sengor, Gurvit & Guzelis, 2007), así como de la activación adecuada de regiones cinguladas y de circuitos frontotemporales (Pliszka, Liotti & Woldorff, 2000; Hervey et al., 2006; Depue et al., 2010; Mullane et al., 2010; Sjowall, Roth, Lindqvist & Thorell, 2012). Por lo que personas con daño frontal o alteraciones en el neurodesarrollo tienden a presentar un marcado incremento de errores en esta tarea, así como dificultad para advertir el error, lo que se describe como un estilo de respuesta “descuidado”, característico de pacientes con TDAH (Sørensen et al., 2013).

En contraposición, la presencia de autocorrecciones (Miller, 2010), con el consecuente incremento en los tiempos de resolución de la prueba, puede definirse como un modo de respuesta “reflexivo”, o asociado a la función ejecutiva de automonitoreo (Tirapu-Ustárroz, García-Molina, Ríos-Lago & Ardila-Ardila, 2012; Gutiérrez-Martínez, Ramos-Ortega & Vila-Chaves, 2018).

Estudios con adaptaciones computarizadas del Stroop suelen utilizar conjuntamente mediciones de latencias de respuesta y/o de número de errores (Salo Henik & Robertson 2001; King et al., 2007; Christiansen & Oades, 2010; Zarghi et al., 2012; Wilbourn, Kurtz & Kalia, 2012; Wright, 2017, entre otros), señalando que ambas medidas son altamente sensibles a las características individuales (Repovs, 2004; Heine, 2010; Sørensen et al., 2013).

Dichos formatos suelen incluir estímulos neutros, congruentes e incongruentes, aunque no se corresponden necesariamente con formas análogas al Stroop tradicional y, en algunos casos, se omite la tarea de lectura.

### **Examen de la Equivalencia Considerando las Mediciones Básicas**

Podría decirse que las versiones tradicional y computarizada en este estudio corresponden efectivamente a dos formatos paralelos, en virtud al diseño y presentación de los ítems en cada tarea.

En concordancia con Jensen y Rohwer (1966), siendo una prueba sencilla, una auténtica variación se logra con un arreglo diferente de los mismos ítems, con la que se espera obtener una alta fiabilidad.

En este contexto, la estimación empírica de la fiabilidad es igual al cálculo del coeficiente de correlación lineal entre los puntajes observados en ambas formas del test, entendiendo que, de ser ambas formas equivalentes, tendría que obtenerse un coeficiente de correlación elevado (Argibay, 2006; Hernández Sampieri et al., 2014).

Sin embargo, tal como afirma Melia (2000), si bien la correlación entre dos mediciones paralelas iguala el coeficiente de fiabilidad, este resultado no expresa ni prueba en ningún grado la equivalencia entre éstas. Para contrastar la paralelidad, el autor propone satisfacer las siguientes condiciones observables: medias iguales, varianzas iguales y correlaciones positivas.

Respecto del análisis de los coeficientes de correlación entre sus respectivas tareas se encuentran coeficientes positivos y muy significativos, para las latencias de respuesta ( $r > .74$  en P y C,  $r > .68$  en PC), lo que corrobora una alta consistencia intrasujeto ( $H_4$ ). Es decir que aquellos participantes con un buen desempeño en la versión tradicional también tienden a mostrar valores altos en la computarizada para las mismas tareas y viceversa.

Aunque, curiosamente el registro del número de aciertos y de errores presenta una baja consistencia ( $r < .40$ ), lo que no necesariamente implica que estas mediciones son independientes del formato de aplicación del Stroop, pero sí que existe una alta discrepancia en ambas versiones.

De otra parte, al comparar el rango medio de las versiones tradicional y computarizada se hallan diferencias significativas en las tres tareas (pruebas de Wilcoxon:

$p < .01$ ), tanto en la medición de errores, como en aciertos y latencias de respuesta. Por tanto, no es posible confirmar la equivalencia entre formatos de respuesta verbal y manual ( $H_3$ ). En el mismo sentido de Penner et al. (2012), la interpretación de cada resultado y la equiparación entre ambos debe hacerse de forma cautelosa.

De todo ello se desprenden varias consideraciones: a) cada tipo de medición aporta información diferenciable, b) puede resultar equivocado o insuficiente el análisis del desempeño en el Stroop desde una única medición, y c) las distintas mediciones pueden dar lugar a interpretaciones sobre la velocidad en la respuesta, la precisión y la eficiencia, de modo independiente para las tareas de lectura, nominación del color y resolución del conflicto palabra-color, o de modo conjunto sobre el control de la interferencia.

En el contraste entre las versiones aplicadas debe considerarse la existencia de efectos diferenciales de Stroop asociados a la modalidad de la respuesta (Ludwig et al., 2010) y al tipo de puntaje analizado (Jensen, 1965; Chafetz & Matthews, 2004).

Los resultados permiten adherir a la idea que la versión computarizada aporta una medida "más pura" de la capacidad de resistir las respuestas de lectura automática y prepotente (véase la comisión de errores), mientras que la versión tradicional evalúa aún más la capacidad de resistir la interferencia causada por la distracción ante otros estímulos circundantes (véanse las latencias de respuesta y número de aciertos en la tarea de conflicto, respecto del desempeño en los estímulos neutros).

En otras palabras, el formato de respuesta manual requiere superar la interferencia proactiva, además de resistirse a un procesamiento automático.

### ***Hallazgos por Sexo***

La versión tradicional no evidencia discrepancias entre varones y mujeres de la muestra analizada, en ninguna de las tres mediciones básicas. Esto coincide con distintos autores que han observado paridad por sexo en la ejecución de las tareas del Test (Houx, Jolles & Vreeling, 1993; Fuentes et al., 2003; Fernández & Bendisky, 2003; Sisco et al., 2016; Wright, 2017).

Aunque otros reportan leves diferencias individuales en las que, por ejemplo, las mujeres tienden a nombrar el color más ágilmente (Stroop, 1935; MacLeod 1992; Van

Mourik, Osterlaan & Sergeant, 2005; Van Der Elst et al., 2006; Van Mourik et al., 2009). Empero, éstas no tienen significancia estadística o presentan un tamaño del efecto despreciable (Rivera et al., 2015).

En comparación, para la aplicación manual del Test se observan algunas diferencias estadísticamente significativas. Los resultados muestran que las mujeres participantes son más rápidas y presentan más aciertos durante la tarea PC del Stroop computarizado, adicionalmente en las tareas P y C cometen menos errores.

Hallazgo contrario al estudio de Zarghi, Zali, et al. (2012), que informa un desempeño similar por sexo, con un registro levemente más alto de errores congruentes en las mujeres de 9 a 14 años, pero no en los grupos participantes de más edad.

Otros estudios en grupos de edad escolar corroboran la insignificancia en los contrastes entre varones y mujeres, aunque revelan algunas diferencias asociadas al nivel educativo (Van Der Elst et al., 2006; Zalonis et al., 2009; Rivera et al., 2015; Brugnolo et al., 2016), las desventajas socioeconómicas (Conca Binfa & Ibarra González, 2004; Aran Filippetti & Mías, 2009), o la capacidad intelectual y la presencia de déficits neuropsicológicos diversos (Homack & Riccio, 2004), que pueden covariar con el factor sexo.

Por tanto, se acepta que el rendimiento es independiente del sexo ( $H_6$ ), en lo que respecta a la versión tradicional, pero no así en la computarizada, dado que las mediciones de latencias, aciertos y errores muestran diferencias entre varones y mujeres, particularmente en la tarea de conflicto.

### ***Hallazgos por Rango Etario***

Se corrobora la hipótesis sobre los cambios en el rendimiento del Stroop en función de la edad de los participantes ( $H_7$ ). Todos los grupos etarios se diferencian entre sí en las tres tareas, siendo mejor el desempeño general de los participantes de mayor edad, tanto en la versión tradicional como en la computarizada. Dichos resultados concuerdan con estudios precursores que han descrito una mejoría progresiva en la resolución de las tareas del Stroop (Jensen & Rohwer, 1966; Golden, 1978; De Losh, Davalos & Davis, 2007).

Se ha sugerido que cualquiera que pueda leer mostrará un Efecto Stroop robusto (MacLeod, 1991a); sin embargo, la tarea de conflicto palabra-color genera una menor interferencia antes de los 7 años (Comalli, Wapner & Werner 1962), en tanto las habilidades lectoras comúnmente se adquieren y fortalecen con la escolarización.

De ahí que para personas con poco entrenamiento lector o de poca edad suelen implementarse adaptaciones pictóricas del Stroop (Wilbourn, Kurtz & Kalia, 2012), mediante las cuales se reconoce igualmente la presencia de un efecto de interferencia al responder a estímulos duales e incongruentes.

Como se menciona antes, si bien existen cambios en la relación entre las tareas de lectura y nominación de color, en todos los participantes de la muestra se verifica el Efecto Stroop, destacándose la marcada diferencia en el desempeño del grupo de 7 a 10 años, comparativamente más lentos, con menor número de aciertos y más errores en promedio.

Dichos resultados concuerdan con las baremaciones de la prueba tradicional que definen las puntuaciones esperables en cada tarea y sus incrementos por edad (véanse Golden, 1978, 1999; Fernández & Bendisky, 2003; Van Der Elst et al, 2006; Martín et al, 2012; Morrow, 2013; Rivera et al., 2015; Rodríguez, Pulido & Pineda, 2017).

Y, aunque por ahora no se encuentran datos normativos para versiones computarizadas, distintos estudios suelen coincidir en el patrón de un desempeño particularmente disminuido en la tercera tarea, que mejora con la edad, pero que varía entre grupos clínicos y controles (Homack & Riccio, 2004; Penner et al., 2012), como se describe aquí.

Distintos autores sostienen que el decremento en la comisión de errores y la mejoría general en la velocidad de respuesta para la resolución de tareas de conflicto refiere también el desarrollo gradual en otras habilidades atencionales, como focalización, selectividad y atención sostenida (Bugg, De Losh, Davalos & Davis, 2007; Smoleń & Chuderski, 2010; Wilbourn, Kurtz & Kalia, 2012; Wright, 2017), y de las habilidades ejecutivas, como automonitoreo y control inhibitorio, entre otras.

Tales observaciones han sido presentadas en estudios de grupos con déficit atencional (Barkley, 1997, 2012; Stins et al 2005; Mullane et al, 2010, Sørensen et al., 2013), e igualmente en estudios que involucran evaluación de capacidades lingüísticas y

aprendizaje (Lazarus, Ludwig & Aberson, 1984; Heine et al., 2010; Wilbourn, Kurtz & Kalia, 2012).

En el contexto escolar, la capacidad de "permanecer en la tarea" (Golden, 2001) exige la inhibición de distracciones e influencias externas más la recuperación de tales influencias, en el mismo sentido que exige la resolución de las tareas del Stroop. De modo que un sistema atencional menos maduro causa no sólo un enlentecimiento en la respuesta sino una mayor comisión de errores (Wright, 2017).

### ***Hallazgos entre Grupos Clínicos y Grupo Control***

Las revisiones de Homack y Riccio (2004), Van Mourik, Oosterlaan y Sergean (2005) y Lanbergen, Keneman y Van Engeland (2007) reseñan diferencias entre grupos clínicos y controles, al tiempo que también advierten discrepancias entre los formatos de aplicación de la prueba, en donde la modalidad de respuesta verbal tiende a mostrar mayor interferencia, como se corrobora aquí.

Los análisis realizados permiten corroborar que, como era esperable, los grupos clínicos TDAH-DA, TDAH-C y TAPZ presentan un rendimiento disminuido en ambos formatos del Stroop, en comparación con el grupo control o de desarrollo normotípico ( $H_8$ ), aunque igualmente se corroboran variaciones entre las versiones en estudio, vistas sus correspondientes mediciones.

Asimismo, otros estudios se han ocupado de la incidencia de factores personales, como la edad, el sexo o el nivel educativo, sobre el control atencional y la cognición, con resultados disímiles respecto de una versión u otra (Ludwig et al., 2010; Zarghi et al., 2012; Penner et al., 2012), que asimismo verifican sistemáticamente un performance descendido en los grupos clínicos (Lansbergen, Kenemans & Van Engeland, 2007).

Tal como presentan los resultados, en la versión tradicional del Stroop se registran diferencias significativas por condición clínica en las tareas de nominación de color, C y PC, y en la versión computarizada las diferencias son notorias en la tarea de lectura, tanto en las latencias de respuesta como en el número de aciertos. Mientras que para la medición de los errores se hallan diferencias en las tres tareas.

El grupo control muestra un mejor desempeño general en ambas versiones aplicadas, siendo marcada la diferencia con los grupos clínicos para los participantes más

pequeños. Tal como presentan López Villalobos, et al. (2010) y Rubiales, Bakker y Urquijo (2013), pacientes con diagnóstico de TDAH tienen un peor desempeño en el Stroop en comparación con el grupo control. Aunque en dichos estudios los subtipos clínicos analizados no muestran diferencias entre ellos.

La presente investigación reconoce que cada grupo presenta un desempeño característico. En el rango de 7 a 10 años, por ejemplo, el grupo con diagnóstico de TAPZ presenta menos aciertos en las tareas P y C, seguidos por los participantes con TDAH-DA. Mientras que los TDAH-C, si bien tienen un mejor desempeño en estas mismas tareas, tienden a cometer más errores en la tarea PC.

Vale mencionar que la lectura imprecisa o lenta y con esfuerzo, propia del diagnóstico de TAPZ, así como la sintomatología predominante de déficit atencional o de impulsividad en los subtipos inatento y combinado del TDAH (APA, 2002, 2014), pueden resultar útiles para comprender estas variaciones entre las tareas, en ambas versiones.

En concordancia con Van Mourik, Oosterlaan y Sergeant (2005) y con Christiansen y Oades (2010), puede afirmarse que la severidad de los síntomas, así como la presencia de comorbilidades, está asociada con un peor desempeño general en el Stroop. Por su parte, King, et al. (2007) señalan que grupos clínicos de adultos (18 a 45 años) igualmente presentan un efecto de interferencia, aunque en menor proporción.

Los autores presumen que las diferencias se pueden atribuir a la gravedad del diagnóstico y la remisión de los síntomas con el crecimiento. Por ejemplo, su adaptación computarizada registra una media de 841 ms en el grupo control, frente a 999 ms del grupo con TDAH, lo que se asemeja con los valores parciales de 20 a 22 años de la presente muestra, con 912 ms en el grupo control, frente a 947 y 991 ms para los participantes TDAH-DA y TDAH-C, respectivamente.

### **Cuestiones Relativas al Cálculo de los Índices de Interferencia**

Cada uno de los métodos de cuantificación de la interferencia arroja valores diferentes, en virtud a que operan con un cierto tipo de puntuación y puede incluir los resultados de las tres tareas del Stroop o sólo de una parte de éstas; de ahí que para cada índice se propuso identificar la capacidad de discriminación entre grupos clínicos y



controles, establecer los puntos de corte con la sensibilidad y la especificidad más alta, para luego comparar la capacidad discriminativa entre ambas versiones.

### *Hallazgos en $I_D$*

Respecto del índice  $I_D$ , se obtienen valores distintos según la naturaleza del cálculo realizado (por número de aciertos  $I_D=C-PC$  en la versión tradicional y por latencias de respuesta  $I_D=PC-C$  en la versión computarizada). En el primer caso la diferencia de aciertos es de 23.21 en promedio, que se ubica sobre el punto de corte identificado. Y, en el segundo, la diferencia de tiempo es de 205 ms en promedio, muy por debajo del correspondiente umbral.

En términos generales, puede advertirse que el cálculo de  $I_D$  sugiere mayor interferencia en la versión tradicional que en la computarizada, no obstante, al tratarse de magnitudes distintas esta apreciación se hace desde el análisis de su capacidad de clasificación de la condición clínica, que en ambos casos presenta buenos niveles de especificidad, pero baja sensibilidad.

Estos resultados coinciden con autores que señalan cómo el índice  $I_D$  es proporcionalmente más alto para la respuesta verbal. El estudio de Redding & Gerjets (1977), en una pequeña muestra de adultos, se obtiene para la respuesta oral un  $I_D=177$  y para la respuesta manual un valor de  $I_D=98$ .

Y por su parte Penner et al. (2012) reporta igualmente índices diferenciables entre los formatos tradicional y computarizado, tanto para la muestra de adultos ( $I_D=315$  e  $I_D=310$ ) como para la de niños ( $I_D=665$  e  $I_D=390$ ), respectivamente, señalando que existe menos interferencia, pero más facilitación en la respuesta manual.

Desde Stroop (1935) y los estudios que le siguieron se defiende el cálculo de  $I_D$  como una medida adecuada de la interferencia. No obstante, distintas revisiones plantean que este índice puede fallar al capturar adecuadamente la complejidad del Efecto Stroop (Salo, Henik & Robertson, 2001; Lansbergen, Keneman & Van Engeland, 2007; Sisco et al., 2016).

Se le critica a  $I_D$  que tiende a presentar baja fiabilidad en mediciones Test-Retest (Jensen, 1965; Jensen & Rohwer, 1966; Homack & Riccio, 2004) y que no ofrece una adecuada capacidad de discriminación (Van Mourik, Oosterlaan & Sergeant, 2005).

### ***Hallazgos en $I_R$***

El estudio de Lansbergen, Keneman y Van Engeland (2007) sobre  $I_R$  presume que el cálculo debe mostrar el mismo valor por latencias ( $I_R=C/PC$ ) y por aciertos ( $I_R=PC/C$ ), tratándose de dos puntuaciones hechas a un test. En la presente investigación, el índice se obtiene para una u otra versión, de modo que no puede verificarse que esto pudiera ocurrir.

La versión computarizada aquí analizada arroja un cociente de 0.79, mientras que en la versión tradicional se obtiene 0.57, que es levemente más bajo, pero que igualmente ubican a los participantes en un nivel adecuado de control de la interferencia para estímulos incongruentes.

De la misma manera, Lansbergen, Keneman y Van Engeland (2007) sugieren que el incremento de la interferencia es independiente de variables moderadoras, como la edad y el tiempo de respuesta en pacientes con diagnóstico de TDAH y controles.

Se corroboran leves diferencias entre grupos clínicos y controles, y se obtienen puntos de corte distintos para cada grupo de edad. Sin embargo, los resultados del análisis de curvas ROC indican una moderada a baja capacidad de discriminación de  $I_R$ , tanto en la versión tradicional como en la computarizada.

Es importante tener en cuenta que tanto  $I_R$  como  $I_D$  omiten el resultado obtenido en la tarea P, por lo que puede pensarse que son insuficientes en la interpretación del desempeño general en el Stroop.

Sin embargo, según Protopapas, Archonti y Skaloumbakas (2007),  $I_D$  e  $I_R$  resultan adecuados para identificar diferencias significativas entre grupos de lectores hábiles y con problemas en lectura; planteando que tanto la precisión como la velocidad en la respuesta contribuyen a la varianza en la interferencia, dado que la interferencia parece estar determinada por la experiencia general de lectura y no sólo por la habilidad para atender a las tareas del Stroop.

### ***Hallazgos en $I_V$***

Asimismo, en el contexto hispanoamericano resulta poco conocida la revisión científica al puntaje de interferencia  $I_V$  (Caffara et al., 2002; Valentijn et al., 2005; Van Mourik, Oosterlaan & Sergeant, 2005; Van Der Elst et al., 2006, 2008), que ofrece una

alternativa basada en la relación de variación de las tres tareas del Stroop ( $I_V = PC - [(P+C)/2]$ ). Este índice puede calcularse mediante las latencias medias de respuesta o también con el número de errores registrados, como fue analizado aquí.

El cálculo de  $I_V$  permite corroborar como grupos de pacientes con déficit atencional tienden a mostrar mayor interferencia producida por el conflicto (Van Mourik et al., 2009). En este sentido, los resultados del presente estudio muestran importantes diferencias para el cálculo de  $I_V$  en la versión tradicional del Stroop que indica mayor interferencia (2.03) de los grupos clínicos, en comparación con la versión computarizada en donde se observa un mejor nivel de control de la interferencia (0.97), mientras que en el grupo control ocurre lo contrario (0.73 y 1.01), siendo levemente más alto el índice en la respuesta manual.

El estudio de Van Der Elst et al. (2008) ofrece algunas posibilidades interesantes, evitando el uso de  $I_D$  o de  $I_G$ . No obstante, su enfoque basado en la regresión, requiere un procedimiento de tres pasos para convertir las puntuaciones sin procesar en residuos estandarizados, lo que no resulta práctico de usar, por lo que proporcionan tablas de datos normativos para el cálculo  $I_V$ , pero basados en los puntajes del tiempo total usado para completar 100 ítems de cada tarea.

### ***Hallazgos en $I_{READ}$ , $I_{NAME}$ e $I_{REL}$***

Es más infrecuente todavía la mención a los índices  $I_{READ}$  e  $I_{NAME}$  en la literatura científica. Ludwig et al. (2010) informan cocientes hallados en una muestra de adultos jóvenes, ante una adaptación del Stroop con estímulos presentados en bloque e ítem a ítem. En el primer caso ( $I_{READ} = (PC-P)/P$ ) reportan un valor de 0.65 que resulta muy superior a lo observado aquí entre los grupos clínicos (0.15) y controles (0.19). Por el contrario, respecto del segundo cálculo ( $I_{NAME} = (PC-C)/C$ ), coincidimos con dichos autores, ya que tanto grupos clínicos como controles obtienen el mismo valor de 0.30 en promedio.

En acuerdo con Ludwig et al. (2010), deben reconocerse las significativas diferencias en las mediciones asociados con el formato de aplicación del Stroop, entendiendo que las diferencias observadas con la edad e informadas en ambas versiones pueden ser de naturaleza diferente, esto es, implican procesos atencionales diferenciables

y/o activaciones neurocognitivas relativas a la modalidad de respuesta (verbal o manual) y a la exigencia de la tarea (lectura o nominación del color).

Por su parte, el cálculo de  $I_{REL}$ , obtenido únicamente para la versión tradicional, presenta una adecuada capacidad de clasificación de la interferencia, con buenos niveles de sensibilidad y especificidad para los rangos de 11 a 15 años y de 16 a 22 años, pero no así para los más pequeños. En concordancia con Vitkovitch et al. (2002) y Macniven et al. (2008), los grupos clínicos tienden a presentar valores más altos en este índice.

### ***Hallazgos en $I_G$***

Si bien el índice propuesto por Golden (1978, 1999, 2001) es uno de los más ampliamente difundidos a nivel mundial, en tanto que está sustentado en reconocidos estudios de baremación, y tiene como ventaja el reconocimiento del desempeño en las tres tareas ( $I_G = PC - [(P \times C) / (P + C)]$ ), distintos autores sugieren una revisión a su validez de constructo.

Algunos autores señalan errores en la interpretación del Efecto Stroop al implementar  $I_G$  dado el manejo artificioso del factor de corrección en vez de usar las mediciones brutas (Chafetz & Matthews, 2004; Lansbergen, Kenemans y van Engeland, 2007), como se menciona anteriormente.

Al respecto, Sisco et al. (2016) advierten que los resultados de  $I_G$  deben tomarse con precaución ya que no suelen producir diferencias entre grupos. Tal es el caso de los hallazgos en una muestra argentina (Rubiales, Bakker & Urquijo, 2013) en donde inicialmente se señalan diferencias entre los grupos TDAH y Control para las mediciones individuales de las tareas del Stroop, pero una vez que se realiza el cálculo de  $I_G$  desaparecen las diferencias entre éstos y se informa que todos se ubican en un rango normal.

Vale aquí también una crítica a los estudios de Rivera et al. (2015, 2017) y de Rodríguez, Pulido y Pineda (2017), por citar ejemplos de datos normativos recientes en el contexto latinoamericano, en tanto que se asume la validez *de facie* de  $I_G$ , sin cuestionar o indagar otras posibles interpretaciones, aun cuando hallan confiabilidad en mediciones Test-Retest para sus mediciones básicas pero no para el cálculo de  $I_G$ , o presentan resultados disímiles entre países que no son explicados o que muestran incluso valores

muy superiores en adultos mayores que contradicen la observación de un decremento en la capacidad de control inhibitorio asociada al envejecimiento (Houx, Jolles & Vreeling, 1993; Davidson, Zacks & Williams, 2003; Bugg et al., 2007).

### **Examen de la Equivalencia Considerando los Índices Calculados**

Contrario a lo esperado, el análisis de las correlaciones de cada índice entre ambas versiones del Stroop arroja coeficientes bajos, no significativos, e incluso de signo negativo ( $H_0$ ). E igualmente, a partir de las comparaciones de rango medio se confirma que no existe equivalencia entre las versiones, en ninguno de los índices de interferencia calculados.

Los índices analizados presentan una capacidad de discriminación de nivel moderado a bajo, en ambas versiones, con valores de sensibilidad y especificidad que varían por rango etario. Por lo tanto, debe rechazarse igualmente la hipótesis que sugería qué método de cuantificación de la interferencia Stroop habría de mostrar diferencias significativas por condición clínica ( $H_{10}$ ).

Por ejemplo, en la versión tradicional el índice  $I_R$  presenta una adecuada capacidad de discriminación entre la condición clínica y control en los rangos etarios entre 11 a 15 años, seguido por 16 a 22 años, pero no así para los más pequeños en los que se observa mayor dispersión en las mediciones básicas como en estos cálculos de la interferencia. Mientras que este mismo índice, aplicado en la versión computarizada, no resulta concluyente.

Los resultados de la presente investigación sugieren que el cálculo del índice  $I_V$ , seguido de  $I_D$  particularmente en niños entre 7 y 10 años, resultan útiles en la evaluación del Efecto Stroop. Aunque sus resultados no son plenamente convincentes en términos de sensibilidad.

Por otro lado, debe destacarse de manera importante cómo las mediciones básicas de la tarea PC arrojan valores muy significativos en el análisis de curvas ROC, muy superiores a lo que fuera hallado en alguno de los índices analizados.

En especial el registro de aciertos y de errores, en ambas versiones del test, hallan valores buenos de sensibilidad y valores altos de especificidad, al tiempo que la identificación de los puntos de corte por cada rango etario tiende a situar correctamente

los grupos clínicos en niveles altos de interferencia, mientras el grupo control tiende a mostrar mejor capacidad de inhibición de la interferencia. Siguiendo a López-Villalobos et al. (2010), la definición de umbrales para mediciones directas en la tarea de conflicto es útil y tiene validez de criterio complementaria en la discriminación del déficit atencional.

PARTE V:  
CONCLUSIONES

### Conclusiones

Los resultados y contrastes de hipótesis, consecuentes a los planteamientos realizados en esta investigación, derivan en consideraciones de orden teórico, psicométrico y aplicado.

En primer lugar, referidas a la comprensión del constructo sobre la interferencia y los mecanismos cognitivos involucrados. En segundo, relativos a la estabilidad y consistencia de las mediciones intra e interindividuales, al considerar las formas desarrolladas y la obtención de puntajes básicos y derivados. Y, por último, respecto de su administración tanto en el contexto de la evaluación clínica como en la investigación.

Desde estas tres perspectivas se concluye que:

- Las tres mediciones analizadas confirman la persistencia del Efecto Stroop en ambas versiones, lo que coincide con la revisión a la literatura científica.
- Cada una de las versiones en estudio presenta resultados claramente diferenciables. Se corroboran discrepancias intrasujeto e intragrupo para las mediciones de aciertos, de errores y de latencias de respuesta. Dicha variabilidad puede asociarse al logro del control de la interferencia en el curso del neurodesarrollo individual.
- La variación de las mediciones básicas evidencia que las modalidades de respuesta verbal y manual de las versiones en estudio implican diferentes niveles del control atencional de interferencia.
- No puede afirmarse que se trata verdaderamente de dos formas paralelas del Stroop, si bien las versiones tradicional y computarizada de este estudio son similares en su diseño, no aportan mediciones equivalentes y, por tanto, no deben interpretarse sus resultados del mismo modo.
- Es evidente la dificultad que entraña lograr formas estrictamente paralelas, por lo que parece más razonable hablar de formas alternativas, donde se entiende que tales formatos no son por completo equivalentes, sino que constituyen intentos de que lo sean.
- En el análisis del desempeño en el Stroop es importante tener en cuenta la edad y la habilidad de lectura, que pueden afectar el desempeño en la tarea de



interferencia en un sentido inverso a lo esperado; así también se debe el impacto de factores educativos y de desventaja social.

En acuerdo con Van Der Elst et al. (2006), aunque el Efecto Stroop es robusto y se ha observado en numerosas adaptaciones de la prueba, se necesita todavía más investigación con respecto a la comparabilidad e intercambiabilidad de las versiones.

Del mismo modo, la naturaleza problemática de los cálculos de la interferencia (Lanbergen, Keneman & Engeland, 2007) como la heterogeneidad de operaciones e interpretaciones posibles (Scarpina & Tagini, 2017) obligan a un análisis minucioso sobre los resultados del Stroop, cuestionando cuál es la interpretación más adecuada.

A partir de los resultados obtenidos en este trabajo, se concluye que la interpretación del Stroop debería basarse en el análisis de los puntajes brutos en las tres tareas (P, C, y PC), en cuanto a la velocidad, la precisión y eficiencia, ya que aportan información acerca de distintos procesos.

Tal como se ha observado, los resultados obtenidos a partir de los mismos varían significativamente, mientras que los índices que agrupan estas medidas son menos informativos, y su validez y sensibilidad para la detección de alteraciones es cuestionable. Igualmente, se concluye que cualquiera sea el análisis elegido (sobre una o varias de las mediciones e índices) debe considerarse el nivel de maduración neurocognitivo o de desarrollo “esperable” por edad.

Finalmente, resulta esencial realizar ajustes en la interpretación del control de la interferencia según la modalidad de respuesta utilizada (manual o verbal). Al no ser equivalentes, los resultados obtenidos con ambas versiones no son comparables y, por tanto, cualquier estudio que utilice el Stroop como medida de interferencia debe tener en cuenta este aspecto al momento de realizar inferencias y comparar sus resultados con estudios previos.

### **Limitaciones y Prospectiva**

Deben señalarse algunas limitaciones y obstáculos en el proceso de investigación.

En primer lugar, la amplitud y complejidad en la evaluación de los procesos atencionales plantea un primer obstáculo, en tanto que no existe una definición unificada de atención, ni de control inhibitorio; como se reseña en la revisión teórica, existen distintos modelos explicativos sobre el efecto de interferencia Stroop, así como distintas propuestas de cuantificación e interpretación de los resultados de cada tarea y de la prueba en su totalidad. Por lo que es necesario advertir que no se incluyen todas las mediciones posibles, ni todos los modelos de cálculo de la interferencia, sino una selección intencionada de éstos.

Otra limitación está asociada a la elección de la muestra, en tanto que los participantes son de la Ciudad de Buenos Aires y pertenecen a un nivel socioeconómico medio, por lo que se haría necesaria la replicación del estudio en otros grupos para asegurar la validez externa de la investigación. No obstante, se trata de una muestra de 240 participantes en la que se cuidó la conformación de grupos, al haberlos equilibrado por edad y condición clínica, así como el contrabalanceo en la aplicación de las versiones del Stroop, por lo que se presume su validez interna en el diseño implementado.

La investigación realizada es de corte transversal, por lo que las inferencias se limitan al momento de la evaluación y, por consiguiente, son relativas al nivel de desarrollo neurocognitivo alcanzado individualmente, valores que pueden variar con el curso de su maduración y las vivencias propias. La propuesta de estudios longitudinales intrasujeto pueden arrojar información sobre la magnitud de los avances logrados respecto de las habilidades lectoras, de nominación de color y de control de la interferencia, en la perspectiva de las mediciones del Stroop.

Asimismo, se presume la existencia de errores de medición no controlados, bien por las características de los instrumentos usados o por otros factores intervinientes o situacionales que pueden afectar el performance general en el test, en un momento dado.

Finalmente, más allá del aspecto psicotécnico del Stroop, se plantea la necesidad de abrir el debate en una dimensión ecológico-cultural y funcional. Debe tenerse en cuenta que la complejidad de los procesos atencionales y de control de interferencia

evaluados por el test pueden estar influidos por múltiples factores individuales y contextuales.

## REFERENCIAS

### Referencias

- Adam, N. C. & Jarrold, C. J. (2009). Inhibition and the Validity of the Stroop Task for Children with Autism. *Journal of Autism and Developmental Disorders*, 39 (8), 1112-1121. doi: 10.1007/s10803-009-0721-8
- Adleman, N., Menon, V., Blasey, C., White, C., Warsofsky, I., Glover, G., et al. (2002). A Developmental fMRI study of the Stroop color-word task. *NeuroImage*, 16, 61-75. doi: 10.1006/nimg.2001.1046
- Amato, M. P. et al. (2006). The Rao's brief repeatable battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. *Multiple Sclerosis Journal*, 12, 787-793. doi: 10.1177/1352458506070933
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8 (2), 71-82. doi: 10.1076/chin.8.2.71.8724
- APA. (2002). *Manual Diagnóstico y Estadístico de los Trastornos Mentales. DSM IV-TR*. Barcelona: Masson.
- APA. (2014). *Guía de consulta de los criterios diagnósticos del DSM-V*. Arlington: Asociación Americana de Psiquiatría.
- Aran Filippetti, V. & Mías, C. M. (2009). Neuropsicología del Trastorno por Déficit de Atención/Hiperactividad: subtipos predominio Déficit de Atención y predominio Hiperactivo-Impulsivo. *Revista Argentina de Neuropsicología*, 13, 14-28. Recuperado de <https://ri.conicet.gov.ar/handle/11336/56235>
- Argibay, J. C. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. *Subjetividad y Procesos Cognitivos*, 8, 15-33. Recuperado de <https://www.redalyc.org/pdf/3396/339630247002.pdf>
- Artigas-Pallarés, J. (2009). Modelos cognitivos en el trastorno por déficit de atención/hiperactividad. *Revista de Neurología*, 49 (11), 587-593. doi: 10.33588/rn.4911.2009369
- Augustinova, M., Parris, B. A. & Ferrand, L. (2019). The Loci of Stroop Interference and Facilitation Effects with Manual and Vocal Responses. *Frontiers in Psychology*, 10, 1786, 60-72. doi: 10.3389/fpsyg.2019.01786

- Ball, G., Stokes, P. R., Rhodes, R. A., Bose, S. K., Rezek, I., Wink, A. M., et al. (2011). Executive functions and prefrontal cortex: a matter of persistence? *Frontiers in Systems Neuroscience*, 5, 1-13. Recuperado de [www.frontiersin.org](http://www.frontiersin.org). doi:10.3389/fnsys.2011.00003
- Banich, M. T. (2019). The Stroop Effect occurs at multiple points along a cascade of control: evidence from cognitive neuroscience approaches. *Frontiers in Psychology*, 10, 2164. doi: 10.3389/fpsyg.2019.02164
- Banich, M. T., et al. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, 12 (6), 988-1000. doi: 10.1162/08989290051137521
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, 121 (1), 65-94. doi: 10.1037/0033-2909.121.1.65
- Barkley, R. A. (2012). *Executive Functioning and Self-Regulation: Extended Phenotype, Synthesis, and Clinical Implications*. New York: Guilford.
- Barkley, R. A., Fischer, M., Smallish, I., & Fletcher, K. (2002). The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *Journal of Abnormal Psychology*, 111, 279-289. doi: 10.1037/0021-843X.111.2.279
- Barnett, M., Sawyer, J. & Moore, J. (2020). An experimental investigation of the impact of rapport on Stroop test performance. *Applied Neuropsychology: Adult*. Advance online publication. doi: 10.1080/23279095.2020.1828081
- Bausela Herreras, E. & Santos Cela, J. L. (2006). Utilidad del Stroop en la Psicología Clínica. *Avances en Salud Mental Relacional*, 5 (1). Recuperado de <https://psiquiatria.com/bibliopsiquis/utilidad-del-stroop-en-la-psicologia-clinica/>
- Bench, C. J., Frith, C., Grasby, P., Friston, K., Paulesu, E., Frackowiak, R., et al. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia* 31, 907-922. doi: 10.1016/0028-3932(93)90147-R

- Ben-David, B. M., Nguyen, L. L. & Van Lieshout, P. (2011). Stroop effects in persons with traumatic brain injury: selective attention, speed of processing, or color-naming? A meta-analysis. *Journal of the International Neuropsychological Society* 17, 354–363. doi: 10.1017/S135561771000175X
- Bergström, J. (1894). The relation of the interference of the practice effect of an association. *American Journal of Psychology*, 6, 433-442.
- Besner, D., Stolz, J. A., & Boutiuer, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, 4 (2), 221-225. doi: 10.3758/BF03209396
- Best, J. R., Miller, P. H. & Jones, L. L. (2009). Executive Functions after Age 5: Changes and Correlates. *Developmental Review*, 29 (3), 180-200. doi: 10.1016/j.dr.2009.05.002
- Brugnolo, A. et al. (2016). An updated Italian normative dataset for the Stroop color word test (SCWT). *Neurological Sciences*, 37 (3), 365-372. doi: 10.1007/s10072-015-2428-2.
- Bugg, J. M., Delosh, E. L., Davalos, D. B. & Davis, H. P. (2007). Age differences in Stroop interference: contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition*, 14, 155-167. doi: 10.1080/138255891007065
- Bugg, J. M., Jacoby, L. L. & Toth, J. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, 36 (8), 1484-1494. doi: 10.3758/MC.36.8.1484
- Bush, G., Whalen, P., Rosen, B., Jenike, M., McInerney, S. & Rauch, S. (1998). The counting Stroop: an interference task specialized for functional neuroimaging-validation study with functional MRI. *Human Brain Mapping*, 6, 270-282. doi: 10.1002/(SICI)1097-0193
- Busigó, S., Martínó, L., Mercado, C., Colón, L. & Pérez, W. (1998). Validación de la Prueba Neuropsicológica Stroop con la Escala de Inteligencia Wechsler para niños – Revisada de Puerto Rico. *Revista Puertorriqueña de Psicología*, 11, 149-156.  
Recuperado de  
<https://pdfs.semanticscholar.org/d511/fc33e981c4085473bc0cfaa8edb39bdf429.pdf>

- Cabeza, R. & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neurosciences*, 12 (1), 1-47. doi: 10.1162/08989290051137585
- Caffarra, P., Vezzadini, G., Francesca, D., Zonato, F. & Veneri, A. (2002). A short version of the Stroop test: Normative data in an Italian population sample. *Nuova Rivista di Neurologia* 12 (4), 111-115. Recuperado de [https://www.researchgate.net/publication/279903068\\_A\\_short\\_version\\_of\\_the\\_Stroop\\_test\\_Normative\\_data\\_in\\_an\\_Italian\\_population\\_sample](https://www.researchgate.net/publication/279903068_A_short_version_of_the_Stroop_test_Normative_data_in_an_Italian_population_sample)
- Calleja, N. & Hernández-Pozo, M. (2009). Prueba Stroop computarizada de riesgo tabáquico para adolescentes. *Revista Mexicana de Análisis de la Conducta*, 35, 91-107. Recuperado de [https://www.researchgate.net/publication/269784924\\_Prueba\\_Stroop\\_computarizada\\_de\\_riesgo\\_tabaquico\\_para\\_adolescentes](https://www.researchgate.net/publication/269784924_Prueba_Stroop_computarizada_de_riesgo_tabaquico_para_adolescentes)
- Campanholo, K., Romão, M, Machado, M., et al. (2014). Performance of an adult Brazilian sample on the Trail Making Test and Stroop Test. *Dementia & Neuropsychologia*, 8, 26-31. doi: 10.1590/S1980-57642014DN81000005
- Cardo, E., Nevot, A., Redondo, M., Melero, A., de Azua, B., García-De la Banda, G., et al. (2010). Trastorno por déficit de atención/hiperactividad: ¿un patrón evolutivo? *Revista de Neurología*, 50 (3), 143-147. doi: 10.33588/rn.50S03.2010005
- Castellanos, F. X. & Acosta, M. T. (2002). El síndrome de déficit de atención con hiperactividad como expresión de un trastorno funcional orgánico. *Revista de Neurología*, 35 (1), 1-11. doi: 10.33588/rn.3501.2002231
- Castellanos, F. X. & Acosta, M. T. (2004). Neuroanatomía del trastorno por déficit de atención con hiperactividad. *Revista de Neurología*, 38 (1), 131-136. doi: 10.33588/rn.38S1.2004086
- Catell, J. (1886). The time taken up by cerebral operations. *Mind*, 11 (41), 220-242. Recuperado de <https://www.jstor.org/stable/2247078?seq=1>
- Chafetz, M. D. & Matthews, L. H. (2004). A New interference score for the Stroop test. *Archives of Clinical Neuropsychology*, 19, 555-567. doi: 10.1016/j.acn.2003.08.004



- Christiansen, H. & Oades, R. (2010). Negative priming within a Stroop task in children and adolescents with attention-deficit hyperactivity disorder, their siblings, and independent controls. *Journal of Attention Disorders*, *13*, 497-504. doi: 10.1177/1087054708325974
- Coderre, E., Filipi, C., Newhouse, P. & Dumas, J. (2008). The Stroop Effect in Kana and Kanji Scripts in Native Japanese Speakers: An fMRI Study. *Brain Lang*, *107* (2), 124–132. doi:10.1016/j.bandl.2008.01.011.
- Cohen, J., Dunbar, K. & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop Effect. *Psychological Review*, *97* (3), 332-361. doi: 10.1037/0033-295X.97.3.332
- Cohen, R. & Swerdlik, M. (2001). *Pruebas y Evaluación Psicológica*. México: McGraw-Hill.
- Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *100*, 47-53. doi: 10.1080/00221325.1962.10533572
- Compton, R. & Banich, M. (2003). Paying attention to emotion: An fMRI investigation of cognitive and emotional Stroop tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *3* (2), 81-96. doi: 10.3758/CABN.3.2.81
- Conca Binfá, 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*. Universidad de Chile, Carrera de Psicología, Facultad de Ciencias Sociales. Recuperado de <http://repositorio.uchile.cl/bitstream/handle/2250/113385/cs39-ibarram53.pdf?sequence=1>
- Córdova, C., Karnikowski, M., Pandossio, J. & Nóbrega, O. (2008). Caracterização de respostas comportamentais para o teste de Stroop computadorizado - TESTINPACS. *Neurociências*, *4* (2), 75-9. Recuperado de [https://www.researchgate.net/publication/291045892\\_Caracterizacao\\_de\\_respostas\\_comportamentais\\_para\\_o\\_teste\\_de\\_stroop\\_computadorizado\\_-\\_testinpacs](https://www.researchgate.net/publication/291045892_Caracterizacao_de_respostas_comportamentais_para_o_teste_de_stroop_computadorizado_-_testinpacs)

- Cuevas, K., Calkins, S. & Bell, M. (2016). To Stroop or not to Stroop: Sex-related differences in brain-behavior associations during early childhood. *Psychophysiology*, 53, 30-40. doi: 10.1111/psyp.12464
- Davidson, D. J., Zacks, R. T. & Williams, C. C. (2003). Stroop Interference, Practice, and Aging. *Neuropsychology, development, and cognition*, 10 (2), 85-98. doi: 10.1076/anec.10.2.85.14463
- Davidson, M. C., Amso, D., Anderson, L. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44 (11), 2037–2078. doi: 10.1016/j.neuropsychologia.2006.02.006
- Denney, D. R. & Lynch, S.G. (2009) The impact of multiple sclerosis on patients' performance on the Stroop Test: processing speed versus interference. *Journal of the International Neuropsychological Society*, 15 (3), 451-458. doi: 10.1017/S1355617709090730
- Depue, B. E., Burgess, G. C., Willcutt, E. G., Bidwell, L. C., Ruzic, L. & Banich, M. T. (2010). Symptom-correlated brain regions in young adults with combined-type ADHD: Their organization, variability, and relation to behavioral performance. *Psychiatry Research*, 182 (2), 96-102. doi: 10.1016/j.psychresns.2009.11.011
- Dos Santos Assef, E. C., Seabra Capovilla, A. G. & Capovilla, F. C. (2007). Avaliação do controle inibitório em TDAH por meio do Teste de Geração Semântica. *Psicologia: Teoria e Prática*, 9 (1), 61-74. Recuperado de <http://pepsic.bvsalud.org/pdf/ptp/v9n1/v9n1a05.pdf>
- Duchek, J., Balota, J., Thomas, J., Snyder, A., Rich, P., Benzinger, T., Fangan, A., Holtzman, D., Morris, J. & Ances, B. (2013). Relationship between Stroop performance and resting state functional connectivity normal older adults. *Neuropsychology*, 27 (5), 516-528. doi: 10.1037/a0033402
- Dunbar, K. & MacLeod, M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, 10 (5), 622-639. doi: 10.1037/0096-1523.10.5.622
- Durgin, F. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review*, 7 (1), 121-125. doi: 10.3758/BF03210730

- Erdodi, L., Sagar, S., Seke, K., Zuccato, B., Schwartz, E. & Roth, R. (2018). The Stroop Test as a measure of performance validity in adults clinically referred for neuropsychological assessment. *Psychological Assessment*, 30 (6), 755-766. doi: 10.1037/pas0000525.
- Etchepareborda, M. C. & Comas, F. (1994). Citado por Etchepareborda, M. C. (2001). *Evaluación del sistema atencional a través de la batería computarizada MCC-94*. Primeras Jornadas Uruguayas de Neuropsicología Infantil. Montevideo.
- Etchepareborda, M. C. (2000). Evaluación y clasificación del trastorno por déficit de atención con hiperactividad. *Revista de neurología clínica. Revista de Neurología*, 1, 171-180. doi: 10.33588/rn.101.2100028
- Etchepareborda, M. C. (2004). Bases experimentales para la evaluación de la atención en el trastorno por déficit de atención con hiperactividad. *Revista de Neurología*, 38 (1), 137-144. doi: 10.33588/rn.38S1.2004042
- Etchepareborda, M. C., Paiva-Barón, H. & Abad, L. (2009). Ventajas de las baterías de exploración neuropsicológica en el trastorno por déficit de atención/hiperactividad. *Revista de Neurología*, 48 (2), 89-93. doi: 10.33588/rn.48S02.2009073
- Faraone, S. V. & Biederman, J. (1998). Neurobiology of attention-deficit hyperactivity disorder. *Biological Psychiatry Journal*, 44, 951-958. doi: 10.1016/S0006-3223(98)00240-6
- Fernández, A. L. & Bendersky, V. (2004). Valores normativos para el Test de Stroop en una muestra de hispano-parlantes. *Psicodiagnosticar*, 13, 63-72. Recuperado de <https://www.researchgate.net/publication/280494346>  
\_Valores\_normativos\_del\_Test\_de\_Stroop\_para\_una\_muestra\_de\_adultos\_hispano-parlantes
- Fernández, A. L. (2014). Neuropsicología de la atención. Conceptos, alteraciones y evaluación. *Revista Argentina de Neuropsicología* 25, 1-28. Recuperado de [https://www.researchgate.net/publication/273970215\\_Neuropsicologia\\_de\\_la\\_atencion\\_Conceptos\\_alteraciones\\_y\\_evaluacion](https://www.researchgate.net/publication/273970215_Neuropsicologia_de_la_atencion_Conceptos_alteraciones_y_evaluacion)

- Fernández-Duque, D. & Posner, M.I. (2001). Brain imaging of attentional networks in normal and pathological states. *Journal of Experimental and Clinical Neuropsychology*, 23,74-93. doi: 10.1076/jcen.23.1.74.1217
- Ferrand, L., et al. (2019). Stroop interference is a composite phenomenon: Evidence from distinct developmental trajectories of its components. *Developmental Science*, 23 (2). doi: 10.1111/desc.12899
- Flores Lázaro, J. C., Ostrosky Shejet, F. & Lozano Gutiérrez, A. (2014). Bateria Neuropsicológica de Funciones Ejecutivas y Lóbulos Frontales. Manual de aplicación. 2 Ed. México: Manual Moderno.
- Freud, S. (1895). *Proyecto de una Psicología para neurólogos*. En: Obras completas. (1997). Buenos Aires: Biblioteca Nueva / Editorial Losada.
- Fuentes, L., González, C., Estévez, A., Carranza, J., Daza, M., Galián, M. & Álvarez, D. (2003). Sensitivity of certain standardized test to executive attention functioning in seven-year-old children. *Electronic Journal of Research Educational Psychology*, 1 (2), 23-36. doi:
- Fuster, J. M. (1999). Synopsis of function and dysfunction of the frontal lobe. *Acta Psychiatrica Scandinavica*, 99, 51-57. doi: 10.1111/j.1600-0447.1999.tb05983.x
- Fuster, J. M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology*, 31, 373 – 385. doi: 10.1023/a:1024190429920
- Fuster, J. M. (2008). *The prefrontal cortex*. Amsterdam: Academic Press – Elsevier.
- Gantiva, C., Toro, R., Ballesteros, L., Salcedo, D., Vargas, R., Gutiérrez, Y., et al. (2012). Atención y motivación ante estímulos afectivos en personas con ansiedad social. *Revista CES Psicología*, 4 (2), 79-87. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=4174127>
- Gao, Q., Chen, Z. & Russell, P. (2007). Working memory load and the Stroop interference effect. *New Zealand Journal of Psychology*, 36 (3), 146-153. Recuperado de [https://www.psychology.org.nz/wp-content/uploads/Gao-Chen-Russell\\_pg146.pdf](https://www.psychology.org.nz/wp-content/uploads/Gao-Chen-Russell_pg146.pdf)
- García Valls, J. M., Blasco González, P., Gil Collados, A., Martínez Jiménez, A. & Laguía Moreno, C. (2007). Validación y concepto de esquizofrenia deficitaria mediante el rendimiento en tareas ejecutivas utilizando el test de Stroop.

- Psiquiatria.com*, 11 (1). Recuperado de <http://www.psiquiatria.com/revistas/index.php/psiquiatriacom/article/viewFile/244/229/>
- García-Molina, A., Tirapu-Ustárrroz, J., Luna-Lario, P. & Duque, P. (2010). ¿Son lo mismo inteligencia y funciones ejecutivas? *Revista de Neurología*, 50, 738-746. doi: 10.33588/rn.5012.2009713
- García-Oqueta, M. I. (2001). Mecanismos atencionales y síndromes neuropsicológicos. *Revista de Neurología*, 32, 463-467. doi: 10.33588/rn.3205.2000169
- Gardner, R. W., Holzman, P., Klein, G., Linton, H. & Spence, D. (1959). Cognitive control: a study of individual consistencies in cognitive behaviour. *Psychological Issues*, 1 (4), 1-186. Recuperado de <https://psycnet.apa.org/record/1961-02266-001>
- Giménez-Donoso, S., Treen-Calvo, D., Setién-Suero, E., Tirapu-Ustárrroz, J., Crespo-Facorro, B. & Ayesa-Arriola, R. (2018). Evaluación del control inhibitorio verbal con el test de Stroop en primeros episodios de psicosis con experiencia de trauma en la infancia. *Revista de Neurología*, 67, 115-120. doi: 10.33588/rn.6704.2017499
- Golden, C. J. & Freshwater, S. (2002). *Stroop Color and Word Test*. Wood Dale: Stoelting Company.
- Golden, C. J. (1976). Identification of brain disorders by the Stroop color and word test. *Journal of Clinical Psychology*, 32 (3), 654-658. doi: 10.1002/1097-4679(197607)32:3<654::aid-jclp2270320336>3.0.co;2-z
- Golden, C. J. (1978). *Stroop color and word test*. Wood Dale: Stoelting Company.
- Golden, C. J. (1999). *Stroop. Test de Colores y Palabras*. Manual. (2 Ed.). Madrid: TEA.
- Golden, C. J. (2001). *Stroop. Test de Colores y Palabras*. Manual. (3 Ed.). Madrid: TEA.
- Golden, C. J. (2007). *Stroop. Test de Colores y Palabras*. Manual. (5 Ed.). Madrid: TEA.
- Golden, C. J. (2020). *Stroop. Test de colores y palabras – Edición revisada* (Ruiz-Fernández, B., Luque, T. & Sánchez-Sánchez, F., adaptadores). Madrid: TEA.
- Gualtieri, C. T. & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, 21, 623-643. doi: 10.1016/j.acn.2006.05.007

- Gutiérrez-Martínez, F., Ramos-Ortega, M. & Vila-Chaves, J. (2018). Eficacia ejecutiva en tareas de interferencia tipo Stroop. Estudio de validación de una versión numérica y manual (CANUM). *Anales de psicología*, 34 (1), 184-196. doi: 10.6018/analesps.34.1.263431
- Hanslmayr, s., Pastötter, B., Bäuml, K., Gruber, S., Wimber, M. & Klimesch, W. (2008). The electrophysiological Dynamics of Interference during the Stroop Task. *Journal of Cognitive Neuroscience*, 20 (2), 215-225. doi: 10.1162/jocn.2008.20020
- Harrison, B. J. (2005). Functional connectivity during Stroop task performance. *NeuroImage*, 24, 181-191. doi: 10.1016/j.neuroimage.2004.08.033
- Heine, A., Tamm, S., De Smedt, B., Schneider, M., Thaler, V., Torbeyns, J., et al. (2010). The numerical Stroop effect in primary school children: a comparison of low, normal, and high achievers. *Child Neuropsychology*, 16, 461–477. doi: 10.1080/09297041003689780
- Henik, A. & Salo, R. (2004). Schizophrenia and the Stroop Effect. *Behavioral and Cognitive Neuroscience Reviews*, 3 (1), 42-59. doi: 10.1177/1534582304263252
- Hepp, H., Maier, S., Hermle, L. & Spitzer, M. (1996). The Stroop effect in Schizophrenic patients. *Schizophrenic Research*, 22, 187-195. doi: 10.1016/S0920-9964(96)00080-1
- Herd, S. A., Banich, M. T. & O'Reilly, R. (2006). Neural mechanisms of cognitive control: an integrative model of Stroop task performance and fMRI data. *Journal of Cognitive Neuroscience*, 18 (1), 22-32. doi: 10.1162/089892906775250012
- Hernández Sampieri, R., Fernández Collado, C. & Baptista, L. (2014). *Metodología de la Investigación*. 6 ed. México: McGraw-Hill.
- Hershman, R. & Henik, A. (2020) Pupillometric contributions to deciphering Stroop conflicts. *Memory & Cognition*, 48, 325–333. doi: 10.3758/s13421-019-00971-z
- Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Arnold, L. E., Conners, C. K., et al. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology*, 12 (2), 125-140. doi: 10.1080/09297040500499081

- Homack, S. & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology, 19*, 725-743. doi: 10.1016/j.acn.2003.09.003
- Horn, W., Lukesch, H., Mayrhofer, S. & Kormann, A. (1969). *Prüfsystem für Schul- und Bildungsberatung*. Manual. Göttingen: Hogrefe Verlag.
- Houx, P., Jolles, J. & Vreeling, F. (1993). Stroop Interference: Aging Effects Assesed with Stroop Color-Word Test. *Experimental Aging Research, 19*, 209-224. doi:
- Huang, Y., Su, L. & Ma, Q. (2020). The Stroop effect: An activation likelihood estimation meta-analysis in healthy young adults. *Neuroscience Letters, 716* (18). doi: 10.1016/j.neulet.2019.134683
- Hunt, R. & Ellis, H. (2004). *Fundamentos de Psicología Cognitiva*. Bogotá: Manual Moderno.
- Ikeda, Y., Okuzumi, H. & Kokubun, M. (2014). Stroop-Like Interference in the Real Animal Size Test and the Pictorial Animal Size Test in 5- to 12-Year-Old Children and Young Adults. *Journal Applied Neuropsychology: Child, 3* (2), 115-125. Doi: 10.1080/21622965.2012.725185
- Jaensch, E. (1929). *Grundformen menschlichen seins*. Berlin: Otto Elsner Verlagsgesellschaft. Disponible en <https://catalog.hathitrust.org/Record/000618118>
- James, W. (1909). *Principios de la psicología*. Barnés, D. (Trad.) Recuperado de <https://archive.org/details/HDeposito0051210/page/n4>
- Jensen, A. R. & Rohwer, W. D. (1966). The Stroop Color-Word Test: A review. *Acta Psychologica, 25*, 36-93. doi: 10.1016/0001-6918(66)90004-7
- Jensen, A. R. (1965). Scoring the Stroop Test. *Acta Psychologica, 24*, 398-408. Recuperado de <http://arthurjensen.net/wp-content/uploads/2014/06/Scoring-the-Stroop-Test-1965-by-Arthur-Robert-Jensen.pdf>
- Jiménez, J. E. (2012). Trastorno por déficit de atención con o sin hiperactividad (TDAH): Prevalencia y evaluación de las Funciones Ejecutivas. *Revista Europea de Psicología de la Educación, 5* (1), 5-11. Recuperado de <https://www.redalyc.org/pdf/1293/129324775001.pdf>



- Juvina, I., Taatgen, N. A., & Dickison, D. (2007). *Cognitive Control as Alternation of Activation and Suppression in the Stroop Task*. Disertación presentada en el Proceedings of the 29th Annual Conference of the Cognitive Science Society, Nashville. Recuperado de <http://act-r.psy.cmu.edu/wordpress/wp-content/uploads/2012/12/719p1133.pdf>
- Kahneman, D. (1973). *Attention and Effort*. New Jersey: Prentice-Hall International. Capítulo 6 (Attention to Attributes), Capítulo 9 (Speeded responses to simultaneous and to immediately successive signals) y Capítulo 10 (Attention and task interference). Recuperado de <https://pdfs.semanticscholar.org/a07f/fad799cffee3ef6a2b33f4a56bffc5b747d.pdf>
- Kang, C., Lee, G. J., Yi, D., McPherson, S., Rogers, S., Tingus, K., et al. (2013). Normative data for healthy older adults and an abbreviated version of the Stroop test. *The Clinical Neuropsychologist*, 27, 276–289. doi: 10.1080/13854046.2012.742930
- Kaplan, B. G., Sengor, N. S., Gurvit, H. & Guzelis, C. (2007). Modelling the Stroop effect: A connectionist approach. *Neurocomputing*, 70, 1414-1423. doi: 10.1016/j.neucom.2006.05.009
- Kaufmann, L., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *NeuroImage*, 25 (3). 888–898. doi: 10.1016/j.neuroimage.2004.12.041
- King, J., Colla, M., Brass, M., & Von Cramon, D. (2007). Inefficient cognitive control in adult ADHD: evidence from trial-by-trial Stroop test and cued task switching performance. *Behavioral and Brain Functions*. 26 p. doi: 10.1186/1744-9081-3-42
- Kinoshita, S., & Mills, L. (2020). Phonological encoding in the oral but not manual Stroop task: Evidence for the role of a speech production process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46 (8), 1494-1504. doi: 10.1037/xlm0000828
- Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology*, 77, 576–588. doi: 10.2307/1420768



- Koch, C. & Roid, G. (2012). Nonverbal Stroop Card Sorting Test (NSCST). Wood Dale, IL, USA, Stoelting Company. Disponible en <https://www.stoeltingco.com/nonverbal-stroop-card-sorting-test-nscst-kit-2337.html>
- Koch, C., Gobell, J. & Roid, G. H. (1999). Exploring individual differences in Stroop processing with cluster analysis. *Psychology*, 10 (25). Recuperado de <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?10.025>
- Koizumi, A., Ikeda, K., Tanaka, A. & Takano, Y. (2007). Emotional Stroop Task with Facial Expressions and Emotional Words. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29, 1792. Recuperado de <http://csjarchive.cogsci.rpi.edu/Proceedings/2007/docs/p1792.pdf>
- Lamers, M., Roelofs, A. & Rabeling-Keus, I. (2010). Selective attention and response set in the Stroop task. *Memory & Cognition*, 38 (7), 893-904. doi: 10.3758/MC.38.7.893
- Langleben, D., Monterosso, J., Elman, I., Ash, B., Krikorian, G., & Austin, G. (2006). Effect of methylphenidate on Stroop Color-Word task performance in children with attention deficit hyperactivity disorder. *Psychiatry Research*, 141 (3), 315-320. doi: 10.1016/j.psychres.2005.09.007
- Lansbergen, M. M., Kenemans, J. L. & Van Engeland, H. (2007). Stroop Interference and Attention-Deficit/Hyperactivity Disorder: A Review and Meta-Analysis. *Neuropsychology*, 21 (2), 251-262. doi: 10.1037/0894-4105.21.2.251
- Lansbergen, M. M., Van Hell, E. & Kenemans, J. L. (2007). Impulsivity and conflict in the Stroop task. An ERP study. *Journal of Psychophysiology*, 21, 33-50. doi: 10.1027/0269-8803.21.1.33
- Larsen, R. J., Mercer, K. A. & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, 6 (1), 62-72. doi: 10.1037/1528-3542.6.1.62
- Lazarus, P. J., Ludwig, R. P., & Aberson, B. (1984). Stroop Color-Word Test: A screening measure of selective attention to differentiate LD from non-LD children. *Psychology in the Schools*, 21 (1), 53-60. doi: 10.1002/1520-6807(198401)

- Leao, A., Antunes, I., Martins, P., Almeida, A., Machado-Piñeiro, W. (2014). Order of stimulus presentation modulates interference in Stroop matching tasks: a reaction time study. *Revista Psicologia: Teoria e Prática*, 16 (3), 143-154. doi: 10.15348/1980-6906.
- León, O. & Montero, I. (1993). *Diseño de investigaciones. Introducción a la lógica de la investigación en psicología y educación*. Madrid: McGraw-Hill.
- León, O. & Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2×2. *Psicothema*, 13 (1), 159-165. Recuperado de <http://www.psicothema.com/pdf/428.pdf>
- León-Carrión, J., García-Orza, J. & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *The International Journal of Neuroscience*, 114, 1291-1311. doi: 10.1080/00207450490476066
- Leung, H., Skudlaski, P., Gatenby, J., Peterson, B. & Gore, J. (2000). An Event-related Functional MRI of the Stroop color Word Interference Task. *Cerebral Cortex*, 10, 552-560. doi: 10.1006/nimg.2001.1046
- Levin, Y. & Tzelgov, J. (2014). Conflict components of the Stroop effect and their “control”. *Frontiers in psychology*, 5, 1-5. doi: 10.3389/fpsyg.2014.00463
- Liefooghe, B., Hughes, S., Schmidt, J. R., & De Houwer, J. (2020). Stroop-like effects of derived stimulus–stimulus relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46 (2), 327–349. doi: 10.1037/xlm0000724
- Liotti, M., Woldorff, M., Pérez, R. & Mayberg, H. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38, 701-711. doi: 10.1016/s0028-3932(99)00106-2
- Logan, G. D. & Zbrodoff, N. J. (1998). Stroop-type interference: congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance*, 24 (3), 978-992. doi: 10.1037/0096-1523.24.3.978
- Long, C. & Prat, C. (2002). Working memory and Stroop interference: an individual differences investigation. *Memory & Cognition*, 30 (2), 294-301. doi: 10.3758/bf03195290

- López-Villalobos, J. A., Andrés-de Llano, J. M., Delgado Sánchez-Mateos, J., Rodríguez-Molinero, L., Garrido-Redondo, M., Martínez-Rivera, M. T., et al. (2014). Validez de un modelo reducido de ítems del DSM-IV según respuesta de padres y profesores en el diagnóstico del Trastorno por Déficit de Atención con Hiperactividad Combinado. *Anales de Psicología*, 30 (3), 898-907. doi: 10.6018/analesps.30.3.152561
- López-Villalobos, J. A., Serrano-Pintado, I., Andrés De Llano, J. M., Delgado Sánchez-Mateos, J., Alberola-López, S. & Sánchez-Azón, M. I. (2010). Utilidad del test de Stroop en el trastorno por déficit de atención/hiperactividad. *Revista de Neurología*, 50, 333-340. doi: 10.33588/rn.5006.2009418
- Lorentz, E., Mckibben, T., Ekstrand, C., Gould, L., Anton, K. & Borowsky, R. (2016). Disentangling genuine semantic stroop effects in reading from contingency effects: on the need for two neutral baselines. *Frontiers in Psychology*, 7, 1-32. doi:10.3389/fpsyg.2016.00386
- Lovett, M. C. (2005). A strategy-based interpretation of Stroop. *Cognitive Science*, 29, 493-524. doi: 10.1207/s15516709cog0000\_24
- Ludwig, C., Borella, E., Tettamanti, M. & de Ribaupierre, A. (2010). Adult age differences in the color Stroop test: a comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics*, 51, 135-142. doi: 10.1016/j.archger.2009.09.040 doi: 10.1016/j.archger.2009.09.040
- Luria, A. R. (1970). The functional organization of the human brain. *Scientific American*, 222, 66-79. Disponible en <https://www.jstor.org/stable/24925755?seq=1>
- Luria, A. R. (1974). *El cerebro en acción*. Barcelona: Fontanella.
- Luria, A. R. (1995). *Las funciones corticales superiores del hombre*. México: Fontamara.
- MacLeod, C. M. & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4 (10), 383-390. doi: 10.1016/S1364-6613(00)01530-8
- MacLeod, C. M. & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4 (10), 383-391. doi: 10.1093/mind/os-XI.41.63

- MacLeod, C. M. (1986). [A replication of the classic Stroop effect]. Manuscrito no publicado. Citado por MacLeod, C. M. (1991a). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychological Bulletin*, *109* (2), 163-203. doi: 10.1037/0033-2909.109.2.163
- MacLeod, C. M. (1991a). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychological Bulletin*, *109* (2), 163-203. doi: 10.1037/0033-2909.109.2.163
- MacLeod, C. M. (1991b). The Stroop Task in Cognitive Research. En: A. Wenzel & D. C. Rubin (Eds.). *Cognitive methods and their application to clinical research*. (pp. 17-40). Washington: American Psychological Association. Recuperado de <https://psycnet.apa.org/record/2004-19025-002>
- MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology*, *121* (1), 12-14. doi: 10.1037/0096-3445.121.1.12
- MacLeod, C. M. (2015). The Stroop Effect. *Encyclopedia of Color Science and Technology*. New York: Springer Science+Business Media. doi 10.1007/978-3-642-27851-8\_67-1
- Macniven, J., Davis, C., Ho, M., Bradshaw, C., Szabadi, E. & Constantinescu, C. (2008). Stroop performance in multiple sclerosis: information processing, selective attention, or executive functioning? *Journal of the International Neuropsychological Society*, *14* (5), 805-814. doi: 10.1017/S1355617708080946
- 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. doi: 10.30552/ejep.v5i1.76
- Martínez-Morga, M., Quesada-Rico, M.P., Bueno, C. & Martínez, S. (2018). Bases neurobiológicas del trastorno del espectro autista y del trastorno por déficit de atención/hiperactividad: diferenciación neural y sinaptogénesis. *Revista de Neurología*, *66* (1), 1-5. doi: 10.33588/rn.66S01.2018033
- 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. *Revista Gaceta de Psiquiatría Universitaria*, 10 (3), 344-349. Recuperado de [http://revistagpu.cl/2014/GPU\\_sept\\_2014\\_PDF/INV\\_Propiedades.pdf](http://revistagpu.cl/2014/GPU_sept_2014_PDF/INV_Propiedades.pdf)
- Max J. E., et al. (2013). Psychiatric disorders in children and adolescents in the first six months after mild traumatic brain injury. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 25 (3), 187-197. doi: 10.1176/appi.neuropsych.12010011
- McKenna, F. & Sharma, D. (2004). Reversing the Emotional Stroop Effect Reveals That It Is Not What It Seems: The Role of Fast and Slow Components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30 (2), 382-392. doi: 10.1037/0278-7393.30.2.382
- Mead, L., Mayer, A. R., Bobholz, J. A., Woodley, S. J., Cunningham, J. M., Hammeke, T. A., et al. (2002). Neural basis of the Stroop interference task: Response competition or selective attention? *Journal of the International Neuropsychological Society*, 8 (6), 735-742. doi: 10.1017/S1355617702860015
- Meier, M. E. & Kane, M. J. (2013). Working memory capacity and Stroop interference. Global vs. Local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39 (3), 748-759. doi: 10.1037/a0029200
- Melia, J. L. (2000). La cuestión de las mediciones paralelas. *Psicothema*, 12 (2), 386-388. Recuperado de <http://www.psicothema.com/pdf/588.pdf>
- Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28, (5), 597-613. doi: 10.1002/ana.410280502
- Miller, A. (2010). Examining the Errors and Self-Corrections on the Stroop Test. ETD Archive, 606, Tesis. Recuperado de <https://engagedscholarship.csuohio.edu/etdarchive/606>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H. & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis. *Cognitive Psychology*, 41, 49-100. doi: 10.1006/cogp.1999.0734

- Monahan, J. S. (2001). *Effects of practice on Stroop congruity*. Paper presented at the 17th annual meeting of the International Society for Psychophysics, Leipzig. Recuperado de [https://www.researchgate.net/publication/2548145\\_Effects\\_Of\\_Practice\\_On\\_Stroop\\_Congruity](https://www.researchgate.net/publication/2548145_Effects_Of_Practice_On_Stroop_Congruity)
- Montero, I. & León, O. (2007). A guide for naming research studies in Psychology. *Journal of Clinical and Health Psychology*, 7 (3), 847-862. Recuperado de [http://www.aepc.es/ijchp/GNEIP07\\_es.pdf](http://www.aepc.es/ijchp/GNEIP07_es.pdf)
- Morrow, S. A. (2013). Normative Data for the Stroop Color Word Test for a North American Population. *Canadian Journal of Neurological Sciences*, 40, 842-847. doi: 10.1017/s0317167100015997
- Mullane, J. C., Corkum, P. V., Klein, R. M., McLaughlin, E. N. & Lawrence, M. A. (2010). Alerting, orienting, and executive attention in children with ADHD. *Journal of Attention Disorders*, 7. Recuperado de <http://jad.sagepub.com/content/early/2010/05/27/1087054710366384>. doi:10.1177/1087054710366384
- Nicosia, J. & Balota, D. (2020). The consequences of processing goal-irrelevant information during the Stroop task. *Psychology and Aging*, 35 (5), 663-675. doi: 10.1037/pag0000371
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, 75, 522-536. doi: 10.1037/h0026699
- Okuzumi, H., et al. (2015). Stroop-Like Interference in the Fruit Stroop Test in Typical Development. *Psychology*, 6, 643-649. doi: 10.4236/psych.2015.65062
- Oliveira, R. Mograbi, D., Gabrig, I. & Charchat-Fichman, H. (2016). Normative data and evidence of validity for the Rey Auditory Verbal Learning Test, Verbal Fluency Test, and Stroop Test with Brazilian children. *Psychology and Neurosciences*, 9, 54-67. doi: 10.1037/pne0000041
- Organización Mundial de la Salud OMS (1995). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE-10)*. Edición de la Organización Panamericana de la salud. Recuperado de <http://ais.paho.org/classifications/Chapters/pdf/Volume1.pdf>

- Paiva Barón, H., Saona Olivera, G. F. & Perna Ramos, A. (2009). Capacidad discriminante de las variables de la batería MCC-94 en el Trastorno por Déficit de Atención con Hiperactividad. *Anales de Psicología*, 25 (1), 52-59. Recuperado de <https://www.redalyc.org/pdf/167/16711594006.pdf>
- Palencia Avendaño, M. L., Mogollón Rincón, C. & Etchepareborada, M. C. (2016). Análisis del tipo de error en el Continuous Performance Test (CPT) en la evaluación del déficit atencional. *Revista Argentina de Neuropsicología*, 29, 16-31. Recuperado de [https://7e5bfcbc-8cab-4aa5-94d4-d55ce46fc649.filesusr.com/ugd/2c1a84\\_9b1497e2782c4fe2a0a9bd5b181aab24.pdf](https://7e5bfcbc-8cab-4aa5-94d4-d55ce46fc649.filesusr.com/ugd/2c1a84_9b1497e2782c4fe2a0a9bd5b181aab24.pdf)
- Pardo, J. V., Pardo, P., Janer, K. & Raichle, M. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences of the USA*, 87, 257-259. doi: 10.1073/pnas.87.1.256
- Parris, B. A., Augustinova, M., Ferrand, L., eds. (2020). *The Locus of the Stroop Effect*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-445-3
- Parsons, T., Courtney, C., Arismendi, B. & Dawson, M. (2011). Virtual Reality Stroop Task for neurocognitive assessment. *Medicine Meets Virtual Reality*, 18, 433-439. doi:10.3233/978-1-60750-706-2-433
- Paus, T., Petrides, M., Evans, A. & Meyer, E. (1993). Role of the human Anterior Cingulate Cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology*, 70 (2), 453-469. doi: 10.1152/jn.1993.70.2.453
- Penner, I. K., Kobel, M., Stöcklin, M., Weber, P., Opwis, K. & Calabrese, P. (2012). The Stroop task: comparison between the original paradigm and computerized versions in children and adults. *The Clinical Neuropsychologist*, 1-12. doi: 10.1080/13854046.2012.713513
- Petersen, S. E. & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu Rev Neurosci*, 21 (35), 73-89. Doi: 10.1146/annurev-neuro-062111-150525



- Pineda, D., Lopera, F., Henao, G., Palacios, J. & Castellanos, F. X. (2001). Confirmación de la alta prevalencia del trastorno por déficit de atención en una comunidad colombiana. *Revista de Neurología*, 32 (3), 217-222. doi: 10.33588/rn.3203.2000499
- Pliszka, S. R., Glahn, D. C., Semrud-Clikeman, M., Franklin, C., Pérez, R., Xiong, J., et al. (2006). Neuroimaging of inhibitory control areas in children with Attention Deficit Hyperactivity Disorder who were treatment naive or in long-term treatment. *American Journal of Psychiatry*, 163 (6), 1052–1060. doi: 10.1176/appi.ajp.163.6.1052
- Pliszka, S. R., Liotti, M. & Woldorff, M. G. (2000). Inhibitory control in children with Attention-Deficit/Hyperactivity Disorder: Event-related potentials identify the processing component and timing of an impaired right-frontal response inhibition mechanism. *Biological psychiatry*, 48, 238–246. doi: 10.1016/s0006-3223(00)00890-8
- Polanczyk, G., Silva de Lima, M., Lessa Horta, B., Biederman, J. & Rohde, L. A. (2007). The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. *American Journal of Psychiatry*, 164, 942-948. doi: 10.1176/appi.ajp.164.6.942
- Posner, M. I. & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25-42. doi: 10.1146/annurev.ne.13.030190.000325
- Posner, M. I. & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology*, 12, 427-441. doi: 10.1017/S0954579400003096
- Posner, M. I. & Rothbart, M. K. (2009). Toward a physical basis of attention and selfregulation. *Physics of Life Reviews*, 6, 103-120. doi: 10.1016/j.plrev.2009.02.001
- Posner, M. I., & Snyder, C. (1975). Attention and cognitive control. En: Solso, R. (Ed.), *Information processing and cognition: The Loyola symposium*. (p. 55-85). Recuperado de: <https://philpapers.org/rec/POSAAC>
- Prieto, G. & Delgado, A. (2010). Fiabilidad y Validez. *Papeles del psicólogo*, 3 (1), 67-74.



- Protopapas, A., Archonti, A. & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, 54, 251-282. doi: 10.1016/j.cogpsych.2006.07.0003
- Ramírez-Benítez, Y. & Díaz-Bringas, M. (2011). Efecto Stroop y sus limitaciones ejecutivas en la práctica de la neuropsicología infantil. *Cuadernos de Neuropsicología*, 5 (2), 163-172. Recuperado de <http://www.cnps.cl/index.php/cnps/article/view/127>
- Ramos-Galarza, C. & Pérez-Salas, C. (2017). Control inhibitorio y monitorización en población infantil con TDAH. *Avances en Psicología Latinoamericana*, 35 (1), 117-130. doi: 10.12804/revistas.urosario.edu.co/apl/a.4195
- Ravnkilde, B., Videbech, P., Rosenberg, R., Gjedde, A. & Gade, A. (2002). Putative test of frontal lobe function: A PET-study of brain activation during a Stroop's test and verbal fluency. *Journal of Clinical and Experimental Neuropsychology*, 24, 534-547. doi: 10.1076/jcen.24.4.534.1033
- Razumiejczyk, E., Macbeth, G. & Adrover, J. F. (2011). Comparación de la interferencia Stroop intermodal entre representaciones gustativas y visuales por imágenes y por palabras. *Boletín de Psicología*, 101, 7-20. Recuperado de <https://www.academica.org/000-054/193.pdf>
- Redding, G. M. & Gerjets, D. A. (1977). Stroop effect: interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills*, 45, 11-17. doi: 10.2466/pms.1977.45.1.11
- Repovs, G. (2004). The mode of response and the Stroop effect: A reaction time analysis. *Psiholoska obzorja / Horizons of Psychology*, 13 (2), 105-114. Recuperado de [http://psiholoska-obzorja.si/arhiv\\_clanki/2004\\_2/repovs2.pdf](http://psiholoska-obzorja.si/arhiv_clanki/2004_2/repovs2.pdf)
- Ríos Lago, M., & Periañez Morales, J. A. (2010). *Attention and Speed of Information Processing*. En: Koob, G., Thompson R. F. & Le Moal, M. (Eds.), *Encyclopedia of Behavioral Neuroscience*. Elsevier
- Risko, E., Schmidt, J. & Besner, D. (2006). Filling a gap in the semantic gradient: Color associates and response set effects in the Stroop task. *Psychonomic Bulletin & Review*, 13 (2), 310-315. doi: 10.3758/BF03193849

- Rivera D., et al. (2015). Stroop color-word interference test: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, 37 (4), 591–624. doi: 10.3233/NRE-151281
- Rivera, D., Morlett Paredes, a., Peñalver Guia, A., Irías Escher, M., et al. (2017). Stroop Color-Word Interference Test: Normative Data on 10 Neuropsychological Tests on a Pediatric Population from 10 Spanish-Speaking Countries. *NeuroRehabilitation*, 41 (3), 605-616. doi: 10.3233/NRE-172246
- Rodríguez Barreto, L. C., Pulido, N. & Pineda Roa, C. (2016). Propiedades psicométricas del Stroop, test de colores y palabras en población colombiana no patológica. *Universitas Psychologica*, 15 (2), 255-272. doi: 10.11144/Javeriana.upsy15-2.ppst
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42 (1-3), 107-142. doi: 10.1016/0010-0277(92)90041-F
- Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, 110 (1), 88–125. doi: 10.1037/0033-295X.110.1.88
- Rognoni, T., Casals-Coll, M., Sánchez-Benavides, G., Quintana, M., Manero, R. M., Calvo, L., et al. (2012). Estudios normativos españoles en población adulta joven (proyecto NEURONORMA jóvenes): normas para las pruebas Stroop Color-Word Interference Test y Tower of London. *Revista de Neurología*, 28 (2), 73-80. doi: 10.1016/j.nrl.2011.12.020
- Roselli, M., Ardila, A., Santisi, M., Arecco, M., Salvatierra, J., Conde, A., et al. (2002). Stroop effect in Spanish–English bilinguals. *Journal of the International Neuropsychological Society*, 8, 819-827. doi: 10.1017/S1355617702860106
- Roselló, J., Munar, E., Obrador, P. & Cardell, E. (2007). Historia conceptual de la atención. *Revista Historia de la Psicología*, 28 (2), 59-65. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2383008>
- Rubiales, J. (2012). *Análisis de la flexibilidad cognitiva y la inhibición en niños con TDAH*. Tesis doctoral. Universidad Nacional de Mar del Plata. Repositorio de la Facultad de Psicología. Recuperado de <http://www.mdp.edu.ar/psicologia/psico/doctorado/tesis/rubiales.pdf>

- Rubiales, J., Bakker, L. & Urquijo, S. (2013). Estudio comparativo del control inhibitorio y la flexibilidad cognitiva en niños con Trastorno por déficit de atención con hiperactividad. *Cuadernos de Neuropsicología*, 7 (1), 50-69. doi: 10.7714/cnps/7.1.203
- Salgado-Pineda, P., Román, F., Sánchez-Navarro, J., López-Hernández, F., Bargalló, N., Falcón, C., et al. (2003). Activación cerebral durante el test de Stroop en un caso de lesión cerebral focal temprana. *Revista de Neurología*, 36, 343-346. doi: 10.33588/rn.3604.2002399
- Salo, R., Henik A. & Robertson, L. C. (2001). Interpreting Stroop interference: an analysis of differences between task versions. *Neuropsychology*, 15 (4), 462-471. doi: 10.1037//0894-4105.15.4.462
- Sánchez-Cubillo I, Periáñez J. A., Adrover-Roig D., Rodríguez-Sánchez J. M., Ríos-Lago M., Tirapu, J. & Barceló, F. (2009) Construct validity of the Trail Making Test: role of task switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15 (3), 438-450. doi: 10.1017/S1355617709090626
- Sánchez-López, R., quinto-Guillen, R., Pérez-Lucas, J., Jurado-Barba, R., Martínez-Grass, I., Ponce-alfaro, G. & Rubio-Valladolid, G. (2015). Validación de la versión española del Test Stroop de Alcohol. *Anales de Psicología*, 31 (2), 504-523. doi: 10.6018/analesps.31.2.167491
- Scarpina, F. & Tagini, S. (2017). The Stroop Color and Word Test. *Frontiers in Psychology Review*, 8, 1-8. doi: 10.3389/fpsyg.2017.00557
- Schachar, R., Tannock, R., Marriott, M. & Logan, G. (1995). Deficient Inhibitory Control in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, 23 (4), 411-437. doi: 10.1007/bf01447206
- Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology*, 72 (1), 105-108. doi: 10.1037/h0023358
- Schweiger, A., Doniger, G., Dwolatzky, T., Jaffe, D. & Simon, E. (2003). Reliability of a novel computerized neuropsychological battery for mild cognitive impairment. *Acta Neuropsychologica*, 4, 407-413. Recuperado de

- [https://www.researchgate.net/publication/284506248\\_Reliability\\_of\\_a\\_novel\\_computerized\\_neuropsychological\\_battery\\_for\\_mild\\_cognitive\\_impairment](https://www.researchgate.net/publication/284506248_Reliability_of_a_novel_computerized_neuropsychological_battery_for_mild_cognitive_impairment)
- Scurtu, M. C., Manzano-Arrondo, V. & Rodríguez Testal, J. F. (2016). Stroop Test Software. La propuesta de Tativa para pruebas Stroop. *Escritos de Psicología*, 9 (2), 47-52. doi: 10.5231/psy.writ.2016.2303
- Shepard, J. & Fogelsonger, H. (1913). Studies in association and inhibition. *Psychological Review*, 20 (4), 291-311. Recuperado de <https://zenodo.org/record/1922551#.XjhlBGhKiUk>
- Sisco, S., Slonena, E., Okun, M. S., Browsers, D. & Price, C. C. (2016). Parkinson's Disease and the Stroop Color Word Test: Processing Speed and Interference Algorithms. *The Clinical Neuropsychologist*, 30 (7), 1104-1117. doi: 10.1080/13854046.2016.1188989
- Sjowall, D., Roth, L., Lindqvist, S. & Thorell, L. (2012). Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry*. doi:10.1111/jcpp.12006
- Smoleń, T. & Chuderski, A. (2010). *Modeling strategies in Stroop with a general architecture of executive control*. Trabajo preliminar, presentado como ponencia en: 32nd Annual Conference of the Cognitive Science Society, Austin: Cognitive Science Society, 931-936. Recuperado de [https://www.researchgate.net/publication/252183663\\_Modeling\\_strategies\\_in\\_Stroop\\_with\\_a\\_general\\_architecture\\_of\\_executive\\_control](https://www.researchgate.net/publication/252183663_Modeling_strategies_in_Stroop_with_a_general_architecture_of_executive_control)
- Sobel, K., Puri, A. & York, K. (2020). Visual search inverts the classic Stroop asymmetry. *Acta Psychologica*, 205. Advance online publication. doi: 10.1016/j.actpsy.2020.103054
- Sociedad Argentina de Pediatría. Comité de Crecimiento y Desarrollo. (2017). Guía para el seguimiento del desarrollo infantil en la práctica pediátrica. *Archivos Argentinos de Pediatría*, 115 (3), 53-62. Recuperado de [https://www.sap.org.ar/uploads/consensos/consensos\\_guia-para-el-seguimiento-del-desarrollo-infantil-en-la-practica-pediatica-68.pdf](https://www.sap.org.ar/uploads/consensos/consensos_guia-para-el-seguimiento-del-desarrollo-infantil-en-la-practica-pediatica-68.pdf)
- Soprano, A. M. (2003). Evaluación de las funciones ejecutivas en el niño. *Revista de Neurología*, 37, 44-50. doi: 10.33588/rn.3701.2003237

- Sørensen, L., Plessen, K., Adolphsdottir, S & Lundervold, A. (2013). The specificity of the Stroop interference score of errorsto ADHD in boys. *Child Neuropsychology*, 1-18. doi: 10.1080/09297049.2013.855716
- Spinelli, G., & Lupker, S. J. (2020). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/xlm0000820
- Stafford, T. & Gurney, K. (2007). Biologically constrained action selection improves cognitive control in a model of the Stroop task. *Philosophical Transaction of the Royal Society*, 365, 1671–1684. doi: 10.1098/rstb.2007.2060
- Steinhauser, M. & Hubner, R. (2009). Distinguishing response conflict and task conflict in the Stroop task: evidence from ex-gaussian distribution analysis. *Journal of Experimental Psychology*, 35 (5), 1398–1412. doi: 10.1037/a0016467
- Stins, J. F., Tollenaar, M. S., Slaats-Willemse, D., Buitelaar, J. K., Swaab-Barneveld, H., Verhulst, F., et al. (2005). Sustained attention and executive functioning performance in Attention-Deficit/Hyperactivity Disorder. *Child Neuropsychology*, 11, 285-294. doi: 10.1080/09297040490916938
- Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology*, 31 (1) 121-132. doi: 10.1080/14640747908400712
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662. doi: 10.1037/h0054651
- Stuss, D. T., Floden, D., Alexander, M. P., Levine, B. & Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia*, 39 (8), 771-786. doi: 10.1016/S0028-3932(01)00013-6
- Sugg, M. J. & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 20(3), 647-675. doi: 10.1037//0096-1523.20.3.647

- Swick, D. & Javanovic, J. (2002). Anterior cingulate cortex and Stroop task: neuropsychological evidence for topographic specificity. *Neuropsychologia*, *40*, 1240-1253. doi: 10.1016/S0028-3932(01)00226-3
- Tamm, L., Menon, V. & Reiss, A. (2002). Maturation of brain function associated with response inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry* *41* (10), 1231-1238. doi: 10.1097/01.CHI.0000020272.43550.5E
- Tirapu-Ustárrroz, J., García-Molina, A., Luna-Lario, P., Roig-Rovira, T., & Pelegrín-Valero, C. (2008). Modelos de funciones y control ejecutivo (II). *Revista de Neurología*, *46*, 742-750.
- Tirapu-Ustárrroz, J., García-Molina, A., Ríos-Lago, M., & Ardila-Ardila, A. (2012). *Neuropsicología de la corteza prefrontal y las funciones ejecutivas*. Barcelona: Viguera Editores.
- Tirapu-Ustárrroz, J., Muñoz-Céspedes, J. M., Pelegrín-Valero, C. & Albéniz-Ferreras, A. (2005). Propuesta de un protocolo para la evaluación de las funciones ejecutivas. *Revista de Neurología*, *41*, 177-186. doi: 10.33588/rn.4103.2005054
- Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought processes*. New York: The MacMillan Company. Recuperado de: <https://archive.org/details/lecturesonexperi00tite>
- Tremblay, M. P., Potvin, O., Belleville, S., Bier, N., Gagnon, L., Blanchet, S., et al. (2016). The Victoria Stroop Test: normative data in Quebec-French adults and elderly. *Archives of Clinical Neuropsychology*, *31*, 926–933. doi: 10.1093/arclin/acw029
- Valgimigli, S., Padovani, R., Budriesi, C., Leone, M. E., Lugli, D. & Nichelli, P. (2010). The Stroop test: a normative Italian study on a paper version for clinical use. *Giornale Italiano di Psicologia*, *37* (4), 945-956. doi: 10.1421/33435
- Van der Elst, W., Van Boxtel, M., Van Breukelen, G. & 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. doi: 10.1177/1073191105283427
- Van der Elst, W., Van Boxtel, M., Van Breukelen, G. & Jolles, J. (2008). Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's

- Verbal Learning Test, and the Letter Digit Substitution Test: The regression-based change approach. *Journal of the International Neuropsychological Society*, *14*, 71–80. doi: 10.1017/S1355617708080028
- Van Maanen, L., Van Rijn, V. & Borst, J. P. (2009). Stroop and picture–word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, *16* (6), 987-999. doi: 10.3758/PBR.16.6.987
- Van Mourik, R., Oosterlaan, J. & Sergeant, J. (2005). The Stroop revisited: a meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry*, *46*(2), 150-165. doi: 10.1111/j.1469-7610.2004.00345.x
- Van Mourik, R., Papanikolau, A., Van Gellicum-Bijlhout, J., Van Oosterbruggen, J., Veugelers, D., Post-Uiterweer, A., Sergeant, J. & Oosterlaan, J. (2009). Interference Control in Children with Attention Deficit/Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *37*, 293-303. Recuperado de <https://link.springer.com/article/10.1007/s10802-008-9277-x>
- Vanier, M. (1991). *Test de Stroop*. Québec: SAAQ et FRSQ. Recuperado de [http://automobileevaluation.com/Stroop-M\\_Vanier.pdf](http://automobileevaluation.com/Stroop-M_Vanier.pdf)
- Vitkovitch, M., Bishop, S., Dancy, C. & Richards, A. (2002). Stroop interference and negative priming in patients with multiple sclerosis. *Neuropsychologia*, *40* (9), 1570-1576. doi: 10.1016/s0028-3932(02)00022-2
- Washburn, D. (2016). The Stroop Effect at 80: the competition between stimulus control and cognitive control. *Journal of the experimental analysis of behavior*, *105* (1), 3-13. doi: 10.1002/jeab.194
- Watanabe, Y., Sumitani, S., Hosokawa, M. & Ohmori, T. (2015). Prefrontal activation during two Japanese Stroop task revealed with multi-channel near-infrared spectroscopy. *The Journal of Medical Investigation*, *62*, 51-55. doi: 10.2152/jmi.62.51
- Weiler, M. D., Bernstein, J. H., Bellinger, D. & Waber, D. (2002). Information processing deficits in children with attention-deficit/hyperactivity disorder, inattentive type, and children with reading disability. *Journal of Learning Disabilities*, *35* (5), 448-461. doi: 10.1177/00222194020350050501



- West, R. & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*, 179–189. doi: 10.1111/1469-8986.3720179
- West, R. & Stanovich, K. (1979). The Development of Automatic Word Recognition Skills. *Journal of Reading Behavior*, *11* (3), 211-219. doi: 10.1080/10862967909547325
- Wilbourn, M. P., Kurtz, L. E. & Kalia, V. (2012). The Lexical Stroop Sort (LSS) picture-word task: A computerized task for assessing the relationship between language and executive functioning in school-aged children. *Behavior Research Methods*, *44* (1), 270–286. doi: 10.3758/s13428-011-0142-4
- Wright, B. C. (2017). What Stroop task can tell us about selective attention from childhood to adulthood. *British Journal of Psychology*, *108*, 583-607. doi: 10.1111/bjop.12230
- Wühr, P. & Waszak, F. (2003). Object-based attentional selection can modulate the Stroop effect. *Memory & Cognition*, *31* (6), 983-994. doi: 10.3758/BF03196450
- Wundt, W. (1904). Principles of physiological psychology. Recuperado de [http://elibrary.bsu.az/books\\_163/N\\_152.pdf](http://elibrary.bsu.az/books_163/N_152.pdf)
- Yoshihara, M., et al. (2020). Orthographic properties of distractors do influence phonological Stroop effects: Evidence from Japanese Romaji distractors. *Memory & Cognition*. Advance online publication. doi: 10.3758/s13421-020-01103-8
- Zalonis, I., Christidi, F., Bonakis, A., Kararizou, E., Triantafyllou, N. I., Paraskevas, G., et al. (2009). The stroop effect in Greek healthy population: normative data for the Stroop Neuropsychological Screening Test. *Archives of Clinical Neuropsychology*, *24*, 81–88. doi: 10.1093/arclin/acp011
- Zarghi, A, Zali, A., Tehranidost, M., Ashrafi, F., Zarindast Mohammad, R., Moazzezi, M., & Khodadadi Seyed, M. (2012). Assessment of selective attention with CSCWT (Computerized Stroop Color-Word Test) among children and adults. *US-China Education Review*, *1*, 121-127. Recuperado de [https://archive.org/details/ERIC\\_ED530676/page/n5/mode/2up](https://archive.org/details/ERIC_ED530676/page/n5/mode/2up)



- Ziaka, L., Moirou, D., Vlahou, E. & Protopapas, A. (2014). Word Reading Practice Reduces Stroop Interference in Children. *Acta Psychologica, 148*, 204-208. doi: 10.1016/j.actpsy.2014.02.006
- Zimmermann, N., Cardoso, C., Trentini, C., Grassi-Oliveira, R. & Fonseca, R. (2015). Brazilian preliminary norms and investigation of age and education effects on the Modified Wisconsin Card Sorting Test, Stroop Color and Word test and Digit Span test in adults. *Dementia & Neuropsychologia, 9*, 120–127. doi: 10.1590/1980-57642015DN92 000006

## ANEXOS

Anexo 1

Formato de presentación del Stroop

1.1 Aplicación del Stroop tradicional

**Tabla 1 (LP-a)**  
Leer las palabras escritas

NEGRO	ROJO	VERDE
ROJO	AZUL	
VERDE	VERDE	
NEGRO	ROJO	
AZUL	NEGRO	
VERDE	AZUL	

**Tabla 2 (NBC-a)**  
Nombrar el color de las barras

[Barra Verde]	[Barra Negra]	[Barra Roja]
[Barra Azul]	[Barra Verde]	
[Barra Roja]	[Barra Negra]	
[Barra Verde]	[Barra Azul]	
[Barra Negra]	[Barra Verde]	
[Barra Roja]	[Barra Verde]	

**Tabla 3 (INT-a)**  
Nombrar el color de las palabras impresas (no leer)

ROJO	ROJO	AZUL
AZUL	VERDE	VERDE
ROJO	NEGRO	ROJO
VERDE	AZUL	AZUL
NEGRO	VERDE	NEGRO
ROJO	ROJO	VERDE

1.2 Aplicación del Stroop computarizado

The diagram illustrates the computerized Stroop application through a sequence of three screens:

- Screen 1: A white box containing the word "AZUL" on a dark background.
- Screen 2: A solid blue horizontal bar on a dark background.
- Screen 3: A white box containing the word "AZUL" in green text on a dark background.

## Anexo 2

## Autorización institucional y Consentimiento informado

## 2.1 Autorización institucional para recolección de datos en archivo

LABORATORIO PARA EL ESTUDIO DE LAS  
FUNCIONES CEREBRALES SUPERIORES

Director: Prof. Dr. Máximo C. Etchepareborda

Especialista Universitario en Neurología Infantil – Doctor en Medicina de la Universidad Nacional de Buenos Aires  
Diploma de Honor de la Facultad de Medicina de la U.N.B.A. – ExNeuropediatra del Instituto Central para la Salud Mental de la  
Universidad de Heidelberg de la República Federal Alemana  
Ex Profesor Auxiliar Extraordinario de Patología del Habla y del Lenguaje de la Universidad del Salvador  
Profesor de la Maestría de Neuropsicología del Hospital Italiano de Buenos Aires, Argentina  
Director científico del Centro de Neurodesarrollo Interdisciplinar Red-Cemil Valencia y Castellón de la Plana  
Presidente de la Academia Iberoamericana de Neuropediatría (AINP) 2009-2011

## A quien corresponda.

Buenos Aires, 14 de enero de 2015

Motivo: Autorización institucional para uso de datos con fines de investigación.

De mi mayor consideración,

Por la presente me dirijo a Ustedes, para autorizar a la psicóloga **María Luisa Palencia** a realizar una investigación que involucra el registro, documentación y manejo de los datos obtenidos en una selección de pacientes atendidos en nuestra Institución.

Dichos datos serán destinados al desarrollo del proyecto titulado:

"COMPARACIÓN DE DOS VERSIONES DEL TEST DE PALABRAS Y COLORES STROOP".

Previo al inicio, la investigadora entregará para su aprobación:

- Constancia de recepción y revisión (Adjunto 1)
- Cronograma de trabajo (Adjunto 2)

La investigadora se compromete a dar un uso cuidadoso, responsable y ético de los datos recabados en los archivos de nuestra institución. Para ello deberá:

- Presentar para su aprobación el formato de recolección, donde defina qué variables ingresará
- Asignar códigos alfa numéricos que reemplacen los datos de identificación de las personas
- Abstenerse de realizar modificaciones a los datos consignados en los documentos utilizados (Historias clínicas, informes de otros profesionales, etc.)
- Presentar un informe de resultados a su culminación

## Referentes

## Legislación en Argentina:

- Sobre los derechos del paciente en su relación con los profesionales e instituciones de la salud señalados en la Ley 26.529 (Octubre 21 de 2009) y Ley 26.742 (Mayo 24 de 2012).
- Guía para investigaciones en salud humana consignada en la Resolución Ministerial 1480/2011.

## Acuerdos internacionales:

- Declaración de Helsinki (1964)
- Declaración de Tokio (1975)
- Declaración bioética de Gijón. Sociedad Internacional de Bioética (2000).
- Declaración Internacional sobre Bioética y Derechos Humanos, UNESCO (2005)
- Principios éticos de los psicólogos y código de conducta. American Psychological Association (2010)

Sin otro motivo en particular, aprovecho para saludar cordialmente,

Dr. Máximo C. Etchepareborda  
Director de LAFUN

Dr. MÁXIMO C. ETCHEPAREBORDA  
Especialista Universitario en  
Neurología Infantil  
M.N. 57.296

Laboratorio para el Estudio de las Funciones Cerebrales Superiores  
Estados Unidos 3402 – 1228 Buenos Aires, Argentina, Tel / Fax: (54) 11 – 4931-2335 ó 4957-4990  
Email: [lafun.ar@gmail.com](mailto:lafun.ar@gmail.com), Web: [www.lafun.com.ar](http://www.lafun.com.ar)

## 2.2 Autorización institucional para toma de la muestra control

### LABORATORIO PARA EL ESTUDIO DE LAS FUNCIONES CEREBRALES SUPERIORES

**Director: Prof. Dr. Máximo C. Etchepareborda**

Especialista Universitario en Neurología Infantil – Doctor en Medicina de la Universidad Nacional de Buenos Aires  
Diploma de Honor de la Facultad de Medicina de la U.N.B.A. – ExNeuropediatra del Instituto Central para la Salud Mental de la  
Universidad de Heidelberg de la República Federal Alemana  
Ex Profesor Auxiliar Extraordinario de Patología del Habla y del Lenguaje de la Universidad del Salvador  
Profesor de la Maestría de Neuropsicología del Hospital Italiano de Buenos Aires, Argentina  
Director científico del Centro de Neurodesarrollo Interdisciplinar Red-Cenit Valencia y Castellón de la Plana  
Presidente de la Academia Iberoamericana de Neuropediatría (AINP) 2009-2011  
Especialista en Estimulación Magnética Transcraneana por la Harvard Medical School, USA  
European Clinical TMS Certification. Miembro internacional de la Clínica TMS Society

#### A quien corresponda.

Buenos Aires, 14 de mayo de 2017

**Motivo:** Permiso para recolección de información a voluntarios.

De mi mayor consideración,

Por la presente me dirijo a Ustedes, para autorizar a la investigadora **María Luisa Palencia** para la toma de pruebas de evaluación a voluntarios convocados a nuestra institución y que serán referidos como “muestra control” en el desarrollo la investigación titulada “Comparación de dos versiones del Test de Palabras y Colores Stroop”.

La investigadora entregará para su revisión y aprobación:

- Constancia de recepción y revisión (Adjunto 1)
- Formato de consentimiento informado (Adjunto 2)
- Cuestionario para seleccionar la muestra control (Adjunto 3)

La investigadora se compromete a dar un uso cuidadoso, responsable y ético de los datos recabados de los voluntarios convocados.

#### Referentes

##### Legislación en Argentina:

- Sobre los derechos del paciente en su relación con los profesionales e instituciones de la salud señalados en la Ley 26.529 (Octubre 21 de 2009) y Ley 26.742 (Mayo 24 de 2012).
- Guía para investigaciones en salud humana consignada en la Resolución Ministerial 1480/2011.

##### Acuerdos internacionales:

- Declaración de Helsinki (1964)
- Declaración de Tokio (1975)
- Declaración bioética de Gijón. Sociedad Internacional de Bioética (2000).
- Declaración Internacional sobre Bioética y Derechos Humanos, UNESCO (2005)
- Principios éticos de los psicólogos y código de conducta. American Psychological Association (2010)

Sin otro motivo en particular, aprovecho para saludar cordialmente,

  
**Dr. Máximo C. Etchepareborda**  
Director de LAFUN

**Dr. MÁXIMO C. ETCHEPAREBORDA**  
Especialista Universitario en  
Neurología Infantil  
M.N. 57.296

Laboratorio para el Estudio de las Funciones Cerebrales Superiores  
Estados Unidos 3402 – 1228 Buenos Aires, Argentina, Tel / Fax: (54) 11 – 4931-2335 ó 4957-4990  
Email: lafun.ar@gmail.com , Web: www.lafun.com.ar , WhatsApp: +54911-30025885

**2.3 Formato aprobado de consentimiento informado**

**LABORATORIO PARA EL ESTUDIO DE LAS FUNCIONES CEREBRALES SUPERIORES**

**Director: Prof. Dr. Máximo C. Etchepareborda**

Especialista Universitario en Neurología Infantil – Doctor en Medicina de la Universidad Nacional de Buenos Aires  
 Diploma de Honor de la Facultad de Medicina de la U.N.B.A. – ExNeuropediatra del Instituto Central para la Salud Mental de la  
 Universidad de Heidelberg de la República Federal Alemana  
 Ex Profesor Auxiliar Extraordinario de Patología del Habla y del Lenguaje de la Universidad del Salvador  
 Profesor de la Maestría de Neuropsicología del Hospital Italiano de Buenos Aires, Argentina  
 Director científico del Centro de Neurodesarrollo Interdisciplinar Red-Cenit Valencia y Castellón de la Plana  
 Presidente de la Academia IberoAmericana de Neuropediatría (AINP) 2009-2011  
 Especialista en Estimulación Magnética Transcraneana por la Harvard Medical School, USA  
 European Clinical TMS Certification. Miembro internacional de la Clínica TMS Society

**Adjunto 2. Formato de consentimiento informado.**

CONSENTIMIENTO INFORMADO

Por la presente yo, \_\_\_\_\_, en mi calidad de madre/padre del menor \_\_\_\_\_, manifiesto que:

He recibido información clara y satisfactoria sobre los objetivos de la investigación “Comparación de dos versiones del Test de Colores y Palabras Stroop”, así como sobre el procedimiento y los instrumentos que serán usados, la duración de la aplicación de las pruebas y la devolución de resultados.

Se me ha informado que los datos otorgados por mí serán estrictamente confidenciales y no se usarán para ningún otro propósito fuera de los de la investigación; las respuestas al cuestionario y los puntajes obtenidos en el proceso de evaluación de mi hijo/hija se codificarán y resguardarán bajo la forma de anonimato.

He acordado que podré hacer las preguntas que considere necesarias, que podremos retirarnos en cualquier momento, y que me será entregada una copia de esta ficha de consentimiento, así como un informe individual de resultados de cada prueba.

Por lo anterior, acepto que mi hijo/hija \_\_\_\_\_, participe en este proceso.

En constancia de ello, doy mi consentimiento.

\_\_\_\_\_  
 Firma

\_\_\_\_\_  
 Aclaración

Buenos Aires, \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Firmas autorizadas:  
 COMITÉ DE ÉTICA LAFUN

  
 Dr. MÁXIMO C. ETCHEPAREBORDA  
 Especialista Universitario en  
 Neurología Infantil  
 M.N. 57.296

Dirección científica

  
 JULIETA TEJADA  
 MÉDICA  
 MN 161632  
 Supervisión

  
 Secretaria

Laboratorio para el Estudio de las Funciones Cerebrales Superiores  
 Estados Unidos 3402 – 1228 Buenos Aires, Argentina, Tel / Fax: (54) 11 – 4931-2335 ó 4957-4990  
 Email: lafun.ar@gmail.com , Web: www.lafun.com.ar , WhatsApp: +54911-30025885

**2.4 Formato aprobado de cuestionario para seleccionar la muestra control**

**LABORATORIO PARA EL ESTUDIO DE LAS FUNCIONES CEREBRALES SUPERIORES**

**Director: Prof. Dr. Máximo C. Etchepareborda**

Especialista Universitario en Neurología Infantil – Doctor en Medicina de la Universidad Nacional de Buenos Aires  
 Diploma de Honor de la Facultad de Medicina de la U.N.B.A. – ExNeuropediatra del Instituto Central para la Salud Mental de la Universidad de Heidelberg de la República Federal Alemana  
 Ex Profesor Auxiliar Extraordinario de Patología del Habla y del Lenguaje de la Universidad del Salvador  
 Profesor de la Maestría de Neuropsicología del Hospital Italiano de Buenos Aires, Argentina  
 Director científico del Centro de Neurodesarrollo Interdisciplinar Red-Cenit Valencia y Castellón de la Plana  
 Presidente de la Academia Iberoamericana de Neuropediatría (AINP) 2009-2011  
 Especialista en Estimulación Magnética Transcraneana por la Harvard Medical School, USA  
 European Clinical TMS Certification. Miembro internacional de la Clinical TMS Society

**Adjunto 3. Cuestionario para seleccionar la muestra control.**

**INFORMACIÓN DEL PARTICIPANTE**

Sexo: Mas.  Fem.  Fecha de nacimiento: \_\_\_ / \_\_\_ / \_\_\_\_\_

¿Con quién vive? \_\_\_\_\_

Número de hermanos: \_\_\_ Lugar entre hermanos: Mayor \_\_\_ Medio \_\_\_ Menor \_\_\_

Lugar donde vive actualmente: Barrio \_\_\_\_\_ Ciudad \_\_\_\_\_

Lateralidad: Diestro  Zurdo  Ambidiestro  Grado escolar actual: \_\_\_\_\_

Rendimiento académico: Bajo \_\_\_ Medio \_\_\_ Alto \_\_\_ Último promedio: \_\_\_\_\_

¿Ha perdido algún año escolar? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Tuvo que repetir alguna materia? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Se le dificulta aprender? Si \_\_\_ No \_\_\_ ¿En qué área? \_\_\_\_\_

¿Tiene mal comportamiento? Si \_\_\_ No \_\_\_ ¿Dónde? \_\_\_\_\_

¿Presenta problemas de sueño? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Presenta problemas de salud? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Toma alguna medicación? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Tiene dificultades para oír o ver? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Tuvo complicaciones al nacer? Si \_\_\_ No \_\_\_ ¿Cuál? \_\_\_\_\_

¿Tuvo dificultades para caminar? Si \_\_\_ No \_\_\_

¿Tuvo dificultades para hablar? Si \_\_\_ No \_\_\_

¿Convulsión alguna vez? Si \_\_\_ No \_\_\_ ¿Cuándo? \_\_\_\_\_

¿Se desmaya habitualmente? Si \_\_\_ No \_\_\_ ¿Cuándo? \_\_\_\_\_

¿De hábito, fuma, bebe o se droga? Si \_\_\_ No \_\_\_ ¿Qué? \_\_\_\_\_



**Dr. MÁXIMO C. ETCHEPAREBORDA**  
 Especialista Universitario en  
 Neurología Infantil  
 M.N. 57.296

Laboratorio para el Estudio de las Funciones Cerebrales Superiores  
 Estados Unidos 3402 – 1228 Buenos Aires, Argentina, Tel / Fax: (54) 11 – 4931-2335 ó 4957-4990  
 Email: [lafun.ar@gmail.com](mailto:lafun.ar@gmail.com) , Web: [www.lafun.com.ar](http://www.lafun.com.ar) , WhatsApp: +54911-30025885

## Anexo 3

## Descriptivos generales de latencias estimadas en la versión tradicional y pruebas de ajuste

## 3.1 Latencias estimadas en la tarea de lectura de palabras (P) en la versión tradicional

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	333	4500	683.26	607.320	4.221	20.537
Varones	319	4091	630.44	407.666	5.234	36.777
7-10 años	459	4500	977.25	736.812	2.920	9.701
11-15 años	372	818	512.00	85.734	0.931	1.623
16-22 años	319	643	458.84	75.392	0.427	-0.382
TDAH-DA	349	2368	629.18	344.163	3.018	11.190
TDAH-C	319	2250	621.32	368.561	3.303	11.963
TAPZ	333	4500	776.05	785.994	3.632	13.472
CONTROL	344	2250	570.90	262.849	4.694	28.578
Total	319	4500	649.36	488.069	4.855	28.984

## 3.2 Normalidad y Homocedasticidad para las latencias de respuesta estimadas en P

		Prueba de normalidad		Homogeneidad de la varianza	
		<i>Kolmogorov-Smirnov<sup>a</sup></i>	<i>p</i>	<i>Estadístico de Levene</i>	<i>p</i>
Sexo	Mujeres	.348	.000**	2.445	.119
	Varones	.252	.000**		
Edad	7-10	.255	.000**	41.108	.000**
	11-15	.068	.200		
	16-22	.073	.200		
Grupo	TDAH-DA	.234	.000**	5.464	.001**
	TDAH-C	.274	.000**		
	TAPZ	.345	.000**		
	CONTROL	.229	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)



**3.3 Latencias estimadas en la tarea de nominación de color (C) en la versión tradicional**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	500	2250	880.29	340.577	2.168	5.357
Varones	450	1667	853.10	232.058	1.083	1.360
7-10 años	726	2250	1108.94	319.269	1.625	3.112
11-15 años	500	1125	797.61	137.987	0.276	-0.320
16-22 años	450	1154	683.31	118.215	0.848	2.080
TDAH-DA	529	2250	893.92	315.191	1.854	4.992
TDAH-C	450	1406	855.47	225.318	0.640	0.014
TAPZ	529	2045	914.68	307.827	1.576	2.556
CONTROL	469	2025	789.08	241.061	3.848	22.567
Total	450	2250	863.29	277.653	1.967	5.874

**3.4 Normalidad y Homocedasticidad para las latencias de respuesta estimadas en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		<i>Kolmogorov-Smirnov<sup>a</sup></i>	<i>p</i>	<i>Estadístico de Levene</i>	<i>p</i>
Sexo	Mujeres	.209	.000**	5.009	.026*
	Varones	.106	.000**		
Edad	7-10	.172	.000**	30.668	.000**
	11-15	.087	.200		
	16-22	.096	.066		
Grupo	TDAH-DA	.144	.003**	3.019	.031*
	TDAH-C	.109	.071		
	TAPZ	.171	.000**		
	CONTROL	.173	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

### 3.5 Latencias estimadas en la tarea de interferencia palabra-color (PC) en la versión tradicional

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	763	2813	1376.94	468.721	1.023	0.279
Varones	738	3750	1498.40	541.227	1.317	2.287
7-10 años	1000	3750	1908.28	511.564	1.029	2.131
11-15 años	763	2500	1353.26	382.226	1.234	1.314
16-22 años	738	2368	1103.09	249.514	2.103	7.918
TDAH-DA	789	3214	1524.73	542.916	0.959	0.452
TDAH-C	738	3750	1586.80	615.375	1.345	2.223
TAPZ	763	2813	1451.93	451.514	0.963	0.723
CONTROL	746	2143	1256.03	386.072	0.986	-0.078
Total	738	3750	1454.88	518.713	1.262	1.964

### 3.6 Normalidad y Homocedasticidad para las latencias de respuesta estimadas en PC

		Prueba de normalidad		Homogeneidad de la varianza	
		<i>Kolmogorov-Smirnov<sup>a</sup></i>	<i>p</i>	<i>Estadístico de Levene</i>	<i>p</i>
Sexo	Mujeres	.171	.000**	1.238	.075
	Varones	.140	.000**		
Edad	7-10	.161	.000**	11.038	.000**
	11-15	.170	.000**		
	16-22	.132	.002**		
Grupo	TDAH-DA	.144	.004**	4.231	.006**
	TDAH-C	.139	.005**		
	TAPZ	.194	.000**		
	CONTROL	.183	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 4

**Análisis de varianza factorial o univariante para las mediciones de latencias estimadas de respuesta en las tareas de la versión tradicional**

**4.1 Pruebas de efectos intersujetos para la tarea P de la versión tradicional**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	4.197	.000	.309
Intersección	1	420.55	.000	.661
Sexo	1	.215	.643	.001
Edad	2	29.280	.000	.213
Grupo <sup>b</sup>	3	2.472	.063	.033
Sexo × Edad	2	.424	.655	.004
Sexo × Grupo	3	.191	.902	.003
Edad × Grupo	6	2.098	.055	.055
Sexo × Edad × Grupo	6	.352	.908	.010

<sup>a</sup>  $R^2 = .309$  ( $R^2$  ajustada = .235).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

**4.2 Pruebas de efectos intersujetos para la tarea C de la versión tradicional**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	9.758	.000	.510
Intersección	1	3256.23	.000	.938
Sexo	1	1.107	.294	.005
Edad	2	83.491	.000	.436
Grupo <sup>b</sup>	3	4.596	.004	.060
Sexo × Edad	2	2.958	.054	.027
Sexo × Grupo	3	.572	.634	.008
Edad × Grupo	6	2.279	.037	.060
Sexo × Edad × Grupo	6	.409	.873	.011

<sup>a</sup>  $R^2 = .510$  ( $R^2$  ajustada = .457).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

**4.2 Pruebas de efectos intersujetos para la tarea PC de la versión tradicional**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	10.130	.000	.519
Intersección	1	2592.182	.000	.923
Sexo	1	2.768	.098	.013
Edad	2	72.975	.000	.403
Grupo <sup>b</sup>	3	4.926	.002	.064
Sexo × Edad	2	.909	.404	.008
Sexo × Grupo	3	.776	.508	.011
Edad × Grupo	6	.372	.896	.010
Sexo × Edad × Grupo	6	1.121	.351	.030

<sup>a</sup>  $R^2 = .519$  ( $R^2$  ajustada = .468).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 5

## Descriptivos generales del número de aciertos en la versión tradicional y pruebas de ajuste

## 5.1 Aciertos registrados en la tarea de lectura de palabras (P) en la versión tradicional

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	10	135	83.75	26.970	-0.757	0.788
Varones	11	141	82.79	24.125	-0.589	0.367
7-10 años	10	98	59.34	22.700	-0.488	-0.624
11-15 años	54	121	89.65	14.375	0.069	-0.339
16-22 años	70	141	100.41	16.432	0.258	-0.510
TDAH-DA	17	129	82.55	25.529	-0.594	-0.027
TDAH-C	20	141	84.52	25.336	-0.588	0.548
TAPZ	10	135	78.67	26.945	-0.946	0.787
CONTROL	20	131	86.79	22.411	-0.210	0.305
Total	10	141	83.13	25.130	-0.654	0.544

## 5.2 Normalidad y Homocedasticidad para los aciertos en P

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.095	.055	0.431	.512
	Varones	.094	.002**		
Edad	7-10	.119	.007**	8.710	.000**
	11-15	.077	.200		
	16-22	.061	.200		
Grupo	TDAH-DA	.090	.200	0.370	.774
	TDAH-C	.117	.041*		
	TAPZ	.164	.000**		
	CONTROL	.085	.200		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**5.3 Aciertos registrados en la tarea de nominación de color (C) en la versión tradicional**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	20	90	54.98	15.815	-0.367	-0.356
Varones	25	100	54.99	14.599	0.443	0.269
7-10 años	20	62	41.65	10.486	-0.247	-0.863
11-15 años	39	90	56.71	10.654	0.546	0.375
16-22 años	37	100	66.60	11.777	0.334	0.322
TDAH-DA	20	85	53.90	15.750	-.036	-.695
TDAH-C	25	100	54.53	15.499	.726	.845
TAPZ	22	81	51.78	14.567	-.231	-.693
CONTROL	21	95	59.73	13.333	.115	.798
Total	20	100	54.99	15.013	.109	-.007

**5.4 Normalidad y Homocedasticidad para los aciertos en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.098	.039*	.950	.331
	Varones	.063	.200		
Edad	7-10	.089	.190	0.446	.641
	11-15	.064	.200		
	16-22	.062	.200		
Grupo	TDAH-DA	.079	.200	0.808	.491
	TDAH-C	.097	.200		
	TAPZ	.079	.200		
	CONTROL	.084	.200		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en  $**p < .01$  y  $*p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**5.5 Aciertos registrados en la tarea de interferencia palabra-color (PC) en la versión tradicional**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	10	55	33.44	11.177	-0.040	-.835
Varones	5	61	30.86	11.328	0.241	-0.451
7-10 años	5	42	21.80	7.011	0.536	0.646
11-15 años	13	55	32.70	9.032	-0.013	-0.144
16-22 años	18	61	40.86	8.462	-0.045	0.138
TDAH-DA	11	56	30.67	11.085	.041	-.871
TDAH-C	5	61	29.27	11.518	.301	-.198
TAPZ	9	55	29.57	10.409	.482	.039
CONTROL	19	58	37.63	10.391	-.152	-.863
Total	5	61	31.79	11.319	.136	-.638

**5.6 Normalidad y Homocedasticidad para los aciertos en PC**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.083	.200	.023	.879
	Varones	.065	.200		
Edad	7-10	.114	.012*	2.016	.135
	11-15	.082	.200		
	16-22	.064	.200		
Grupo	TDAH-DA	.115	.045*	.645	.581
	TDAH-C	.110	.070		
	TAPZ	.118	.036*		
	CONTROL	.102	.189		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en  $**p < .01$  y  $*p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 6

## Análisis de varianza factorial o univariante para las mediciones de aciertos en las tareas de la versión tradicional

## 6.1 Pruebas de efectos intersujetos para la tarea P de la versión tradicional

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	10.423	.000	.526
Intersección	1	3865.614	.000	.947
Sexo	1	.384	.536	.002
Edad	2	92.935	.000	.463
Grupo <sup>b</sup>	3	2.553	.057	.034
Sexo × Edad	2	.252	.778	.002
Sexo × Grupo	3	.859	.463	.012
Edad × Grupo	6	1.023	.410	.028
Sexo × Edad × Grupo	6	.771	.592	.021

<sup>a</sup>  $R^2 = .526$  ( $R^2$  ajustada = .476).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

## 6.2 Pruebas de efectos intersujetos para la tarea C de la versión tradicional

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	10.901	.000	.537
Intersección	1	4818.244	.000	.957
Sexo	1	.054	.817	.000
Edad	2	95.633	.000	.470
Grupo <sup>b</sup>	3	5.721	.001	.074
Sexo × Edad	2	1.247	.289	.011
Sexo × Grupo	3	.376	.771	.005
Edad × Grupo	6	1.119	.352	.030
Sexo × Edad × Grupo	6	.411	.871	.011

<sup>a</sup>  $R^2 = .537$  ( $R^2$  ajustada = .488).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.



**6.2 Pruebas de efectos intersujetos para la tarea PC de la versión tradicional**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	13.331	.000	.587
Intersección	1	3205.617	.000	.937
Sexo	1	1.682	.196	.008
Edad	2	103.923	.000	.490
Grupo <sup>b</sup>	3	13.569	.000	.159
Sexo × Edad	2	.405	.667	.004
Sexo × Grupo	3	.257	.856	.004
Edad × Grupo	6	.223	.969	.006
Sexo × Edad × Grupo	6	.575	.750	.016

<sup>a</sup>  $R^2 = .587$  ( $R^2$  ajustada = .543).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 7

## Descriptivos generales de número de errores en la versión tradicional y pruebas de ajuste

## 7.1 Errores registrados en la tarea de lectura de palabras (P) en la versión tradicional

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	6	0.36	0.944	3.759	16.997
Varones	0	16	0.51	1.483	7.826	78.404
7-10 años	0	6	0.54	1.113	2.699	8.306
11-15 años	0	16	0.55	1.882	7.271	59.097
16-22 años	0	3	0.29	0.640	2,333	5.021
TDAH-DA	0	16	0.73	2.261	5.637	36.420
TDAH-C	0	4	0.35	0.755	2.731	8.922
TAPZ	0	4	0.57	0.981	1.983	3.798
CONTROL	0	2	0.18	0.431	2.312	4.937
Total	0	16	0.46	1.315	7.692	82.89

## 7.2 Normalidad y Homocedasticidad para los errores en P

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.451	.000**	1.139	.287
	Varones	.365	.000**		
Edad	7-10	.410	.000**	2.327	.100
	11-15	.385	.000**		
	16-22	.473	.000**		
Grupo	TDAH-DA	.373	.000**	5.502	.001**
	TDAH-C	.445	.000**		
	TAPZ	.385	.000**		
	CONTROL	.498	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**7.3 Errores registrados en la tarea de nominación de color (C) en la versión tradicional**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	8	1.47	1.767	1.840	3.588
Varones	0	7	1.35	1.453	1.026	0.615
7-10 años	0	8	1.61	1.899	1.703	2.680
11-15 años	0	5	1.43	1.439	0.782	-0.424
16-22 años	0	4	1.14	1.290	0.934	-0.174
TDAH-DA	0	7	1.33	1.526	1.394	2.241
TDAH-C	0	8	1.72	1.823	1.218	1.680
TAPZ	0	7	1,82	1.662	1.195	1.419
CONTROL	0	4	0.70	0.889	1.388	2.166
Total	0	8	1.39	1.570	1.458	2.443

**7.4 Normalidad y Homocedasticidad para los errores en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.267	.000**	0.498	.481
	Varones	.212	.000**		
Edad	7-10	.264	.000**	2.691	.070
	11-15	.216	.000**		
	16-22	.249	.000**		
Grupo	TDAH-DA	.209	.000**	7.678	.000**
	TDAH-C	.186	.000**		
	TAPZ	.255	.000**		
	CONTROL	.301	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**7.5 Errores registrados en la tarea de interferencia palabra-color (PC) en la versión tradicional**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	26	2.60	3.799	3.518	17.012
Varones	0	21	2.64	2.985	2.358	9.225
7-10 años	0	26	3.43	4.046	2.672	11.243
11-15 años	0	21	2.86	3.291	2.929	12.168
16-22 años	0	9	1.60	1.953	1.850	3.808
TDAH-DA	0	9	2.30	2.465	1.467	1.673
TDAH-C	0	11	2.82	2.703	1.027	0.388
TAPZ	0	26	4.23	4.921	2.425	7.268
CONTROL	0	4	1.17	1.167	0.988	0.332
Total	0	26	2.63	3.292	3.010	14.291

**7.6 Normalidad y Homocedasticidad para los errores en PC**

	Prueba de normalidad		Homogeneidad de la varianza		
	Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>	
Sexo	Mujeres	.307	.000**	0.507	.477
	Varones	.193	.000**		
Edad	7-10	.229	.000**	7.949	.000**
	11-15	.208	.000**		
	16-22	.233	.000**		
Grupo	TDAH-DA	.215	.000**	14.086	.000**
	TDAH-C	.185	.000**		
	TAPZ	.219	.000**		
	CONTROL	.257	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 8

## Análisis de varianza factorial o univariante para las mediciones de errores en las tareas de la versión tradicional

## 8.1 Pruebas de efectos intersujetos para la tarea P de la versión tradicional

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	1.156	.288	.110
Intersección	1	25.074	.000	.104
Sexo	1	.170	.681	.001
Edad	2	1.921	.149	.017
Grupo <sup>b</sup>	3	2.741	.044	.037
Sexo × Edad	2	1.815	.165	.017
Sexo × Grupo	3	.737	.531	.010
Edad × Grupo	6	.983	.437	.027
Sexo × Edad × Grupo	6	1.388	.221	.037

<sup>a</sup>  $R^2 = .110$  ( $R^2$  ajustada = .015).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

## 8.2 Pruebas de efectos intersujetos para la tarea C de la versión tradicional

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	2.215	.002	.191
Intersección	1	156.125	.000	.420
Sexo	1	.839	.361	.004
Edad	2	4.815	.009	.043
Grupo <sup>b</sup>	3	5.800	.001	.075
Sexo × Edad	2	3.629	.028	.033
Sexo × Grupo	3	.922	.431	.013
Edad × Grupo	6	.767	.596	.021
Sexo × Edad × Grupo	6	1.478	.187	.039

<sup>a</sup>  $R^2 = .191$  ( $R^2$  ajustada = .105).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

**8.2 Pruebas de efectos intersujetos para la tarea PC de la versión tradicional**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	2.658	.000	.221
Intersección	1	149.738	.000	.409
Sexo	1	.259	.612	.001
Edad	2	8.861	.000	.076
Grupo <sup>b</sup>	3	10.384	.000	.126
Sexo * Edad	2	1.245	.290	.011
Sexo * Grupo	3	.943	.421	.013
Edad * Grupo	6	1.249	.283	.034
Sexo * Edad* Grupo	6	.527	.787	.014

<sup>a</sup>  $R^2 = .221$  ( $R^2$  ajustada = .138).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 9

## Índices de interferencia en el Stroop tradicional

9.1 Descriptivos generales ( $n=240$ ) de los índices de interferencia calculados

	Mín.	Máx.	M	DE	Asimetría	Curtosis
$I_D=C-PC^a$	-4.00	57.00	23.21	9.080	0.201	0.785
$I_R=PC/C^a$	0.15	1.18	0.57	0.137	0.382	1.229
$I_G=PC-[(P \times C)/(P+C)]^b$	-16.88	21.00	0.29	6.972	0.317	-0.119
$I_{REL}=[(C-PC)/C] \times 100^a$	-18.18	85.29	42.27	13.860	-0.454	1.436
$I_V=PC-[(P+C)/2]^a$	-6.00	21.50	1.70	3.000	2.592	11.898

<sup>a</sup> El cálculo se realiza con la puntuación bruta de aciertos ( $I_D$ ,  $I_R$ ,  $I_G$  e  $I_{REL}$ ) y de errores ( $I_V$ ).

<sup>b</sup> El cálculo se realiza con la puntuación de aciertos corregida por edad.

## 9.2 Comparación de medias de los índices calculados en el Stroop tradicional por rango etario

	7 a 10	11 a 15	16 a 22	Estadístico
$I_D$	19.85 (8.50)	24.01 (8.04)	25.74 (9.75)	$X^2(2)=22.652$ ( $p=.000$ ) **
$I_R$	0.54 (0.16)	0.58 (0.13)	0.62 (0.11)	$X^2(2)=2.226$ ( $p=.110$ ) ns
$I_G$	0.73 (7.02)	-0.95 (6.45)	1.10 (7.34)	$X^2(2)=0.528$ ( $p=.590$ ) ns
$I_{REL}$	46.30 (16.49)	42.37 (12.69)	38.16 (10.71)	$X^2(2)=2.220$ ( $p=.111$ ) ns
$I_V$	2.35 (3.61)	1.88 (3.14)	0.89 (1.79)	$X^2(2)=1.223$ ( $p=.296$ ) ns

M (DE)

Diferencia significativa: \*\* ( $p<.01$ ) y \* ( $p<.05$ ) Diferencia no significativa: ns ( $p>.05$ )

## 9.3 Comparación de medias de los índices calculados en el Stroop tradicional por grupo diagnóstico

	TDAH-DA	TDAH-C	TAPZ	CONTROL	Estadístico
$I_D$	23.23 (8.44)	25.4 (10.12)	22.13 (8.76)	22.08 (8.86)	$X^2(3)=0.772$ ( $p=.511$ ) ns
$I_R$	0.56 (0.57)	0.53 (0.14)	0.58 (0.15)	0.63 (0.13)	$X^2(3)=0.274$ ( $p=.844$ ) ns
$I_G$	-0.54 (5.32)	-2.26 (7.38)	-0.14 (7.39)	4.06 (6.13)	$X^2(3)=1.109$ ( $p=.346$ ) ns
$I_{REL}$	43.4 (11.08)	47.1 (14.45)	42.2 (14.60)	36.5 (13.24)	$X^2(3)=0.153$ ( $p=.928$ ) ns
$I_V$	1.27 (2.33)	1.84 (2.64)	2.97 (4.42)	0.73 (1.22)	$X^2(3)=1.049$ ( $p=.371$ ) ns

**9.4 Capacidad de clasificación de los índices de interferencia por rango etario**

Índice	Edad	Corte	Sensibilidad	Especificidad	AUC
I <sub>D</sub>	7 a 10	17.50	.40	.85	.637*
	11 a 15	28.50	.36	.80	.577
	16 a 22	28.50	.38	.75	.539
I <sub>R</sub>	7 a 10	0.37	.13	.95	.498
	11 a 15	0.55	.50	.85	.692*
	16 a 22	0.64	.65	.65	.644*
I <sub>G</sub>	7 a 10	-0.26	.48	.65	.521
	11 a 15	0.00	.68	.75	.726*
	16 a 22	6.49	.83	.55	.699*
I <sub>REL</sub>	7 a 10	42.56	.71	.50	.571
	11 a 15	39.45	.75	.75	.779**
	16 a 22	35.48	.66	.70	.700*
I <sub>V</sub>	7 a 10	2.75	.40	.90	.658*
	11 a 15	1.75	.50	.90	.640*
	16 a 22	0.25	.56	.55	.582

AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90



**9.5 Resumen de las mediciones básicas en la tarea PC del Stroop tradicional (n=240)**

	<i>Mín.</i>	<i>Máx.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Latencias estimadas	738	3750	1454.88	518.713	1.262	1.964
Número de aciertos	5	61	31.79	11.319	0.136	-0.638
Número de errores	0	26	2.63	3.292	3.010	14.291

**9.6 Capacidad de clasificación de las mediciones básicas en la tarea PC por rango etario**

	<i>Edad</i>	<i>Corte</i>	<i>Sensibilidad</i>	<i>Especificidad</i>	<i>AUC</i>
Latencias estimadas	7 a 10	1916	.55	.74	.674*
	11 a 15	1139	.74	.64	.727*
	16 a 22	1011	.73	.75	.803**
Aciertos	7 a 10	18.50	.41	1.00	.778**
	11 a 15	33.50	.66	.85	.805**
	16 a 22	39.50	.55	1.00	.823**
Errores	7 a 10	2.50	.50	.85	.707*
	11 a 15	1.50	.66	.70	.717*
	16 a 22	0.50	.71	.55	.670*

*AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90*

## Anexo 10

**Descriptivos generales de latencias medias de aciertos en la versión computarizada y pruebas de ajuste**

**10.1 Latencias medias de aciertos en la lectura de palabras (P) de la versión computarizada**

	<i>Min.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	624	2752	1011.24	378.334	2.331	6.384
Varones	445	3723	1084.38	439.744	2.808	11.485
7-10 años	445	3723	1408.02	541.742	1.771	4.499
11-15 años	589	1451	932.14	182.516	0.868	0.720
16-22 años	624	1241	834.36	121.251	0.866	1.795
TDAH-DA	651	3249	1129.75	440.074	2.483	8.478
TDAH-C	445	3723	1089.60	488.029	3.175	14.101
TAPZ	633	2752	1094.32	434.640	1.832	3.357
CONTROL	624	2133	919.03	254.362	2.335	8.373
Total	445	3723	1058.17	419.447	2.695	10.478

**10.2 Normalidad y Homocedasticidad para las latencias de respuesta en P**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.183	.000**	0.864	.354
	Varones	.181	.000**		
Edad	7-10	.172	.000**	47.533	.000**
	11-15	.123	.003**		
	16-22	.105	.002**		
Grupo	TDAH-DA	.204	.000**	2.750	.043*
	TDAH-C	.220	.000**		
	TAPZ	.206	.000**		
	CONTROL	.143	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**10.3 Latencias medias de aciertos en la nominación de color (C) de la versión computarizada**

	<i>Min.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	530	1873	889.79	266.150	1.501	2.337
Varones	580	2496	940.66	282.316	1.977	6.255
7-10 años	752	2496	1187.79	300.234	1.475	3.705
11-15 años	596	1378	827.91	149.791	1.010	1.439
16-22 años	530	1028	751.60	99.505	0.179	-0.115
TDAH-DA	582	2496	971.05	324.567	2.351	7.829
TDAH-C	580	1687	937.75	269.653	1.073	0.488
TAPZ	607	1948	944.83	293.892	1.436	2.005
CONTROL	530	1500	836.10	189.729	1.066	1.480
Total	530	2496	922.43	277.152	1.812	5.045

**10.4 Normalidad y Homocedasticidad para las latencias de respuesta en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.167	.000**	0.082	.774
	Varones	.133	.000**		
Edad	7-10	.119	.000**	27.912	.000**
	11-15	.094	.001**		
	16-22	.059	.783		
Grupo	TDAH-DA	.195	.000**	2.614	.052
	TDAH-C	.188	.000**		
	TAPZ	.156	.001**		
	CONTROL	.110	.067		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**10.5 Latencias medias de aciertos en el conflicto palabra-color (PC) de la versión computarizada**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	598	2504	1140.20	373.511	1.509	2.453
Varones	628	2643	1235.86	415.287	1.430	1.779
7-10 años	948	2643	1576.09	433.952	0.789	-0.304
11-15 años	736	1812	1071.56	230.133	1.105	1.124
16-22 años	598	1456	957.09	171.905	0.712	0.770
TDAH-DA	598	2623	1263.20	458.081	1.176	0.923
TDAH-C	628	2643	1239.78	446.338	1.435	1.620
TAPZ	695	2199	1215.10	382.561	1.134	0.648
CONTROL	698	2504	1088.23	287.713	2.143	8.694
Total	598	2643	1201.58	402.660	1.455	1.976

**10.6 Normalidad y Homocedasticidad para las latencias de respuesta en PC**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.156	.000**	.871	.352
	Varones	.164	.000**		
Edad	7-10	.135	.000**	33.861	.000**
	11-15	.149	.000**		
	16-22	.105	.022*		
Grupo	TDAH-DA	.143	.004	4.223	.006**
	TDAH-C	.223	.000**		
	TAPZ	.174	.000**		
	CONTROL	.098	.200		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 11

## Análisis de varianza factorial o univariante para las mediciones de latencias media de aciertos en las tareas de la versión computarizada

## 11.1 Pruebas de efectos intersujetos para la tarea P de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	7.047	.000	.429
Intersección	1	1779.051	.000	.892
Sexo	1	2.043	.154	.009
Edad	2	51.443	.000	.323
Grupo <sup>b</sup>	3	3.560	.015	.047
Sexo × Edad	2	.308	.736	.003
Sexo × Grupo	3	.338	.798	.005
Edad × Grupo	6	1.188	.314	.032
Sexo × Edad × Grupo	6	.293	.940	.008

<sup>a</sup>  $R^2 = .429$  ( $R^2$  ajustada = .368).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

## 11.2 Pruebas de efectos intersujetos para la tarea C de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	10.961	.000	.539
Intersección	1	3924.723	.000	.948
Sexo	1	1.915	.168	.009
Edad	2	93.063	.000	.463
Grupo <sup>b</sup>	3	4.881	.003	.063
Sexo × Edad	2	.076	.927	.001
Sexo × Grupo	3	.046	.987	.001
Edad × Grupo	6	1.864	.088	.049
Sexo × Edad × Grupo	6	.210	.973	.006

<sup>a</sup>  $R^2 = .539$  ( $R^2$  ajustada = .489).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

**11.2 Pruebas de efectos intersujetos para la tarea PC de la versión computarizada**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	10.663	.000	.532
Intersección	1	3067.092	.000	.934
Sexo	1	3.921	.049	.018
Edad	2	82.113	.000	.432
Grupo <sup>b</sup>	3	3.101	.028	.041
Sexo × Edad	2	.014	.986	.000
Sexo × Grupo	3	.110	.954	.002
Edad × Grupo	6	2.215	.043	.058
Sexo × Edad × Grupo	6	.739	.618	.020

<sup>a</sup>  $R^2 = .532$  ( $R^2$  ajustada = .482).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 12

## Descriptivos generales del número de aciertos en la versión computarizada y pruebas de ajuste

## 12.1 Aciertos registrados en la tarea de lectura de palabras (P) en la versión computarizada

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	61	72	6.85	2.442	-1.557	1.957
Varones	35	72	68.62	4.695	-3.905	21.146
7-10 años	35	72	66.79	5.955	-2.871	11.315
11-15 años	62	72	69.51	2.093	-1.127	1.647
16-22 años	67	72	70.89	1.222	-1.231	1.047
TDAH-DA	35	72	68.08	6.127	-3.788	16.904
TDAH-C	55	72	68.70	3.572	-2.131	5.654
TAPZ	55	72	68.90	3.261	-1.801	4.624
CONTROL	63	72	70.57	1.671	-2.251	6.792
Total	35	72	69.06	4.072	-4.133	25.863

## 12.2 Normalidad y Homocedasticidad para los aciertos en P

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.251	.000**	4.680	.032*
	Varones	.246	.000**		
Edad	7-10	.191	.000**	26.100	.000**
	11-15	.153	.000**		
	16-22	.262	.000**		
Grupo	TDAH-DA	.295	.000**	4.984	.002**
	TDAH-C	.217	.000**		
	TAPZ	.179	.000**		
	CONTROL	.286	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**12.2 Aciertos registrados en la tarea de nominación de color (C) en la versión computarizada**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	57	72	69.76	3.158	-2.833	8.569
Varones	25	72	68.99	4.779	-5.917	48.499
7-10 años	25	72	67.38	6.405	-4.234	24.421
11-15 años	58	72	69.50	2.561	-2.382	7.442
16-22 años	67	72	70.93	1.156	-1.011	0.604
TDAH-DA	60	72	69.60	2.492	-1.686	3.470
TDAH-C	25	72	67.35	7.253	-3.915	19.763
TAPZ	57	72	69.32	2.765	-2.104	6.241
CONTROL	66	72	70.80	1.205	-1.286	2.732
Total	25	72	69.27	4.278	-5.742	50.035

**12.4 Normalidad y Homocedasticidad para los aciertos en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.289	.000**	1.153	.284
	Varones	.265	.000**		
Edad	7-10	.240	.000**	16.405	.000**
	11-15	.227	.000**		
	16-22	.226	.000**		
Grupo	TDAH-DA	.214	.000**	11.763	.000**
	TDAH-C	.264	.000**		
	TAPZ	.254	.000**		
	CONTROL	.233	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)



**12.5 Aciertos registrados en la tarea de interferencia palabra-color (PC) en la versión computarizada**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	62	72	69.22	2.729	-1.103	0.230
Varones	35	72	67.69	4.442	-3.271	19.143
7-10 años	35	72	66.61	5.053	-3.349	18.589
11-15 años	52	72	68.31	3.271	-1.906	6.497
16-22 años	58	72	69.80	2.543	-1.926	5.142
TDAH-DA	57	72	68.47	3.362	-1.187	1.231
TDAH-C	35	72	66.93	5.757	-3.374	15.902
TAPZ	58	72	67.88	3.390	-0.631	-0.314
CONTROL	64	72	69.68	2.021	-1.309	1.341
Total	35	72	68.24	3.977	-3.232	20.678

**12.6 Normalidad y Homocedasticidad para los aciertos en PC**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.217	.000**	4.796	.029*
	Varones	.166	.000**		
Edad	7-10	.155	.000**	6.331	.002**
	11-15	.183	.000**		
	16-22	.219	.000**		
Grupo	TDAH-DA	.213	.000**	5.826	.001**
	TDAH-C	.190	.000**		
	TAPZ	.150	.002**		
	CONTROL	..229	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 13

## Análisis de varianza factorial o univariante para las mediciones de aciertos en las tareas de la versión computarizada

## 13.1 Pruebas de efectos intersujetos para la tarea P de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.903	.000	.294
Intersección	1	67599.730	.000	.997
Sexo	1	3.028	.083	.014
Edad	2	18.049	.000	.143
Grupo <sup>b</sup>	3	3.702	.013	.049
Sexo × Edad	2	.908	.405	.008
Sexo × Grupo	3	.385	.764	.005
Edad × Grupo	6	1.053	.392	.028
Sexo × Edad × Grupo	6	.466	.833	.013

<sup>a</sup>  $R^2 = .294$  ( $R^2$  ajustada = .218).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

## 13.2 Pruebas de efectos intersujetos para la tarea C de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.366	.000	.264
Intersección	1	58994.504	.000	.996
Sexo	1	.527	.469	.002
Edad	2	15.116	.000	.123
Grupo <sup>b</sup>	3	4.162	.007	.055
Sexo × Edad	2	.106	.899	.001
Sexo × Grupo	3	.459	.711	.006
Edad × Grupo	6	2.171	.047	.057
Sexo × Edad × Grupo	6	.119	.994	.003

<sup>a</sup>  $R^2 = .264$  ( $R^2$  ajustada = .185).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

**13.3 Pruebas de efectos intersujetos para la tarea PC de la versión computarizada**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.394	.000	.265
Intersección	1	66802.776	.000	.997
Sexo	1	6.334	.013	.028
Edad	2	11.312	.000	.095
Grupo <sup>b</sup>	3	3.327	.021	.044
Sexo × Edad	2	.239	.787	.002
Sexo × Grupo	3	.900	.442	.012
Edad × Grupo	6	1.026	.409	.028
Sexo × Edad × Grupo	6	1.436	.202	.038

<sup>a</sup>  $R^2 = .265$  ( $R^2$  ajustada = .187).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 14

## Descriptivos generales del número de errores en la versión computarizada y pruebas de ajuste

## 14.1 Errores registrados en la tarea de lectura de palabras (P) en la versión computarizada

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	11	2.14	2.450	1.549	1.929
Varones	0	37	3.38	4.695	3.905	21.146
7-10 años	0	37	5.21	5.955	2.871	11.315
11-15 años	0	10	2.48	2.105	1.109	1.591
16-22 años	0	5	1.11	1.222	1.231	1.047
TDAH-DA	0	37	3.92	6.127	3.788	16.904
TDAH-C	0	17	3.30	3.572	2.131	5.654
TAPZ	0	17	3.10	3.261	1.801	4.624
CONTROL	0	9	1.42	1.680	2.234	6.676
Total	0	37	2.93	4.074	4.127	25.810

## 14.2 Normalidad y Homocedasticidad para los errores en P

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.249	.000**	4.609	.033*
	Varones	.246	.000**		
Edad	7-10	.191	.000**	26.022	.000**
	11-15	.152	.000**		
	16-22	.262	.000**		
Grupo	TDAH-DA	.295	.000**	4.931	.002**
	TDAH-C	.217	.000**		
	TAPZ	.179	.000**		
	CONTROL	.281	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**14.3 Errores registrados en la tarea de nominación de color (C) en la versión computarizada**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	15	2,24	3,158	2,833	8,569
Varones	0	47	3,01	4,779	5,917	48,499
7-10 años	0	47	4.63	6.405	4.234	24.421
11-15 años	0	14	2.50	2.561	2.382	7.442
16-22 años	0	5	1.08	1.156	1.011	0.604
TDAH-DA	0	12	2.40	2.492	1.686	3.470
TDAH-C	0	47	4.65	7.253	3.915	19.763
TAPZ	0	15	2.68	2.765	2.104	6.241
CONTROL	0	6	1.20	1.205	1.286	2.732
Total	0	47	2.73	5.278	5.742	50.035

**14.4 Normalidad y Homocedasticidad para los errores en C**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.289	.000**	1.153	.284
	Varones	.265	.000**		
Edad	7-10	.240	.000**	16.405	.000**
	11-15	.227	.000**		
	16-22	.226	.000**		
Grupo	TDAH-DA	.214	.000**	11.763	.000**
	TDAH-C	.264	.000**		
	TAPZ	.254	.000**		
	CONTROL	.233	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

**14.5 Errores registrados en la tarea de interferencia palabra-color (PC) en la versión computarizada**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Mujeres	0	11	2.84	2.824	1158	0.435
Varones	0	37	4.36	4.489	3.175	18.146
7-10 años	0	37	5.48	5.101	3.229	17.572
11-15 años	0	20	3.74	3.344	1.886	5.995
16-22 años	0	14	2.23	2.580	1.880	4.741
TDAH-DA	0	15	3.58	3.386	1.134	1.054
TDAH-C	0	37	5.10	5.769	3.338	15.672
TAPZ	0	14	4.25	3.587	0.721	-0.262
CONTROL	0	8	2.32	2.021	1.309	1.341
Total	0	37	3.81	4.074	4.127	25.810

**14.5 Discriminación del tipo de error cometido en la tare PC**

	<i>Mín.</i>	<i>Max.</i>	<i>M</i>	<i>DE</i>	<i>Asimetría</i>	<i>Curtosis</i>
Impresiones	0	20	2.38	2.540	2.318	10.155
Error congruente	0	17	1.44	2.115	3.129	14.825

**14.7 Normalidad y Homocedasticidad para los errores en PC**

		Prueba de normalidad		Homogeneidad de la varianza	
		Kolmogorov-Smirnov <sup>a</sup>	<i>p</i>	Estadístico de Levene	<i>p</i>
Sexo	Mujeres	.210	.000**	4.719	.031*
	Varones	.168	.000**		
Edad	7-10	.153	.000**	6.623	.000**
	11-15	.187	.000**		
	16-22	.222	.000**		
Grupo	TDAH-DA	.202	.000**	6.122	.001**
	TDAH-C	.192	.000**		
	TAPZ	.136	.000**		
	CONTROL	.229	.000**		

<sup>a</sup> Corrección de significación de Lilliefors. *p*-valor: Diferencia significativa en \*\* $p < .01$  y \* $p < .05$ . Diferencia no significativa en  $p > .05$  (i.e. para distribuciones similares a la curva normal y varianzas homogéneas)

## Anexo 15

## Análisis de varianza factorial o univariante para las mediciones de errores en las tareas de la versión computarizada

## 15.1 Pruebas de efectos intersujetos para la tarea P de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.918	.000	.294
Intersección	1	108.618	.000	.335
Sexo	1	3.082	.081	.014
Edad	2	18.070	.000	.143
Grupo <sup>b</sup>	3	3.767	.012	.050
Sexo × Edad	2	.895	.410	.008
Sexo × Grupo	3	.390	.761	.005
Edad × Grupo	6	1.047	.396	.028
Sexo × Edad × Grupo	6	.474	.828	.013

<sup>a</sup>  $R^2 = .294$  ( $R^2$  ajustada = .219).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.

## 15.2 Pruebas de efectos intersujetos para la tarea C de la versión computarizada

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.366	.000	.264
Intersección	1	84.985	.000	.282
Sexo	1	.527	.469	.002
Edad	2	15.116	.000	.123
Grupo <sup>b</sup>	3	4.162	.007	.055
Sexo × Edad	2	.106	.899	.001
Sexo × Grupo	3	.459	.711	.006
Edad × Grupo	6	2.171	.047	.057
Sexo × Edad × Grupo	6	.119	.994	.003

<sup>a</sup>  $R^2 = .264$  ( $R^2$  ajustada = .185).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.  
 $\eta^2$  Eta parcial al cuadrado.



**15.2 Pruebas de efectos intersujetos para la tarea PC de la versión computarizada**

	<i>gl</i>	<i>F</i>	<i>p</i>	$\eta^2$
Modelo corregido <sup>a</sup>	23	3.348	.000	.263
Intersección	1	170.705	.000	.441
Sexo	1	6.162	.014	.028
Edad	2	11.535	.000	.096
Grupo <sup>b</sup>	3	3.583	.015	.047
Sexo × Edad	2	.185	.831	.002
Sexo × Grupo	3	.756	.520	.010
Edad × Grupo	6	1.092	.368	.029
Sexo × Edad × Grupo	6	1.320	.249	.035

<sup>a</sup>  $R^2 = .263$  ( $R^2$  ajustada = .184).

<sup>b</sup> Grupos diagnósticos (TDAH-DA, TDAH-C, TAPZ) y controles.

$\eta^2$  Eta parcial al cuadrado.

## Anexo 16

## Índices de interferencia en el Stroop computarizado

16.1 Descriptivos generales ( $n=240$ ) de los índices de interferencia calculados

	Mín.	Máx.	M	DE	Asimetría	Curtosis
$I_D=PC-C^a$	-8.00	613.00	205.487	126.542	1.038	1.324
$I_R=C/PC^a$	0.57	1.01	0.796	0.095	-0.247	-0.300
$I_{READ}=[(PC-P)/P]^a$	-0.17	0.67	0.152	0.174	0.892	0.904
$I_{NAME}=[(PC-C)/C]^a$	-0.01	0.76	0.275	0.165	0.890	0.750
$I_V=PC-[(P+C)/2]^b$	-3.00	11.50	1.131	2.274	1.767	5.211

<sup>a</sup> El cálculo se realiza con la puntuación de latencias medias de aciertos

<sup>b</sup> El cálculo se realiza con la puntuación de errores

## 16.2 Comparación de medias de los índices calculados en el Stroop computarizado por rango etario

	7 a 10	11 a 15	16 a 22	Estadístico
$I_D$	388.3 (255.3)	243.6 (131.6)	205.4 (126.5)	$X^2(2)=22.652$ ( $p=.000$ ) **
$I_R$	0.76 (0.10)	0.78 (0.08)	0.79 (0.09)	$X^2(2)=2.226$ ( $p=.110$ ) ns
$I_{READ}$	0.18 (0.32)	0.15 (0.12)	0.15 (0.17)	$X^2(2)=0.528$ ( $p=.590$ ) ns
$I_{NAME}$	0.33 (0.18)	0.29 (0.14)	0.27 (0.16)	$X^2(2)=2.220$ ( $p=.111$ ) ns
$I_V$	0.55 (3.81)	1.25 (2.70)	1.13 (2.27)	$X^2(2)=1.223$ ( $p=.296$ ) ns

M (DE)

Diferencia significativa: \*\* ( $p<.01$ ) y \* ( $p<.05$ ) Diferencia no significativa: ns ( $p>.05$ )

## 16.3 Comparación de medias de los índices calculados en el Stroop computarizado por grupo diagnóstico

	TDAH-DA	TDAH-C	TAPZ	CONTROL	Estadístico
$I_D$	292.1(243.1)	302.0(220.3)	270.2(157.1)	252.1(152.6)	$X^2(3)=0.772$ ( $p=.511$ ) ns
$I_R$	0.78 (0.11)	0.77 (0.09)	0.78 (0.08)	0.77 (0.08)	$X^2(3)=0.274$ ( $p=.844$ ) ns
$I_{READ}$	0.12 (0.17)	0.18 (0.31)	0.14 (0.21)	0.19 (0.15)	$X^2(3)=1.109$ ( $p=.346$ ) ns
$I_{NAME}$	0.29 (0.20)	0.30 (0.16)	0.29 (0.14)	0.30 (0.15)	$X^2(3)=0.153$ ( $p=.928$ ) ns
$I_V$	0.42 (3.26)	1.12 (3.18)	1.35 (3.48)	1.00 (1.76)	$X^2(3)=1.049$ ( $p=.371$ ) ns

**16.4 Capacidad de clasificación de los índices de interferencia por rango etario**

Índice	Edad	Corte	Sensibilidad	Especificidad	AUC
I <sub>D</sub>	7 a 10	343.50	.50	.85	.674*
	11 a 15	325.50	.23	.85	.489
	16 a 22	225.50	.40	.70	.437
I <sub>R</sub>	7 a 10	0.70	.35	.90	.587
	11 a 15	0.65	.07	.99	.444
	16 a 22	0.81	.56	.45	.409
I <sub>READ</sub>	7 a 10	0.25	.38	.85	.520
	11 a 15	0.44	.02	.99	.439
	16 a 22	0.62	.02	.99	.308
I <sub>NAME</sub>	7 a 10	0.14	.15	.85	.417
	11 a 15	0.15	.20	.95	.557
	16 a 22	0.14	.30	.90	.597
I <sub>V</sub>	7 a 10	3.75	.21	.85	.396
	11 a 15	2.25	.37	.99	.662*
	16 a 22	2.75	.15	.90	.427

AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90

**16.5 Descriptivos de las mediciones básicas en la tarea PC del Stroop computarizado (n=240)**

	Mín.	Máx.	M	DE	Asimetría	Curtosis
Latencias media de aciertos	598	2643	1201.58	402.660	1.455	1.976
Aciertos	35	72	68.24	3.977	-3.232	20.678
Errores <sup>a</sup>	0	37	3.81	4.074	4.127	25.810

<sup>a</sup> Número total de errores

**16.6 Capacidad de clasificación de las mediciones básicas en la tarea PC por rango etario**

	Edad	Corte	Sensibilidad	Especificidad	AUC
Latencias media de aciertos	7 a 10	1463	.65	.85	.796**
	11 a 15	1088	.86	.35	.577
	16 a 22	987	.41	.75	.530
Aciertos	7 a 10	67.50	.56	.75	.685*
	11 a 15	68.50	.52	.95	.776**
	16 a 22	69.50	.35	.80	.498
Errores <sup>a</sup>	7 a 10	4.50	.56	.75	.689*
	11 a 15	3.50	.51	.95	.779**
	16 a 22	2.50	.35	.80	.499

AUC: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90

<sup>a</sup> Número total de errores

**16.7 Capacidad de clasificación de las mediciones básicas en la tarea PC por sexo**

	Sexo	Corte	Sensibilidad	Especificidad	AUC
Latencias media de aciertos	Mujeres	1199.50	.43	.82	.613*
	Varones	1520.50	.25	.99	.552
Aciertos	Mujeres	67.50	.33	.99	.600*
	Varones	69.50	.65	.59	.628*
Errores <sup>a</sup>	Mujeres	2.50	.64	.59	.629*
	Varones	3.50	.37	.92	.603*

## Anexo 17

Análisis de correlaciones entre las versiones del Stroop ( $n=240$ )17.1 Cocientes  $r$  de Pearson para las mediciones de latencias en las tres tareas

		Versión tradicional		
		P	C	PC
Versión computarizada	P	.740**	.708**	.603*
	C	.564*	.739**	.670**
	PC	.511*	.741**	.682**

Magnitud del coeficiente: \*\*\*muy alta  $>.80$ , \*\*alta  $.61$  a  $.80$ , \*moderada  $.41$  a  $.60$ , baja  $<.40$

17.2 Cocientes  $r$  de Pearson para las mediciones de aciertos en las tres tareas

		Versión tradicional		
		P	C	PC
Versión computarizada	P	.415*	.368	.380
	C	.359	.258	.317
	PC	.336	.279	.311

17.3 Cocientes  $r$  de Pearson para las mediciones de errores en las tres tareas

		Versión tradicional		
		P	C	PC
Versión computarizada	P	.068	.132	.161
	C	.068	-.011	.089
	PC	.045	.049	.114
Imprecisiones		.067	.006	-.023
Errores congruentes		.006	.067	.097

Coefficientes con signo negativo indican relación inversamente proporcional.

#### 17.4 Cocientes $r$ de Pearson para los índices de interferencia calculados en ambas versiones

		Versión tradicional				
		$I_D$	$I_R$	$I_G$	$I_{REL}$	$I_V$
Versión computarizada	$I_D$	-.223	-.063	-.031	.063	.040
	$I_R$	.089	.010	.066	-.012	.046
	$I_{READ}$	.060	-.085	-.178	.084	.028
	$I_{NAME}$	-.102	.000	-.056	.002	-.059
	$I_V$	.012	.036	-.008	-.036	-.010

Magnitud del coeficiente: \*\*\*muy alta  $>.80$ , \*\*alta  $.61$  a  $.80$ , \*moderada  $.41$  a  $.60$ , baja  $<.40$

Correlaciones con signo negativo indican relación inversamente proporcional.

#### 17.5 Cocientes $r$ de Pearson entre las tareas del Stroop tradicional y sus correspondientes índices de interferencia

		$I_D$	$I_R$	$I_G$	$I_{REL}$	$I_V$
	P	-.410*	.176	.365	-.175	.085
Latencia de respuesta	C	-.614**	.069	.154	-.068	.161
	PC	-.092	-.543*	-.351	.545*	.036
	P	.439*	.085	-.254	-.086	-.180
Número de aciertos	C	.659**	.022	-.084	-.023	-.139
	PC	.070	.589*	.454*	-.590*	-.302
	P	-.119	-.005	.017	.005	-.062
Número de errores	C	-.212	-.012	-.076	.013	.173
	PC	.058	-.296	-.219	.296	.940***

**17.6 Cocientes  $r$  de Pearson entre las tareas del Stroop computarizado sus correspondientes índices de interferencia**

		$I_D$	$I_R$	$I_G$	$I_{REL}$	$I_V$
	P	.471*	-.141	-.314	.172	-.068
Latencia de respuesta	C	.426*	-.026	.014	.059	-.135
	PC	.782**	-.434*	.214	.472*	-.064
	P	-.305	.131	.039	-.155	.342
Número de aciertos	C	-.231	.125	.071	-.136	.149
	PC	-.317	.242	.006	-.252	-.481*
	P	.305	-.130	-.040	.153	-.342
	C	.231	-.125	-.071	.136	-.149
Número de errores	PC	.320	-.246	-.004	.255	.493*
	PCImpr.	.248	-.152	-.071	.174	.315
	PCErr. Congr.	.302	-.264	.052	.260	.521*

Magnitud del coeficiente: \*\*\*muy alta  $>.80$ , \*\*alta  $.61$  a  $.80$ , \*moderada  $.41$  a  $.60$ , baja  $<.40$

Correlaciones con signo negativo indican relación inversamente proporcional.

Tipo de error en PC: imprecisiones (Impr.) y errores congruentes (Err.Cong.)

## Anexo 18

**Comparación general del desempeño en las versiones tradicional y computarizada  
entre los grupos clínicos (TDAH-DA, TDAH-C y TAPZ) y el grupo control**

**18.1 Latencias**

		Versión tradicional	Versión computarizada
Grupos clínicos	P	675.45 (540.88)	1104.56 (452.69)
	C	888.02 (285.16)	951.21 (295.57)
	PC	1521.16 (540.57)	1239.36 (428.32)
Grupo control	P	570.90 (262.85)	922.07 (254.37)
	C	789.08 (241.06)	837.99 (189.73)
	PC	1256.03 (386.97)	1090.71 (287.71)

**18.2 Aciertos**

		Versión tradicional	Versión computarizada
Grupos clínicos	P	81.91 (25.92)	68.56 (4.49)
	C	53.41 (15.24)	68.76 (4.78)
	PC	29.83 (10.97)	67.76 (4.34)
Grupo control	P	86,78 (22,41)	70.57 (1.67)
	C	59,75 (13,28)	70.80 (1.20)
	PC	37,63 (10,39)	69.68 (2.02)

**18.3 Errores**

		Versión tradicional	Versión computarizada
Grupos clínicos	P	0.55 (1.49)	3.44 (4.49)
	C	1.62 (1.68)	3.24 (4.79)
	PC	3.12 (3.61)	4.31 (4.40)
Grupo control	P	0.18 (0.43)	1.42 (1.68)
	C	0.70 (0.89)	1.20 (1.20)
	PC	1.17 (1.17)	2.32 (2.02)



**18.4 Índices y capacidad de clasificación**

		Grupos clínicos	Grupo control	<i>AUC</i> <sup>a</sup>
Versión tradicional	I <sub>D</sub>	23.57 (9.18)	22.08 (8.86)	.505
	I <sub>R</sub>	0.56 (0.14)	0.63 (0.13)	.598
	I <sub>G</sub>	-0.97 (6.80)	4.06 (6.13)	.641*
	I <sub>REL</sub>	44.20 (13.56)	36.52 (13.24)	.669*
	I <sub>V</sub>	2.03 (3.33)	0.73 (1.22)	.618*
Versión computarizada	I <sub>D</sub>	288.15 (209.27)	252.13 (152.64)	.529
	I <sub>R</sub>	0.78 (0.10)	0.78 (0.09)	.482
	I <sub>READ</sub>	0.15 (0.24)	0,19 (0.16)	.419
	I <sub>NAME</sub>	0.30 (0.17)	0,30 (0.15)	.521
	I <sub>V</sub>	0.97 (3.32)	1.01 (1.77)	.492

<sup>a</sup> Resultado del análisis de curvas ROC.

Valor *AUC*: malo <0.60, \* moderado .60 a .75, \*\* bueno .75 a .90 y \*\*\* muy bueno >.90

**18.4 Mediciones brutas en la tarea PC y capacidad de clasificación**

		Grupos clínicos	Grupo control	<i>AUC</i> <sup>a</sup>
Versión tradicional	Latencia estimada de respuesta	1521.16 (540.57)	1256.03 (386.97)	.664*
	Número de Aciertos	29.83 (10.97)	37,63 (10,39)	.701*
	Número de Errores	3.12 (3.61)	1.17 (1.17)	.688*
Versión computarizada	Latencia media de aciertos	1239.36 (428.32)	1090.71 (287.71)	.590
	Número de Aciertos	67.76 (4.34)	69.68 (2.02)	.641*
	Número de Errores	4.31 (4.40)	2.32 (2.02)	.643*