



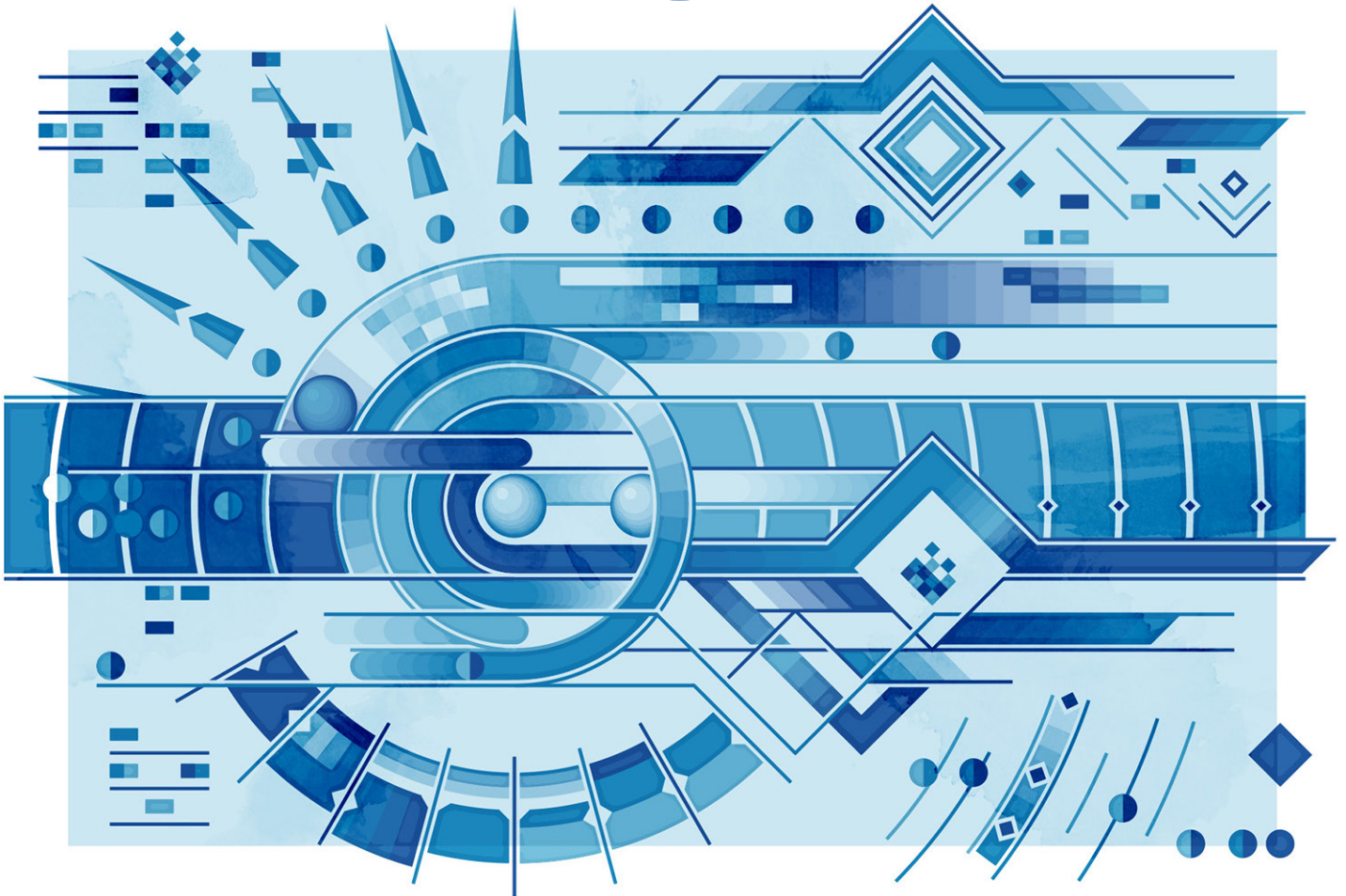
EVALUAR

Revista Evaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología de la Universidad Nacional de Córdoba

2021

VOL 21 - Nº1
ISSN 1667-4545



Conducta financiera del venezolano: Una escala para contextos de adversidad económica

Venezuelan Financial Behavior: A Scale for Economic Adversity Contexts

Antonio Martins * ¹, Luis Rodríguez ²

1 - Escuela de Psicología, Universidad Católica Andrés Bello, Caracas, Venezuela.

2 - Centro de Investigación y Evaluación Institucional, Universidad Católica Andrés Bello, Caracas, Venezuela.

Introducción
Metodología
Resultados
Discusión
Referencias

Recibido: 18/09/2021 Revisado: 27/11/2021 Aceptado: 08/12/2021

Resumen

La conducta financiera es entendida como cualquier acción que implique el uso del dinero. Estos comportamientos pueden variar en función de la realidad económica de los países, siendo los territorios con dificultades financieras los que imponen mayores retos en su medición. Se planteó como objetivo desarrollar una escala de conducta financiera y analizar sus propiedades psicométricas. Los ítems fueron elaborados con la colaboración de jueces expertos, y las dimensiones de la conducta financiera fueron: (a) *conducta de ahorro*, (b) *manejo de deuda* y (c) *manejo del dinero*. La muestra estuvo conformada por 780 participantes (59.74% femeninos) caraqueños con edades comprendidas entre los 18 y 70 años, provenientes de distintos sectores ocupacionales. Se obtuvo una adecuada consistencia interna del instrumento en la mayoría de los factores. Además, se halló evidencia de validez externa al test. Se muestran en el estudio diferencias en las tres dimensiones en función de variables sociodemográficas.

Palabras clave: *conducta financiera, teoría clásica de los test, ahorro, manejo de la deuda, psicología económica*

Abstract

Financial conduct is understood as any action that involves the use of money. These behaviors may vary depending on the economic reality of the countries, being the territories with financial difficulties those that pose greatest challenges in their measurement. The objective was to develop a financial behavior scale and analyze its psychometric properties. The items were prepared with the collaboration of expert judges and the dimensions of financial behavior were: (a) *savings behavior*, (b) *debt management* and (c) *money management*. The sample consisted of 780 participants (59.74% female) from Caracas aged between 18 and 70 years, from different occupational sectors. An adequate internal consistency of the instrument was obtained in most of the factors. In addition, evidence of external validity to the test was found. Differences in the three dimensions depending on sociodemographic variables are shown in the study.

Keywords: *financial behaviour, classical test theory, savings, debt management, economic psychology*

*Correspondencia a: Antonio Martins. Dirección postal: 1020. Teléfono de contacto: +584128104751. Correo electrónico: antoniomartins422@gmail.com

Introducción

En la actualidad existe un gran interés por conocer con mayor detalle cómo las personas piensan y se comportan respecto a sus recursos económicos. Las explicaciones de dichos comportamientos se remontan al desarrollo teórico de la economía en el siglo XVIII; Adam Smith, el padre de la economía clásica, ya realizaba para aquel entonces análisis económicos en los que se consideraban conceptos conductuales (Laibson & List, 2015).

Sin embargo, la aproximación de los economistas se basó en la noción del *homo economicus*, que proponía que los seres humanos han de comportarse económicamente de una manera racional. No obstante, con el paso del tiempo y los avances en la investigación, estos postulados fueron perdiendo popularidad progresivamente. En parte, los estudios del psicólogo Daniel Kahneman (2013) contribuyeron a un cambio de paradigma, a partir del cual se formó la denominada economía conductual y la psicología económica (Kirchler & Hoelzl, 2018).

Esta última es entendida como una rama de la psicología aplicada que busca estudiar la vida mental económica y el comportamiento de los sujetos, teniendo a su vez una estrecha vinculación con la economía, por lo que se puede considerar como un campo interdisciplinar en el que se busca explicar cómo se dan las tomas de decisiones económicas; entre estas, el ahorro, gasto, toma de riesgos financieros, así como sus diversas causas y consecuencias (Ranyard & De Mello-Ferrerira, 2017). Es dentro de esta área que se enmarca la conducta financiera, el constructo principal de la presente investigación.

Definición y teorías de la conducta financiera

Para definir el concepto de conducta financiera, primero es conveniente señalar que no hay un consenso en las teorías, motivo por el cual es necesario precisar: ¿qué es una conducta? La conducta se entiende como la actividad física de un organismo ante estímulos externos o internos; que incluye actividades objetivamente observables, actividades introspectivas observables y procesos no conscientes (VandenBos, 2015). De aquí es importante resaltar que es una acción con algún parámetro observable.

Partiendo de este marco interpretativo, Xiao (2008) define la conducta financiera como cualquier acción humana que es relevante en el manejo de dinero, siendo sus dimensiones aquellas referidas a la administración de efectivo, del crédito (deuda) y del ahorro. También abarca la adquisición de seguros, la planificación de retiro, la inversión y el pago/planificación de impuestos (Van Raaij, 2016).

Existen múltiples teorías en el área de la psicología económica desarrolladas con el fin de explicar la conducta financiera. A continuación, se presentan algunas de estas propuestas, sin pretensión de exponer todas las teorías existentes ni ser extensivo en su tratamiento.

Modelo de la socialización financiera familiar: sostiene que el proceso de interacción con el entorno social es fundamental para el desarrollo de conductas financieras efectivas. Las relaciones entre las características del grupo familiar, los factores sociodemográficos (edad, ingreso y sexo) y los antecedentes personales generan los resultados de la socialización financiera. Estos resultados son actitudes, conocimientos y capacidades económicas que influyen sobre la conducta financiera, y estas, a su vez, sobre el bienestar monetario (Gudmunson & Danes, 2011).

Es importante aclarar sobre este modelo que, si bien dichos procesos inician en la infancia, en la que se adquieren repertorios y se desarrollan patrones de comportamiento financiero, es durante la adultez cuando podría ocurrir una modificación en dichos patrones, en especial en periodos o eventos de grandes cambios (Gudmunson, Ray, & Xiao, 2016). Esto denota que la conducta financiera es de carácter dinámico, ya que puede modificarse en función del contexto económico.

Modelo de Van Raaij: es una propuesta caracterizada por las relaciones entre las variables psicológicas y económicas. Presenta cuatro factores principales que de forma cíclica dan cuenta de la dinámica de la conducta financiera: el ambiente económico (E) influye en el ambiente percibido (E/P), y este, a su vez, sobre la conducta económica (B), que impactará en el bienestar subjetivo (SW), el cual termina afectando al ambiente económico. Adicionalmente hay variables externas que interactúan de forma individual con cada una de estas, siendo correspondientemente el contexto general en E, el factor personal en E/P (edad, sexo e ingreso), el factor situacional en B y el descontento social en SW (Van Raaij, 1981).

De esta manera, la conducta no es solo una consecuencia (del E/P) sino que puede ser también una causa (del SW y del E). Los modelos de socialización financiera y el de Van Raaij permiten dar cuenta de la explicación de la conducta, pero no de su composición; por lo tanto, se consideran como dimensiones las expuestas por Xiao (2008) y Van Raaij (2016).

Formas de medición de la conducta financiera

Se ha advertido que al momento de trabajar con este tipo de conducta se deben precisar ciertos puntos: (a) decidir si trabajar con la conducta

propriadamente o con los productos de la conducta; (b) decidir si trabajar con una conducta particular o generar una categoría de agrupación; (c) definir los elementos de la conducta, tales como la acción, el objetivo, el contexto y el tiempo de duración; (d) decidir si medir mediante autoinforme u observación, y, por último, (e) establecer la medida de conducta, por ejemplo, de forma binaria, categoría o frecuencia. Sobre esta última se aclara que debe ser elegida en función de las preguntas (Xiao, 2008).

Sobre el autoinforme se debe ser cuidadoso, ya que problemas como la deseabilidad social o aquiescencia de respuesta generan que los sujetos distorsionen sus planteamientos, llegando a mentir (Magnusson, 1972). Sin embargo, a veces es la única opción realizable, dado que hay conductas que no se pueden observar directamente.

Consecuentemente con tales planteamientos, se ha advertido que existen formas de disminuir los riesgos que suponen los autoinformes, Mcnair y Crozier (2017) recomiendan invertir algunos ítems o mantener estandarizada la situación de evaluación para evitar este tipo de inconvenientes.

Trabajos previos de medición del constructo

La mayoría de las investigaciones, cuando trabajan con conducta financiera como variable, no cuentan con escalas psicométricamente adecuadas, ya que suelen ser mediciones ad hoc o instrumentos que no informan indicadores psicométricos. Por ende, en la presente investigación se seleccionaron como antecedentes aquellas escalas de conducta financiera que sí cuentan con un mínimo de datos psicométricos (ver Tabla 1).

Tabla 1
Comparación de las escalas de conducta financiera.

Escala	Factores	Confiabilidad	# de ítems	N	Autores y año
Escala de Conducta de Gestión Financiera (ECCF)	Ahorro e inversión	$\alpha = .78$	5	1011	Dew y Xiao (2011)
	Adquisición de seguro	$\alpha = .73$	3		
	Manejo de dinero	$\alpha = .63$	4		
	Crédito	$\alpha = .57$	3		
Habilidades de Gestión del Dinero (EHGD)	General (ad hoc)	$\alpha = .86$	9	271	Garðarsdóttir y Dittmar (2012)
Escala Breve de Gestión del Dinero (EBGD)	Ahorro	$\alpha = .89$	4	1078	Ksendzova, Donnelly y Howell (2017)
	Adquisición de seguro	$\alpha = .82$	5		
	Crédito	$\alpha = .80$	5		
	Manejo de dinero	$\alpha = .79$	4		

Cabe destacar que la EHGD trabaja las conductas de gestión a nivel general y, si bien informa una confiabilidad elevada, no cuenta con indicadores de validez de contenido y criterio. En otro orden de ideas, [Porter y Garman \(1993\)](#) plantean un conjunto de ítems donde aparte de las dimensiones de manejo de dinero, crédito y riesgo (adquisición de seguro), dividen la dimensión *ahorro* en *acumulación de capitales* (ahorro tradicional) y *jubilación/planificación de retiro*. También establecen una dimensión de gestión general que incluye conductas de planificación e impulsividad. Esta propuesta, aunque es una categorización que podría ser relevante a nivel heurístico, no cuenta con indicadores psicométricos que den cuenta de la utilidad del instrumento en distintas muestras.

Esto último invita a que el análisis se centre en la ECCF y la EBGD, ya que son muy similares. Respecto al manejo de crédito, se encuentra que en la segunda escala la confiabilidad para esta dimensión aumenta, lo que se puede deber no solo al aumento del número de ítems ([Magnusson, 1972](#)) sino también a la inclusión de conductas

generales como endeudamiento o mayor cantidad de egresos que de ingresos, los cuales no están supeditados necesariamente a las tarjetas de crédito, aspecto que sí ocurre en la ECCF.

En cuanto a la dimensión *ahorro*, ambas escalas tienen una confiabilidad similar. Igualmente, el porcentaje de varianza de esta dimensión es el mayor. Por otra parte, respecto a la dimensión *manejo de dinero*, es la que menor confiabilidad y porcentaje de varianza explica. Sin embargo, la EBGD tiene una confiabilidad mucho mayor que la ECCF. En otro orden de ideas, la ECCF cuenta con validez convergente adecuada al correlacionarla con los ítems de [Perry y Morris \(2005\)](#).

Respecto a su aplicación en otros países, la escala de ECCF fue validada en Brasil, tomando como muestra 195 estudiantes graduados y profesores de postgrado. En el análisis factorial confirmatorio se obtuvieron los cuatro factores de la escala original; asimismo el instrumento tomado de forma unidimensional resultó confiable (confiabilidad compuesta = .84). Respecto a la validez, la escala correlacionó negativamente con tensión

financiera y con compra impulsiva (Veiga, Avelar, Moura, & Higuchi, 2019).

Por último, en cuanto a las variables que se han relacionado en la literatura con la conducta financiera, se encuentra que la impulsividad de la compra correlaciona negativamente con ella (Dew & Xiao, 2011; Veiga et al., 2019), mientras que los hábitos de compra se asocian de forma positiva con la conducta positiva, dado que es un manejo responsable y cuidadoso del dinero (Van Raaij, 2016).

Crisis económica y conducta financiera

Una vez repasados brevemente los principios de la conducta financiera, se debe aclarar un punto importante sobre esta: los comportamientos financieros de las personas naturalmente se ven afectados por el contexto en el que se desenvuelven los individuos (Gärling, Kirchler, Lewis, & Van Raaij, 2009; Van Raaij, 1981). Es por ello que resulta necesario visualizar el contexto económico del venezolano durante las dos primeras décadas del siglo XXI. Lo primero a mencionar es que revisiones exhaustivas de la historia del país suelen concluir que la economía venezolana posee un carácter impredecible (Sardi, Angelucci, Martins, & Peña, 2020).

Para comprender las condiciones económicas actuales, resultan ilustrativos los hallazgos de la Encuesta Nacional de Condiciones de Vida (ENCOVI), que han reflejado longitudinalmente un empobrecimiento masivo de los venezolanos: mientras que para el año 2015 un 87% de los hogares se hallaba por debajo de la línea de la pobreza, para el año 2019 se estimó que este porcentaje aumentó a 96.2% de la población (UCAB, 2020).

Ante esta realidad, el gobierno nacional ha planteado una serie de políticas económicas que han redundado en el moldeamiento de las con-

ductas financieras que pueden adoptar los venezolanos. Lo más llamativo es que a partir del año 2018, se ha producido un viraje en cuanto al manejo de las divisas en el país. En este año el gobierno facilitó la libre convertibilidad de la moneda extranjera en toda Venezuela, involucrando en estas transacciones al Sistema Bancario Nacional (Olmo, 2019).

Relacionadas con las operaciones cambiarias y el proceso desarrollado por el gobierno para controlar el dólar paralelo, se implementaron una serie de políticas bancarias referidas al encaje legal. Solo desde el año 2018 al 2019 el encaje pasó de ser 21.5% a 100% (Gaceta Oficial 4443.040); esto se tradujo en que los bancos no pudieron otorgar créditos a empresas ni a particulares con el fin de no emitir mayor cantidad de bolívares sin sustento, así como para controlar la inflación (Barbar, Benasayag, & Salomón, 2019). Esto limitó el acceso y uso de tarjetas de crédito en la banca nacional.

Ambas políticas referidas al manejo de las divisas y a la asignación de créditos han implicado cambios en la manera en la que ahorran los venezolanos y en la forma en la que se aproximan a sus deudas mediante los préstamos formales e informales. De manera concreta, aunque la posibilidad de ahorro se ve reducida en esta población empobrecida, cuando esta aparece, se da a través de la acumulación de divisas (y no de bolívares, moneda de curso oficial que ha visto pulverizado su valor frente al dólar) o la obtención de bienes duraderos; mientras que la posibilidad de obtención de créditos se ha visto reducida a aquel pequeño porcentaje de la población que tiene acceso a la banca internacional.

Ante este contexto, resulta necesaria la elaboración de una escala que, aunque parta de las concepciones tradicionales en la medición de la conducta financiera (Dew & Xiao, 2011; Ksendzova et al., 2017) considere las particula-

ridades de las economías inestables tales como el uso de divisas para ahorrar, solicitud de deudas no convencionales (préstamos de familiares o amigos) y consumo impulsivo en moneda local; y proporcione de esta manera una propuesta operacional que resulte aplicable para situaciones que tengan como denominador común la adversidad económica, la cual no es abordada por las escalas tradicionales expuestas anteriormente.

En la literatura se ha encontrado que, en economías inestables, donde predominan la inflación y la recesión, las conductas financieras que adoptan los individuos son disímiles a las adoptadas por sujetos inmersos en economías estables. Por ejemplo, para el caso del ahorro, lo recomendable es que ante tasas inflacionarias elevadas las personas eviten ahorrar, ya que cualquier acumulación de dinero es superada por el aumento de los precios (Van Raaij, 2016). Por tanto, lo favorable resultaría ser el gasto rápido de la moneda local o el ahorro en bienes duraderos o en una moneda con mayor estabilidad.

Con respecto a la deuda, propuestas operacionales como las de Dew y Xiao (2011) consideran que pagar la totalidad de las deudas antes de su vencimiento resulta adaptativo. No obstante, en contextos inflacionarios, lo más eficiente sería acumular la mayor cantidad de deuda posible en la moneda devaluada y retrasar el pago en el tiempo, dado que la tasa de inflación será superior al interés de la deuda.

Por último, en cuanto al manejo del dinero en este tipo de situaciones económicas excepcionales, las personas tienden a realizar un ajuste en su patrón de consumo con el objetivo de mantener el valor de sus ingresos. Por ejemplo, los sujetos ante la adversidad buscan comprar los productos más baratos, de mayor calidad o venden bienes duraderos (Gärling et al., 2009).

Partiendo de estos hallazgos atípicos, que indican que situaciones culturales o contextuales

requieren de una forma de medición del constructo que sea diferente a las que se tienen, se ha considerado necesario construir un test (Fernández, Pérez, Alderete, Richaud, & Liporace, 2010). Dicha construcción está enmarcada en los postulados de la teoría clásica de los test (De Minzi, 2008; Magnusson, 1972), por lo que en la presente investigación se plantean los siguientes objetivos:

1) Construir un instrumento con adecuada validez de contenido para la medición de la conducta financiera en contextos de adversidad económica.

2) Evaluar la estructura factorial del test mediante la realización de un análisis factorial exploratorio, que dilucide si existe una medición de la conducta financiera en términos de las dimensiones constituyentes, útiles para contextos adversos: (a) ahorro, (b) manejo de la deuda y (c) manejo del dinero.

3) Obtener evidencia de confiabilidad mediante el método alfa de Cronbach que evalúa la consistencia interna de las puntuaciones del test.

4) Obtener evidencia de validez externa del test a través de procedimientos de convergencia y divergencia.

Metodología

Tipo y diseño de investigación

En cuanto al grado de control que existe sobre las variables sometidas a estudio, la investigación fue de tipo cuantitativa no experimental (Kerlinger & Lee, 2002), tomando en consideración que no existe una manipulación directa de las variables. Por otro lado, el diseño fue de tipo instrumental (Montero & León, 2007), ya que la investigación tuvo como objetivo el desarrollo de una escala y la obtención de sus propiedades psicométricas.

Participantes

La muestra estuvo conformada por 780 participantes (59.74% femeninos, 40.26% masculinos) oriundos de la Gran Caracas y que fueron seleccionados por medio de un muestreo no probabilístico de tipo intencional, ya que no se consideraron técnicas aleatorizadas al momento de elegir a los sujetos. La edad promedio de los encuestados fue de 32 años, con una desviación estándar de 15 años. Con respecto a su ocupación, 28.72% eran trabajadores del sector privado, 25.77% estudiantes de universidades privadas, 17.69% personal de empresas públicas, 14.74% estudiantes de universidades públicas y un 13.08% se describió como “patronos, empleadores o trabajadores por cuenta propia”. Esta distribución de la muestra permitió conocer que un 72.95% de la muestra mantenía un trabajo, un 19.23% recibía remesas y un 44.6% poseía tarjeta de crédito al momento de la recolección de los datos.

Instrumentos

Cuestionario de datos sociodemográficos. Este cuestionario fue elaborado con la finalidad de facilitar la identificación de los sujetos en función del sexo, la edad, la posibilidad de mantener un trabajo actual y la ocupación que mantenían al momento de la encuesta. También se indagó si los participantes disponían de préstamos o tarjetas de crédito al momento de la investigación. Fue en este documento que además se solicitó el consentimiento para participar en el estudio, al tiempo que se informó del anonimato de las respuestas. No se incluyó el nivel de ingreso, ya que estudios nacionales (ENCOVI) proponen evitar esta medida dada la resistencia de las personas para declararlos, y, por tanto, se toma la ocupación como un indicador indirecto.

Escala de Conducta Financiera (Martins & Rodríguez, 2020). Escala creada en contextos de formación académica para un curso básico de psicometría de la Universidad Católica Andrés Bello (UCAB). El instrumento está conformado por 24 reactivos (8 ítems por factor) cuyo formato de respuesta fue tipo Likert de 4 opciones que van desde *Nunca* a *Siempre*, con la finalidad de equipar las respuestas dadas ante todos los instrumentos del presente estudio. En la construcción de la tabla de especificaciones, que sirvió de fundamento para el proceso de validación de contenido por medio de la valoración de jueces expertos, se estimó la existencia de tres dimensiones de la conducta financiera: (a) ahorros, (b) manejo de la deuda y (c) manejo del dinero. A mayor puntaje en cada uno de los componentes se estima que mayor será la tendencia al ahorro, la inclinación hacia el uso de créditos formales o informales y la inclinación hacia la planificación racional del dinero respectivamente.

Escala de Hábitos de la Compra y Conductas de Consumo (Denegri, Palavecinos, Ripoll, & Yáñez, citado en Ortega & Rodríguez-Vargas, 2004). El instrumento cuenta con 19 reactivos cuyo formato de respuesta es tipo Likert de 4 puntos, que va desde *Nunca* hasta *Siempre*. En una muestra de adultos colombianos provenientes de la ciudad de Barranquilla, con un nivel socioeconómico medio y alto, las puntuaciones de la escala presentaron una estructura de tres dimensiones, cuyos coeficientes de consistencia interna resultaron adecuados para fines diagnósticos y de investigación. Estos tres factores extraídos por Ortega y Rodríguez-Vargas (2004) fueron: (a) calidad del producto ($\alpha = .81$), (b) planificación de la compra ($\alpha = .80$) y (c) uso responsable del crédito ($\alpha = .85$). En la muestra de la presente investigación, se obtuvo una consistencia interna adecuada para fines de investigación en la escala total ($\alpha = .72$).

Por otro lado, se realizó un análisis factorial exploratorio con mínimos cuadrados no ponderados (ULS robusto) para matrices policóricas con rotación oblimin directa, que resultó en la extracción de dos factores que explicaron el 42.64% de la varianza total del test y que fueron denominados como: (a) planificación de la compra ($\alpha = .79$) y (b) uso responsable del crédito y calidad del producto ($\alpha = .76$). A mayor puntaje en cada una de las dimensiones se estima que mayor será la planificación de la compra, el uso responsable del crédito y mejor será la calidad de los productos comprados respectivamente.

Escala de Impulsividad de la Compra (Luna-Arocas, 1998). La escala se encuentra conformada por 17 ítems con un formato de respuesta tipo Likert de 4 puntos, que va desde *Nunca* hasta *Siempre*. En la presente muestra, para las puntuaciones de la escala total se obtuvo un coeficiente de consistencia interna elevado para fines de investigación y diagnóstico ($\alpha = .88$). Por su parte, en cuanto a la estructura factorial de las puntuaciones del test, un análisis factorial exploratorio ULS robusto para matrices policóricas con rotación oblimin directa, permitió extraer un único factor que explicó un 38.55% de la varianza total del instrumento. A mayor puntuación en la escala total, mayor será la impulsividad de la compra.

Procedimiento

Se diseñó una tabla de especificaciones que permitió precisar el contenido o la dimensión que pretendía medir cada uno de los reactivos, su formato de corrección (directo o inverso) y de respuesta. A partir de ella, los profesores de psicometría de la UCAB, en colaboración con sus respectivos estudiantes, procedieron a construir un banco de ítems que fue sometido al juicio

de tres expertos en psicometría, metodología y economía.

El juicio de los expertos se realizó ítem por ítem y se limitó a la consideración de: (a) la pertinencia de los reactivos, (b) su claridad y (c) su adecuación a la muestra. Una vez obtenido el juicio de expertos se procedió a promediar sus decisiones, y aquellos reactivos con baja valoración fueron modificados, siguiendo para ello las sugerencias cualitativas de los jueces.

Una vez obtenida la escala preliminar se procedió a administrar el instrumento para obtener los indicadores psicométricos en la muestra seleccionada. Junto a la escala de conducta financiera, se administró un instrumento de hábitos de compra y otro de impulsividad de la compra para obtener indicadores de validez convergente y divergente. Todas las escalas fueron administradas en formato electrónico a través de la plataforma Google Forms. El proceso de envío de los instrumentos fue realizado por los estudiantes bajo la supervisión de sus profesores y tras ser adiestrados por los investigadores.

Consideraciones éticas

Previo al proceso de aplicación de los instrumentos, se obtuvo el consentimiento explícito de los interesados para participar en la investigación. El anonimato de los sujetos involucrados, y la garantía de la participación voluntaria y sin coerción fueron otros aspectos éticos considerados, siguiendo las directrices de la Declaración de Helsinki (Manzini, 2000).

Análisis de datos

Los cálculos estadísticos necesarios para conseguir los objetivos de investigación fueron

realizados con los softwares estadísticos IBM SPSS 22 (IBM Corp., 2013) y FACTOR Analysis (Lorenzo-Seva & Ferrando, 2011). Para conocer la estructura factorial subyacente a la escala de conducta financiera, se estimaron varios análisis factoriales exploratorios (AFE) con el método de extracción de mínimos cuadrados no ponderados para matrices policóricas (RULS), con rotación oblimin directa y varimax normalizada. Esto se debió a que era la forma de extracción más adecuada frente a la ausencia de normalidad en algunos ítems —evaluada a partir de la asimetría y la curtosis de los reactivos—, al formato de respuesta Likert de cuatro puntos de la escala y a que evita los casos Heywood (Lloret-Segura, Ferreres-Traver, Hernández-Baeza, & Tomás-Marco, 2014). Previamente fueron evaluados y garantizados todos los supuestos requeridos para los AFE. El análisis paralelo, en conjunto con los indicadores de bondad de ajuste, fue empleado como técnica para retener las dimensiones.

En segundo lugar, una vez extraídos los factores subyacentes a las puntuaciones de la escala, se propuso evaluar la consistencia interna de las dimensiones obtenidas a partir del coeficiente alfa de Cronbach. Posteriormente, fueron estimadas correlaciones de Pearson entre las puntuaciones de las dimensiones de la conducta financiera y los indicadores de impulsividad de la compra y hábitos de consumo.

La correlación conducta financiera-impulsividad de la compra fue interpretada como un indicador de validez discriminante, mientras que las asociaciones entre conducta financiera y las dimensiones de los hábitos de la compra sugirieron un proceso de validación convergente, tomando en consideración la inexistencia de escalas previas que evaluaran la conducta financiera tal y como fue propuesta teóricamente en la presente investigación.

Por último, se procedió a realizar un análisis

exploratorio de datos a partir de las puntuaciones obtenidas en cada una de las dimensiones de la conducta financiera, comparando, el manejo de los ahorros, deuda y dinero en función del sexo, edad y ocupación de los participantes. Técnicas como la *t* de Student para muestras independientes, correlaciones momento-producto de Pearson y análisis de varianza de una vía (ANOVA) fueron necesarias para llevar a cabo dichas comparaciones.

Resultados

Análisis psicométricos

Para estudiar la estructura factorial del test, en primer lugar, se realizó un análisis paralelo que permitió estimar el número de factores a extraer en el ULS robusto con rotación oblimin directa. La cantidad de factores a extraer fue de tres, esto tomando en consideración que al comparar el porcentaje de varianza en los datos reales con la proporción del percentil 95, se halló que fue hasta la tercera dimensión donde se superó el criterio del azar.

Posteriormente, se procedió a calcular un nuevo AFE bajo los mismos parámetros, pero solicitando la extracción de tres factores con los 24 ítems iniciales. Bajo este cálculo se eliminaron aquellos reactivos cuyas cargas fueron menores a .30. En este sentido, fueron descartados tres ítems (ítem 6: *Ahorro gran porcentaje de dinero recibido en bolívares*, ítem 8: *Pago atrasado a mis amigos o allegados cuando me toca devolver el dinero*, ítem 15: *Utilizo una proporción de mis ingresos en bolívares para comprar bienes duraderos*); este criterio fue utilizado pese a las sugerencias de Lloret-Segura et al. (2014), dado que Hair, Black, Babin y Anderson (2019) indican que en muestras mayores a 350 sujetos pudiera reducirse el criterio de .40 a .30.

Con los 21 reactivos restantes, fue calculado un nuevo AFE con rotación oblimin directo. No obstante, tras la verificación de una baja correlación entre los factores extraídos se implementó la estimación de un nuevo análisis con rotación varimax normalizada, siguiendo las pautas de interpretación planteadas por Lloret-Segura et al. (2014): no se encontraron diferencias sustanciales en la distribución y monto de las cargas si se comparan las dos rotaciones.

En este último AFE, al igual que en los anteriores, se garantizó el cumplimiento de los supuestos iniciales. En este sentido, se encontró una alta interrelación entre las variables sometidas al análisis ($KMO = .84$), al tiempo que la matriz de correlaciones inicial no se ajustó a la matriz de identidad (test de esfericidad de Bartlett = 7017.70; $p < 0.001$). Por otra parte, los índices de bondad de ajuste demostraron niveles adecuados ($RMSEA = .045$; $GFI = .981$; $CFI = .978$).

Estos tres factores extraídos explicaban el 49.69% de la varianza total del test, y en la Tabla 2 se muestra la configuración de cada una de las dimensiones. El Factor 1, denominado *ahorro*, explicaba el 24.54%, la segunda dimensión, llamada *manejo de la deuda*, representa el 14.61% de la varianza y, por último, la dimensión de *manejo de dinero* explicó un 10.54%, y los autovalores para ellos fueron de 5.15, 3.06 y 2.21 respectivamente.

Con respecto a la consistencia interna de las dimensiones del instrumento, fue calculado el coeficiente alfa de Cronbach para los factores extraídos, obteniendo que la correlación ítem-test presentó un recorrido de -.29 hasta .61, estando la mitad de los reactivos por encima de .30 en su asociación con el test total. Para el caso de la dimensión de ahorro, se obtuvo una consistencia interna adecuada para fines de investigación y diagnóstico ($\alpha = .85$). En segundo lugar, la dimensión de manejo de la deuda resultó adecuada para fines de investigación ($\alpha = .69$). Por último, el factor de

manejo del dinero demostró una baja consistencia interna ($\alpha = .59$).

Además, se calculó una matriz de correlaciones (ver Tabla 3) con la finalidad de obtener evidencia de validez convergente y discriminante a partir de la relación de las puntuaciones obtenidas en las tres dimensiones de la conducta financiera con la impulsividad de la compra y los hábitos de esta. El manejo de la deuda y del dinero obtuvieron evidencia de validez discriminante al presentar correlaciones nulas o inversas con la impulsividad de la compra, e indicadores de validez convergente al relacionarse positiva y significativamente con todos los indicadores de los hábitos de consumo.

Por último, las conductas de ahorro se asociaron de forma positiva y baja con la planificación de la compra, siendo esto reflejo de la validez convergente mientras que el indicador de discriminación para dicha dimensión se mostró afectado al obtenerse una correlación positiva y moderada-baja entre el ahorro y la impulsividad de la compra.

Análisis adicionales

A nivel descriptivo, en las puntuaciones obtenidas por dimensiones de la conducta financiera, se halló que las conductas de manejo del dinero resultaron mayores ($M = 2.93$; $DE = .59$) que las presentes en ahorro ($M = 2.83$; $DE = .71$) y manejo de la deuda ($M = 2.51$; $DE = .68$). Para las conductas de ahorro, se obtuvo que la compra de monedas extranjeras se constituyó como el medio de ahorro más valorado, seguido por la compra abundante de productos que a posteriori pudieran ver aumentado su valor en bolívares.

Por su parte, para el manejo de la deuda, los participantes tendieron a valorar en mayor medida el pago retrasado de los intereses generados

Tabla 2
Matriz rotada.

	Ítems/Factores	1	2	3
1	Cuando cuento con dinero en efectivo, llevo mayor cantidad de dólares antes que bolívares	.609	.072	-.224
2	Con los bancos nacionales, pago los créditos que dispongo en el tiempo estipulado (Inverso)	.053	.706	-.098
3	Compro productos de una manera abundante porque sé que a largo plazo me saldrá más barato	.400	.199	.016
4	Tengo un límite para el monto de mis transacciones semanales	.008	.038	.404
5	Si el interés de mis deudas (con bancos o personas particulares) es en bolívares, pago en las fechas límites	.024	.347	.107
7	Administro mi dinero de forma racional, solo gasto en lo necesario	.017	.104	.788
9	Ahorro en moneda extranjera	.911	-.056	-.040
10	Junto con mi familia, planificamos el uso del dinero que obtenemos para un futuro	.388	.126	.369
11	Abono un poco más de lo requerido en los pagos mínimos de mi(s) tarjeta(s) de crédito	-.010	.681	.123
12	Ahorro dinero en moneda extranjera previniendo emergencias	.932	-.001	.071
13	Chequeo opciones de precios antes de decidir comprar algún producto	.099	.106	.400
14	Pido préstamos en bolívares para pagar algunos gastos inmediatos	.036	.349	-.111
16	Al momento de pagar cosas necesarias, dispongo del dinero para ellas	.337	.209	.074
17	Cuando quiero adquirir algo fuera del alcance de mi presupuesto considero los créditos (formales o informales) como una buena opción	.016	.549	-.038
18	Ahorro parte mi dinero en divisas para distintos fines (ej. educación, viajes, recreación)	.890	.033	-.036
19	Tengo detallados mis gastos mensuales	.115	.163	.527
20	Cuando tengo que pagar algo (con tarjeta de crédito o préstamos) procuro que el interés sea en bolívares	.068	.687	.166
21	Busco ahorrar lo que pueda en moneda extranjera pensando en el futuro	.911	.038	.048
22	Compro impulsivamente cosas que no necesito (Inverso)	-.216	-.030	.527
23	Si aumentaran el límite de mi tarjeta de crédito en bolívares lo gasto inmediatamente	.232	.521	-.192
24	Cuando mi grupo familiar dispone de bolívares y no tenemos un gasto pendiente, los incito a comprar monedas extranjeras	.637	.194	.066

por los créditos solicitados, siempre y cuando estos fueran adquiridos en bolívares y no en cualquier otra moneda.

Para el caso de la edad, se halló que los participantes con menor edad presentaban puntuaciones más elevadas en la dimensión de ahorro ($r = -.27$; $p < 0.001$), mientras que altas puntuaciones

en manejo de la deuda ($r = .43$; $p < 0.001$) o del dinero ($r = .22$; $p < 0.001$) se asociaban con una mayor edad del participante.

Con respecto a las comparaciones por sexo, solo se hallaron diferencias significativas en cuanto a las conductas de ahorro ($t = -5.44$; $gl = 778$; $p < 0.001$), siendo los hombres los que pre-

Tabla 3

Evidencia de validez externa del test.

Correlaciones	MDE	MDI	IC	PC	URC
Ahorro (A)	.16**	.05	.21**	.16**	.09*
Manejo de la Deuda (MDE)	1	.13**	.04	.20**	.45**
Manejo del Dinero (MDI)		1	-.23**	.46**	.31*
Impulsividad de la Compra (IC)			1	.11**	.05
Planificación de la Compra (PC)				1	.51**

Nota. URC = uso responsable del crédito y calidad del producto. Esta dimensión combina dos factores como resultado del análisis factorial de la presente investigación.

** La correlación de producto-momento de Pearson es significativa en el nivel .01 (2 colas).

sentaron mayores comportamientos ligados a la preservación de su patrimonio ($M = 3.00$; $DE = .73$) en comparación con las mujeres ($M = 2.72$; $DE = .63$). No obstante, el tamaño del efecto de esa diferencia fue baja (d de Cohen = .29).

Por último, todas las dimensiones de la conducta financiera presentaron diferencias en función de la ocupación del participante. Concretamente, para las conductas de ahorro ($F = 31.37$; $gl = 779$; $p < 0.001$), los patronos y trabajadores por cuenta propia tienden a presentar una puntuación significativamente mayor ($M = 3.09$) a la de los trabajadores privados o públicos ($M = 2.82$; $M = 2.30$), que son los grupos que manifestaron menor propensión al ahorro.

Por otro lado, para el caso del manejo de la deuda ($F = 28.32$; $gl = 779$; $p < 0.001$), los patronos y trabajadores por cuenta propia puntúan significativamente más alto en el acceso que tienen a dicha oportunidad si se les compara con los estudiantes de universidades privadas y públicas o con los trabajadores del sector privado. Resultados similares se encuentran en el manejo del dinero ($F = 2.81$; $gl = 779$; $p = .025$), ya que los patronos tienden a diferenciarse significativa-

mente si se les contrasta con el resto.

Para finalizar, tomando en consideración que el 54.6% de los sujetos no poseía tarjeta de crédito al momento del estudio producto de las políticas de encaje legal en el país, se procedió a extraer de la escala los ítems que hacían referencia explícita al uso de tarjetas de crédito (ítems 2, 11 y 23).

Lo anterior se hizo con la finalidad de verificar por medio de tres análisis factoriales exploratorios si el comportamiento de las cargas de la escala variaba en monto y factor al comparar tres grupos: personas que tenían tarjetas de crédito ($n = 348$), personas que no tenían ($n = 432$) y la muestra completa ($n = 780$). Se encontró que las cargas de los ítems se mantenían en los mismos factores, pero aumentaban ligeramente. Esto indica que al utilizar estos tres ítems se debe garantizar que la muestra sea homogénea en cuanto al acceso a tarjetas de crédito.

Discusión

Dado que los contextos de adversidad tien-

den a modificar las conductas financieras de los individuos y que además no se encontraron en la literatura propuestas operacionales que permitieran obtener indicadores contextualizados del comportamiento económico de las personas en situaciones problemáticas, se propuso la construcción de una escala que cumpliera con dichos fines. En primer lugar, se garantizó el proceso de validación de contenido del instrumento mediante las siguientes fases: (a) la revisión de la literatura para la familiarización con la conducta financiera como constructo, (b) la elaboración de una matriz de especificaciones que contemplara el contenido de los ítems y su formato de respuesta, (c) la construcción de un banco de ítems que contuvo el doble de afirmaciones estipuladas, (d) la evaluación del banco de ítems por parte de tres jueces expertos que emitieron una valoración en términos de pertinencia, claridad y adecuación de los ítems a la muestra; y (e) la elección de una versión preliminar de la escala que fue aplicada a una muestra piloto con el objetivo de obtener las propiedades psicométricas de las puntuaciones del test (Muñiz & Fonseca-Pedrero, 2019; Pérez & Olaz, 2008).

Con respecto al segundo objetivo de la investigación, se encontraron tres factores que fueron denominados (a) ahorro, (b) manejo de la deuda y (c) manejo del dinero. El primer factor se refiere a la abstención en la utilización del dinero en un determinado momento con la finalidad de gastarlo en un futuro. El segundo factor implica el préstamo de dinero por parte de una entidad financiera (o proveedores informales); evalúa cómo actúan los sujetos ante estas deudas. Por último, el resto de afirmaciones presentó un conjunto de acciones que denotaron la planificación y organización del uso que las personas le dan a su dinero en transacciones y pagos cotidianos (Dew & Xiao, 2011; Van Raaij, 2016).

Esta estructura factorial obtenida a partir de

la muestra del presente estudio mantuvo una semejanza con las propuestas operacionales de Dew y Xiao (2011) y Ksendzova et al. (2017). En paralelo, resultó ser una agrupación de dimensiones cercanas a explicar el 49.69% de la varianza total del test, condición sugerida en la interpretación de análisis factorial exploratorio (Pérez & Olaz, 2008).

En cuanto a la consistencia interna del instrumento, las puntuaciones de las dimensiones fueron adecuadas para fines de investigación siguiendo las sugerencias de la literatura (Prieto & Muñiz, 2008) en el caso del ahorro y el manejo de la deuda. Este planteamiento se asemeja a los hallazgos de la escala de Ksendzova et al. (2017) y se diferenció del instrumento original de Dew y Xiao (2011), el cual presentó inconvenientes en la fiabilidad de la dimensión referida al manejo del crédito.

No obstante, el presente instrumento halló para las puntuaciones de manejo del dinero una consistencia interna que según los parámetros de Prieto y Muñiz (2008) presenta ciertas fallas, siendo entonces una medida susceptible de ser mejorada. Resultados similares fueron informados para dicho factor en la escala de Dew y Xiao (2011), ya que, según los autores, esta propiedad de la conducta financiera tiende a poseer ítems de difícil redacción.

Tomando en consideración la estructura factorial del test y su fiabilidad, se eliminaron tres afirmaciones en la escala (ítems 6: *Ahorro gran porcentaje del dinero que recibo en bolívares*, 8: *Pago atrasado a mis amigos o allegados cuando me toca devolver el dinero*, y 15: *Utilizo una proporción de mis ingresos en bolívares para comprar bienes duraderos*), que según el estudio realizado sobre la base del análisis factorial exploratorio no formaban parte de ninguna dimensión.

La ausencia de carga de estos ítems sugería su eliminación. En primer lugar, en el caso del

sexto ítem, considerando el contexto económico venezolano (España & Ponce, 2018), aquellas personas que deciden ahorrar en bolívares y no en divisas extranjeras son sujetos con poca capacidad de entendimiento de las dinámicas económicas en contextos de adversidad, y, por tanto, esto no se asocia con ningún comportamiento financiero asertivo.

Este bajo entendimiento de las dinámicas inflacionarias también podría explicar la ausencia de carga en el octavo ítem. En estos contextos, percibir una deuda en la moneda a devaluarse y pagarla de forma retrasada en el tiempo, sería una estrategia para que el monto a cancelar sea mucho menor al adquirido en la deuda inicial.

Sin embargo, alrededor de un 80% de la muestra indicó que *Nunca* o *Casi nunca* adopta dicha estrategia, elemento que sugiere que la negativa a dicha posibilidad es propia de sujetos con un peor uso del dinero en un país hiperinflacionario. O bien, también podría simplemente sugerir que los sujetos de la muestra encuestada prefieren no endeudarse ante la realidad actual, no presentando estas conductas; aspecto que es muy similar al del ítem 15.

Con respecto al último objetivo, las puntuaciones de la escala presentaron evidencia de validez convergente en la medida en que todas las dimensiones de la conducta financiera se asociaron positiva y significativamente con otras medidas preexistentes de planificación de la compra, uso responsable del crédito y de la calidad de servicio (Denegri et al., citado en Ortega & Rodríguez-Vargas, 2004), tres de los indicadores subyacentes en la medición adecuada de la dimensión de ahorro, manejo de la deuda y del dinero (Pérez & Olaz, 2008).

En el caso de la validez discriminante, las dimensiones de manejo de la deuda y del dinero obtuvieron un patrón relacional negativo o nulo respecto a la medida *impulsividad de la compra*,

resultado apropiado desde un punto de vista psicométrico (Pérez & Olaz, 2008) en la medida en que se demuestra que estos factores de la conducta financiera se diferencian de aquellos constructos que se suponen teóricamente opuestos.

No obstante, para el caso de la dimensión del *ahorro*, se obtuvo una correlación inesperada entre esta y la *impulsividad de la compra*, ya que la asociación resultó baja y positiva. Es posible que esta relación se sustente en cómo las personas buscan sobrellevar el ritmo inflacionario de la moneda local: debido a que el bolívar no se constituye como una moneda estable, una alternativa de ahorro en Venezuela se presenta en la compra *impulsiva* de divisas o bienes duraderos.

En cuanto a la asociación entre las dimensiones de la conducta financiera, se halló un predominio de relaciones bajas. Específicamente, la asociación entre el *ahorro* y el *manejo de la deuda*, y entre este último y el *manejo del dinero*, aunque significativas, resultaron bajas. Esto se debe a que las personas con recursos económicos abundantes no tienden a preocuparse por la planificación que hacen del uso del dinero en términos de créditos y de manejo del efectivo. No es este el caso de las personas con dificultades económicas, quienes tienden a no endeudarse y a planificar con rigurosidad el uso de sus recursos financieros.

Por otra parte, se encontró que la escala presentó un funcionamiento diferencial de sus dimensiones en función del sexo de los sujetos, la ocupación actual y la edad. Estos hallazgos pueden considerarse como un elemento que permita en posteriores investigaciones obtener indicadores de validez de criterio que no fueron desarrollados en este estudio.

Por último, en cuanto a las limitaciones del estudio, se encontró el tipo de muestreo no probabilístico, la exclusión de participantes con otras ocupaciones laborales (p. ej., cargos obreros), y el hecho de que para utilizar esta escala en una

muestra ajena a la población venezolana se debe realizar una adaptación leve de los ítems, de forma que excluyan elementos propios de la dinámica económica del país (p. ej., la sustitución de la moneda local). Por otra parte, para poder utilizar la escala con los tres ítems que hacen referencia al uso de tarjetas de crédito, se debe garantizar que la mayoría de los sujetos encuestados tengan acceso a este tipo de beneficio financiero.

Referencias

- Barbar, R., Benasayag, S., & Salomón, L. (2019, enero). Nuevo encaje bancario: ¿Cuáles son las implicaciones? PRODAVINCI. Recuperado de <https://prodivinci.com>
- De Minzi, M. (2008). Nuevas tendencias en psicometría. *Revista Evaluar*, 8(1), 1-19. doi: [10.35670/1667-4545.v8.n1.501](https://doi.org/10.35670/1667-4545.v8.n1.501)
- Dew, J., & Xiao, J. J. (2011). The financial management behavior scale: Development and validation. *Journal of Financial Counseling and Planning*, 22(1), 43-59. Recuperado de <https://digitalcommons.uri.edu>
- España, L., & Ponce, M. (2018). *Evolución de la pobreza. IIES/UCAB*. Recuperado de <https://www.ucab.edu.ve>
- Fernández, A., Pérez, E., Alderete, A. M., Richaud, M. C., & Liporace, M. F. (2010). ¿Construir o adaptar tests psicológicos? Diferentes respuestas a una cuestión controvertida. *Revista Evaluar*, 10(1), 60-74. doi: [10.35670/1667-4545.v10.n1.459](https://doi.org/10.35670/1667-4545.v10.n1.459)
- Garðarsdóttir, R. B., & Dittmar, H. (2012). The relationship of materialism to debt and financial well-being: The case of Iceland's perceived prosperity. *Journal of Economic Psychology*, 33(3), 471-481. doi: [10.1016/j.joep.2011.12.00](https://doi.org/10.1016/j.joep.2011.12.00)
- Gärbling, T., Kirchler, E., Lewis, A., & Van Raaij, F. (2009). Psychology, financial decision making, and financial crises. *Psychological Science in the Public Interest*, 10(1), 1-47. doi: [10.1177/1529100610378438](https://doi.org/10.1177/1529100610378438)
- Gudmunson, C. G., & Danes, S. M. (2011). Family financial socialization: Theory and critical review. *Journal of Family and Economic Issues*, 32(4), 644-667. doi: [10.1007/s10834-011-9275-y](https://doi.org/10.1007/s10834-011-9275-y)
- Gudmunson, C., Ray, S. K., & Xiao, J. J. (2016). Financial socialization. En J. J. Xiao (Ed.), *Handbook of Consumer Finance Research* (pp. 61-72). Cham: Springer. doi: [10.1007/978-3-319-28887-1_5](https://doi.org/10.1007/978-3-319-28887-1_5)
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Estados Unidos: Cengage.
- IBM Corp. (2013). IBM. SPSS Statistics for Windows (Version 22.0). [Software de cómputo]. Armonk, NY: IBM Corp.
- Kahneman, D. (2013). *Pensar rápido, pensar despacio*. Estados Unidos: Debolsillo.
- Kerlinger, F., & Lee, H. (2002). *Investigación del comportamiento: Métodos de Investigación en Ciencias Sociales*. (4ª ed.). Ciudad de México, México: McGraw Hill.
- Kirchler, E., & Hoelzle, E. (2018). *Economic Psychology: An Introduction*. Cambridge University Press: United Kingdom.
- Ksendzova, M., Donnelly, G. E., & Howell, R. T. (2017). A brief money management scale and its associations with personality, financial health, and hypothetical debt repayment. *Journal of Financial Counseling and Planning*, 28(1), 62-75. doi: [10.1891/1052-3073.28.1.62](https://doi.org/10.1891/1052-3073.28.1.62)
- Laibson, D., & List, J. A. (2015). Principles of (behavioral) economics. *American Economic Review*, 105(5), 385-390. doi: [10.1257/aer.p20151047](https://doi.org/10.1257/aer.p20151047)
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. doi: [10.6018/analesps.30.3.199361](https://doi.org/10.6018/analesps.30.3.199361)
- Lorenzo-Seva, U., & Ferrando, P. (2011). FACTOR (Versión 8.02). [Software de cómputo]. Recuperado de <http://psico.fcep.urv.es/utilitats/factor/index.html>
- Magnusson, D. (1972). *Teoría de los test*. Ciudad de

México, México: Trillas.

- Manzini, J. L. (2000). Declaración de Helsinki: Principios éticos para la investigación médica sobre sujetos humanos. *Acta Bioethica*, 6(2). doi:[10.4067/S1726-569X2000000200010](https://doi.org/10.4067/S1726-569X2000000200010)
- Mcnair, S., & Crozier, W. R. (2017). Assessing psychological dispositions and states that can influence economic behaviour. En R. Ranyard (Ed.), *Economic Psychology* (pp. 69-87). Hoboken: Wiley-Blackwell. doi: [10.1002/9781118926352.ch5](https://doi.org/10.1002/9781118926352.ch5)
- Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862. Recuperado de <https://www.redalyc.org>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. doi: [10.7334/psicothema2018.291](https://doi.org/10.7334/psicothema2018.291)
- Olmo, D. G. (18 de noviembre de 2019). Dolarización en Venezuela: Cómo Nicolás Maduro cambió de opinión sobre el dólar y su papel en la economía. *BBC News*. Recuperado de <https://www.bbc.com/mundo/noticias-america-latina-50466818>
- Ortega, V., & Vargas, J. C. R. (2004). Escala de hábitos y conductas de consumo: Evidencias sobre dimensionalidad. *International Journal of Clinical and Health Psychology*, 4(1), 121-136. Recuperado de <https://www.redalyc.org>
- Pérez, E., & Olaz, F. (2008). Validez. En S. Tornimbeni, E. Pérez & F. Olaz (Eds.), *Introducción a la Psicometría* (pp.101-136). Buenos Aires, Argentina: Paidós.
- Perry, V. G., & Morris, M. D. (2005). Who is in control? The role of self-perception, knowledge, and income in explaining consumer financial behavior. *Journal of Consumer Affairs*, 39(2), 299-313. doi: [10.1111/j.1745-6606.2005.00016.x](https://doi.org/10.1111/j.1745-6606.2005.00016.x)
- Porter, N. M., & Garman, E. T. (1993). Testing a conceptual model of financial well-being. *Financial Counseling and Planning*, 4, 135-165. doi: [10.1.1.379.6832](https://doi.org/10.1.1.379.6832)
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en España. *Papeles del Psicólogo*, 77, 65-75. Recuperado de <https://www.redalyc.org>
- Ranyard, R., & De Mello-Ferreira, V. R. (2017). Introduction to economic psychology: The science of economic mental life and behaviour. En R. Ranyard (Ed.), *Economic Psychology* (pp. 1-18). doi: [10.1002/9781118926352.ch1](https://doi.org/10.1002/9781118926352.ch1)
- Sardi, G., Angelucci, L., Martins, A., & Peña, G. (2020) Representaciones de la economía como fenómeno psicosocial. En C. Peña (Ed.). *Miradas a la Venezuela del siglo XXI* (pp. 377-406). Caracas, Venezuela: Instituto de Investigaciones Económicas y Sociales: "Dr. Rodolfo Quintero".
- Veiga, R. T., Avelar, C., Moura, L. R. C., & Higuchi, A. K. (2019). Validation of scales to research the personal financial management. *Revista Brasileira de Gestão de Negócios, São Paulo*, 21(2), 332-348. doi: [10.7819/rbgn.v21i2.3976](https://doi.org/10.7819/rbgn.v21i2.3976)
- UCAB. (2020, 13 de julio). UCAB ENCOVI | 2019-2020. Recuperado de https://www.youtube.com/watch?reload=9&v=ph0QyIaWJ_k&feature=youtu.be
- Van Raaij, W. F. (1981). Economic psychology. *Journal of Economic Psychology*, 1(1), 1-24. doi: [10.1016/0167-4870\(81\)90002-7](https://doi.org/10.1016/0167-4870(81)90002-7)
- Van Raaij, W. F. (2016). *Understanding consumer financial behavior money management in an age of financial illiteracy*. doi: [10.1057/9781137544254](https://doi.org/10.1057/9781137544254)
- VandenBos, G. R. (2015). *APA Dictionary of Psychology* (2ª ed.). USA: American Psychological Association. doi: [10.1037/14646-000](https://doi.org/10.1037/14646-000)
- Xiao, J. J. (2008). Applying behavior theories to financial behavior. En J. J. Xiao (Ed.), *Handbook of Consumer Finance Research* (pp. 69-81). New York, USA: Springer. doi: [10.1007/978-0-387-75734-6_5](https://doi.org/10.1007/978-0-387-75734-6_5)



Evaluación de la comprensión del texto científico-académico en el posgrado mediante el procedimiento cloze

Assessment of Reading Comprehension of Scientific-Academic Text at the Postgraduate Level Using Cloze Procedure

Introducción
Método
Resultados
Discusión
Referencias

Hilda Difabio de Anglat *¹, Guadalupe Álvarez²

1 - Centro de Investigaciones Cuyo Dr. Abelardo Pithod, CONICET, Mendoza, Argentina.

2 - Instituto del Desarrollo Humano de la Universidad Nacional de General Sarmiento, CONICET, Buenos Aires, Argentina.

Recibido: 14/10/2020 Revisado: 25/11/2020 Aceptado: 15/12/2020

Resumen

El cloze es un procedimiento que reviste potencial para evaluar de forma parsimoniosa la comprensión del texto científico-académico, para el cual no se registran antecedentes de investigación respecto de la comprensión de textos en español en el nivel de posgrado. Por ello, los objetivos de este trabajo fueron: 1) identificar el nivel de dificultad e índice de discriminación de los ítems de un cloze que se empleó en un curso de posgrado (Facultad de Filosofía y Letras, Universidad Nacional de Cuyo, Argentina); 2) ponderar si la distribución desde ítems muy fáciles a muy difíciles se ajustaba a los porcentajes aconsejados; 3) describir los niveles funcionales de lectura de los participantes, y 4) estimar la incidencia, en la dificultad de los ítems, de los denominados factores intrínsecos o aspectos de la tarea. En un corpus conformado por 82 clozes, mediante análisis estadísticos, se hallaron buenas cualidades psicométricas para el instrumento y niveles diferenciales de dificultad de las palabras lexicales en función de los factores en estudio.

Palabras clave: *comprensión lectora, procedimiento cloze, propiedades psicométricas, nivel de posgrado, español*

Abstract

Cloze is a procedure with the potential of parsimoniously assessing the understanding of the scientific-academic text. There are no previous studies regarding the comprehension of Spanish texts at the postgraduate level. The aims of this study were: 1) to identify difficulty levels and discrimination index of the cloze items used in a postgraduate course (Facultad de Filosofía y Letras, Universidad Nacional de Cuyo, Argentina); 2) to analyse whether the distribution of the items ranging from very easy to very hard corresponds to percentages advised by statistical theory; 3) to describe the participants' functional reading levels and 4) to estimate the incidence of intrinsic factors or aspects of the task in items' difficulty. Eighty-two cloze tests were collected and statistically analysed. The analysis revealed acceptable psychometric qualities for the instrument and acceptable differential levels of difficulty in content words according to the factors under study.

Keywords: *reading comprehension, cloze procedure, psychometric properties, postgraduate level, Spanish*

* **Correspondencia a:** Hilda Difabio de Anglat. Primitivo de la Reta 522, 2° K, 5.500 Mendoza, Argentina. E-mail: ganglat@gmail.com

Introducción

La importancia de la comprensión de textos científico-académicos en los estudios de posgrado es indiscutible en sí misma y en su contribución al desarrollo de los procesos de producción textual (Parodi, 2001) por las “conexiones de tipo psicolingüístico y discursivo” (p. 59) que se verifican entre estos dos procesos y porque, en este nivel educativo, se lee para escribir. En consecuencia, en una instancia educativa orientada a promover la producción escrita (un taller virtual focalizado en familiarizarse con el género para componer un primer capítulo o apartado de la tesis), se pondera la competencia lectora de los cursantes mediante un cloze, procedimiento que suprime palabras de un texto en prosa, reemplazándolas por una línea de extensión constante.

Wilson Taylor (1953) desarrolló el test de cloze como medida de la legibilidad o comprensibilidad de los textos, aunque el procedimiento ha sido empleado en forma ininterrumpida como herramienta de investigación y de evaluación de la competencia lectora, e inclusive como dispositivo de enseñanza (p. ej., Sadeghi, 2008).

El término *cloze* parece ser un neologismo, pues no figura en los diccionarios (Cairney, 2018). Taylor (1953) señala que lo deriva de la palabra inglesa *closure* en tanto constituye una aplicación al lenguaje del principio de cierre de la psicología de la Gestalt o tendencia a percibir como completa una forma fragmentada, a completar un patrón una vez que se ha captado su significado esencial.

Para completar un cloze, el lector debe *cerrar* las percepciones truncadas mediante inferencias (p. ej., Chino-Vilca & Zegarra-Valdivia, 2019) a partir de las claves semánticas (entre ellas, la redundancia textual), sintácticas, discursivas, y las restricciones del lenguaje (Peña-Calvo, 1997), así como de su conocimiento previo y de su expectativa intuitiva (Difabio de Anglat, 2008).

Esto es, el procedimiento requiere que el lector active todos sus recursos cognitivos y lingüísticos a fin de sostener el proceso interactivo entre la información que aporta el texto (proceso de abajo hacia arriba) y la que él anticipa (proceso de arriba hacia abajo); gracias a estos procesos, el lector cumple el objetivo propio de toda tarea de lectura: “dotar de coherencia a lo leído” (León-Cascón, 2004, p. 109). Para ello, el cloze exige *reconstruir el texto* a un nivel de mayor control o monitoreo que la lectura normal (esto es, de un texto completo) ya que demanda atender a una gama amplia de señales adicionales. La comprensión de un cloze también reclama considerable motivación e interés para mantener la atención en los procesos referidos. Por ende, no solo parece contribuir a generar mayor aprehensión del significado, sino a estimular la reflexión metacognitiva sobre el desempeño o *comprensión metacognitiva* (León-Cascón, 2004). Por otra parte, dado que es relativamente resistente al tiempo de exposición, es apropiado para la autoadministración (Datta & Macdonald-Ross, 2002).

Sin embargo, aún hoy se investigan los problemas planteados en los estudios tempranos (Abraham & Chapelle, 1992; Brown, 1980, 1988; Jonz, 1990), que continúan siendo relevantes; entre ellos, su validez, qué evalúa exactamente (p. ej., Gellert & Elbro, 2013), los patrones de eliminación de términos (Kleijn, Pander-Maat, & Sanders, 2019), los métodos de calificación (Kleijn et al., 2019; O’Toole & King, 2010) y, en menor medida, la dificultad de los ítems (Trace, 2020; Trace, Brown, Janssen, & Kozhevnikova, 2017). La primera pregunta conforma un área de ininterrumpido debate, que enfrenta a quienes consideran que solo evalúa comprensión local (por ejemplo, Carlisle & Rice, 2004, en Trace et al., 2017) con aquellos que sostienen que puede ponderar la coherencia global, constituyéndose así en una medida válida y rápida de la competencia lectora

(Gellert & Elbro, 2013; Kleijn et al., 2019; Shah-nazari-Dorcheh, Roshan, & Hesabi, 2012; Trace, 2020).

Esto es, aunque el cloze no produzca pruebas perfectas de la competencia lectora general, reviste potencial para evaluar de forma parsimoniosa las habilidades superiores de lectura (el monitoreo de la comprensión, la anticipación o predicción, la inferencia, la elaboración o relación del texto con el conocimiento previo, la vinculación entre diferentes ideas expuestas en el texto), si las eliminaciones deliberadamente apelan a demandas inferenciales vinculadas con la estructura lógica subyacente al texto (Green, 2001). En otras palabras, se pueden generar lagunas de mayor interés (Gellert & Elbro, 2013) por la supresión de palabras de contenido (palabras léxicas) y conectores que se basan en las relaciones intertextuales.

Para la calificación de un texto cloze, se dispone de dos métodos: aceptar solo la palabra efectivamente omitida (*verbatim*), *puntuación de la respuesta exacta* (Brown, 2013), o computar también los sinónimos, *puntuación de la respuesta aceptable*, calificación semántica (Kleijn, 2018) o conceptual (O'Toole & King, 2011).

Hoy los investigadores (Brown, 2013; Kleijn, 2018; Kleijn et al., 2019; O'Toole & King, 2011) acuerdan en que la segunda es una puntuación más adecuada en términos de discriminación de los ítems (una cualidad importante para nuestra investigación) y, sobre todo, porque proporciona una ponderación más ajustada del desempeño del lector en cuanto la respuesta exacta enmascara la competencia que pueden reflejar reemplazos potencialmente lícitos (Kobayashi, 2002). Al respecto, el estudio meta-analítico de Watanabe y Koyama (2008) confirma la investigación seminal de Brown (1980), poniendo en evidencia que el método de puntuación *verbatim* es menos confiable y válido que el de la respuesta aceptable,

en particular en relación con las palabras léxicas.

La estrategia usual para la interpretación de los resultados consiste en transformar los porcentajes finales en niveles funcionales de lectura: independiente, instruccional y de frustración (p. ej., Condemarin & Milicic, 1988). El primero remite a un lector que comprende acabadamente el texto de modo autónomo. El nivel instruccional significa una aproximación global, con algunas dificultades en el reconocimiento de palabras y en el proceso inferencial. En el de frustración, la comprensión es deficiente porque el lector se apoya solo en el contexto inmediato para completar los blancos. Con relación al criterio para la adscripción de estos niveles, los trabajos en español (entre otros, Calderón-Ibañez & Quijano-Peñuela, 2010; Chino-Vilca & Zegarra-Valdivia, 2019) aplican la categorización porcentual de González-Moreyra (1998).

Con referencia a la investigación de la última década en dicho idioma, varios estudios emplean la técnica para evaluar la competencia lectora de estudiantes de nivel universitario, ya sea como instrumento único (p. ej., Calderón-Ibañez & Quijano-Peñuela, 2010; Márquez, Ibañez, & Pérez, 2011) o en combinación con otros (Chino-Vilca & Zegarra-Valdivia, 2019; Esquivel-Gámez, Martínez-Olvera, Córdoba del Valle, & Reyes-Gutiérrez, 2016; Tabullo et al., 2020). En general, estos estudios han mostrado que el nivel de comprensión de los alumnos se halla por debajo del deseado y que el desempeño en el cloze se relaciona con el rendimiento académico.

En cambio, hasta donde se ha podido indagar, el cloze no se utiliza con frecuencia en el posgrado; los dos trabajos que se identificaron (Costa, 2006; Sadeghi, 2008) lo aplican en portugués e inglés, respectivamente. De allí el interés por ponderar su potencialidad evaluativa en lectores expertos de español como primera lengua; en particular, se intenta analizar los factores que inciden

en la dificultad de los ítems. Este propósito registra trabajos tempranos (Abraham & Chapelle, 1992; Bachman, 1985; Brown, 1989; Jonz, 1990; Kobayashi, 2002) y más recientes desde la teoría de respuesta al ítem (p. ej., Trace, 2020; Trace et al., 2017), análisis inaplicable en el corpus que se analiza por el tamaño de la muestra, dado que “para emplear modelos TRI se requieren muestras grandes de sujetos ($n > 300$)” (Cortada de Kohan, 2004, p. 105).

Abraham y Chapelle (1992) estudian los *factores intrínsecos*, que definen como aspectos de contenido de la tarea factibles de medición con independencia de las características del lector. Las autoras consideran: palabras de contenido o funcionales y, para las primeras, cantidad de contexto requerido para completar los blancos, viabilidad de respuestas alternativas y número de ocurrencias del término en el texto.

A partir de las consideraciones previas, el presente trabajo se plantea cuatro objetivos: 1) identificar el nivel de dificultad e índice de discriminación de los ítems de un cloze que se emplea en un taller virtual de escritura en el posgrado; 2) ponderar si la distribución desde ítems muy fáciles a muy difíciles se ajusta a los porcentajes aconsejados; 3) describir los niveles funcionales de lectura de los participantes, y 4) estimar la incidencia, en la dificultad de los ítems, de los referidos factores intrínsecos, nuestro objetivo principal.

Metodología

Diseño y participantes

Se desarrolló un estudio empírico con metodología cuantitativa, transversal y de profundidad analítica en un corpus de textos cloze proveniente de los cursantes de Estrategias para la producción conceptual y escrita en el posgrado en Ciencias

Sociales y Humanas, taller en línea acreditado por la Secretaría de Posgrado, Facultad de Filosofía y Letras, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina. En este curso, que se implementa en la plataforma Moodle institucional, el cloze forma parte de un diagnóstico integral. A partir de la consigna proporcionada (ver Anexo 1), cada asistente lo completa de modo asincrónico durante la primera semana; luego, sube la resolución al espacio creado con este fin en la plataforma.

La muestra ($N = 82$) se conforma por 76 doctorandos, cinco maestrandos y una doctora en Psicología; 60 mujeres y 22 varones; todos hablantes nativos del español. Los estudiantes se distribuyen entre carreras de posgrado en Educación (el 39%), Letras (21%) y otras disciplinas (Filosofía, Historia, Geografía, Psicología, Sociología, Derecho y Arquitectura) de la UNCuyo pero también de otras universidades argentinas. Diez de los asistentes son extranjeros, todos procedentes de países hispanoamericanos.

La muestra se constituyó de modo progresivo entre 2015 y 2019. Su tamaño se debe a que este taller, en pro de alcanzar su objetivo último de favorecer la producción de un capítulo o apartado de la tesis, es de cupo limitado (hasta 12 estudiantes). Dado que los datos provienen de las implementaciones del cloze en sucesivas versiones del curso referido (en total, siete), por un lado, se recurre al test no paramétrico de Kruskal-Wallis para establecer si se verifican diferencias de rendimiento significativas entre los grupos; por el otro, a la prueba de chi cuadrado a fin de confrontar los niveles funcionales de lectura entre cursantes con formación en Letras ($n = 17$) y sin dicha formación ($n = 65$). Los resultados ($\chi^2 = 3.412$, $gl = 5$, $p = .756$; $\chi^2 = 3.144$, $gl = 2$, $p = .208$, respectivamente) ponen de manifiesto que es posible reunirlos en una sola muestra.

Instrumento

El cloze aplicado (ver Anexo 1) sintetiza el texto “Principios para un final: reflexiones en torno a la escritura de la tesis” (Álvarez, 2011), al que se le agregó como cierre parte de un párrafo de Cubo de Severino y Bosio (2011), capítulo del libro de cabecera del taller. En consecuencia, recurre a un texto ya publicado, en lugar de a uno elaborado *ad hoc*, porque la evidencia empírica (p. ej., O’Toole & King, 2011) ha demostrado que emplear un texto auténtico incrementa la validez de la prueba. Además, conforma un ejemplar representativo del tipo de lectura que los estudiantes efectuarán durante el curso y no depende de conocimiento especializado (palabras *impredecibles*) en cuanto se garantizó al menos una ocurrencia en el texto de los términos técnicos (*situación comunicativa, puesta en discurso, revisión, edición*).

El cloze consta de 484 palabras y 79 blancos, los que se establecieron mediante la combinación de dos criterios o estrategia híbrida (método semejante al de Kleijn et al., 2019): *razón fija*, que consiste en la eliminación del quinto vocablo y sus múltiplos, y *racional*, el cual permite controlar las palabras omitidas y, por esta vía, maximizar el empleo de información fuera del párrafo que contiene el blanco (Abraham & Chapelle, 1992; Bachman, 1985; Gellert & Elbro, 2013). En este sentido, se seleccionó un porcentaje ínfimo (16%) de términos que se podían completar con el solo conocimiento gramatical (palabras funcionales), y el 60% de las palabras léxicas exigía atender a claves entre párrafos o anticipar términos que no aparecen en el texto. El porcentaje total de omisión (también del 16%) se ubicó entre los estándares de la razón fija (20%) y del criterio racional (10%; Kleijn, 2018). La primera oración y la última se reprodujeron completas.

En la calificación se atribuyó un punto a

cada laguna correctamente resuelta (*verbatim* o un sinónimo). El criterio para la determinación de las alternativas correctas fue su adecuación global, esto es, la respuesta debía cumplir “los requisitos de todo el contexto del discurso en el que aparece” (Oller & Jonz, 1994, en Kleijn et al., 2019, p. 8). El total posible fue de 79 puntos, puntuación que se transformó en porcentaje, dado que los niveles funcionales de lectura se expresan porcentualmente (Condemarin & Milicic, 1988; Difabio de Anglat, 2008).

Análisis de datos

Se llevaron a cabo los siguientes análisis. En primer lugar, se elaboró el espacio de alternativas correctas (ver Anexo 2), una tarea importante para una investigación basada en la puntuación de la respuesta aceptable. Por ello, se efectuó mediante acuerdo inter-jueces a través del siguiente procedimiento: la primera autora confeccionó una matriz inicial de todas las respuestas (6.478), distinguiendo entre *verbatim*, alternativa/s correcta/s y error/es; la segunda autora evaluó la codificación y señaló algunos desacuerdos (de alrededor del 15%); finalmente, una especialista en lingüística del español revisó el análisis, lo que permitió resolver las disidencias.

En segundo lugar, se determinó el nivel de dificultad (ND) de cada ítem; como es práctica frecuente (Abraham & Chapelle, 1992; Brown, 1989, 2013; Kobayashi, 2002; Angeli Dos Santos, Primi, Taxa, & Vendramini, 2002), se empleó un método sencillo, aplicado en un trabajo previo (Difabio de Anglat, 2008), que consiste en dividir, en cada ítem, la frecuencia de respuestas correctas por el tamaño de la muestra. En las investigaciones sobre el cloze (p. ej., Abraham & Chapelle, 1992; Brown, 1989) este porcentaje se denomina ND y no nivel de facilidad de los ítems.

Para la interpretación de los resultados, se utilizó la categorización de Backhoff-Escudero, Larrazo-Reyna y Rosas-Morales (2000) en ítems muy fáciles (.87 a 1), fáciles (.74 a .86), moderados (.53 a .73), difíciles (.33 a .52) y muy difíciles (.00 a .32).

En tercer lugar, se estableció el índice de discriminación (ID) de cada ítem. Se aplicó también el procedimiento clásico (Brown, 1989; Angeli Dos Santos et al., 2002) que consiste en restar, en cada ítem, los niveles de dificultad hallados en los grupos extremos (superior, GS, e inferior, GI); estos grupos se conformaron según la puntuación total por el 27% superior y el 27% inferior, 22 cursantes en cada uno.

En cuarto lugar, se calculó el índice de dificultad del *cloze bien adaptado* —*well-tailored cloze* (Brown, 1988, 2013)—. Un *cloze* bien adaptado es aquel que retiene los ítems que manifiestan índices de discriminación $\geq .30$ (Lozano-Fernández & de la Fuente-Solana, 2015). En estos índices se distinguió entre buena potencialidad discriminativa (entre .30 y .39) y muy buena (.40 o más). No obstante, también se conservaron algunos ítems con discriminación marginal (entre .20 y .29).

En quinto lugar, las puntuaciones totales del *cloze* adaptado se transformaron en niveles funcionales de lectura o categorías de competencia lectora. Se adecuó la categorización de Gonzá-

lez-Moreyra (1998) para dicha transformación: *independiente*, puntajes de 75% o más, subcategorizado en tres niveles, excelente (100% - 90%), muy bueno (89% - 80%) y bueno (75% - 79%); *instruccional*, entre 59% y 74%, y *de frustración*, 58% o menos, subcategorizado en de dificultad (58% - 45%) y deficiente (44% o menos).

En sexto lugar, se estimó la incidencia en ND de los factores intrínsecos; se efectuaron cuatro análisis: 1) palabras léxicas vs funcionales, 2) cantidad de contexto requerido para completar el blanco, 3) número de respuestas alternativas y 4) otras ocurrencias del ítem en el texto. Los tres últimos análisis se restringieron a los términos léxicos porque la investigación ha mostrado (Abraham & Chapelle, 1992) que solo se justifican para este tipo de palabras.

De acuerdo con la investigación previa (Abraham & Chapelle, 1992; Bachman, 1985; Jonz, 1990; Kobayashi, 2002; Trace, 2020; Trace et al., 2017), en el presente trabajo se postula que tienden a ser más fáciles los ítems funcionales que los lexicales, los que requieren menor cantidad de contexto para restaurar las palabras suprimidas, los que registran alternativas conceptualmente aceptables y los que manifiestan otras ocurrencias en el texto.

A fin de determinar la cantidad de contexto requerido para completar cada blanco, se adaptó la categorización de la investigación clásica en

Tabla 1

Ejemplificación de las categorías sobre cantidad de contexto requerido para completar los blancos.

Categorías	Ejemplo/s
1. En la oración	Las operaciones de edición podrían describirse como <i>operaciones</i> ...
2. En oraciones consecutivas	La revisión comprende dos subprocesos: <i>evaluación</i> y edición. En la evaluación se...
3. En el párrafo	La escritura, la presentación y la defensa de la tesis son las etapas conclusivas de los estudios de posgrado (...). Así, en el nivel de <i>posgrado</i> , el estudiante...
4. En el texto	La planificación es una evaluación inicial de la situación <i>comunicativa</i> (...).
5. Fuera del texto	El término “situación comunicativa” aparece completo al final del párrafo siguiente. <i>estrategias, elementos</i>

el tema (Abraham & Chapelle, 1992; Bachman, 1985; Jonz, 1990): 1) en la oración; 2) en oraciones consecutivas; 3) en el párrafo; 4) en el texto, y 5) fuera del texto ($n = 28$, el 41% de los blancos). Estas categorías se ejemplifican en la Tabla 1.

Para el análisis de esta variable, como en Abraham y Chapelle (1992), se aplicó un estadígrafo de correlación (ρ de Spearman por el tamaño de la muestra) entre las cinco categorías y el nivel de dificultad de los ítems expresado en porcentaje de respuestas correctas.

Dado que no se dispone de estudios previos sobre la incidencia del número de respuestas alternativas porque tanto Abraham y Chapelle (1992) como Kobayashi (2002) únicamente distinguen entre *una respuesta / más de una respuesta*, se computó dicho número a partir del corpus propio.

Finalmente, en relación con otras ocurrencias de las palabras de contenido en el texto, de los 58 blancos léxicos en el cloze aplicado aparecen 31 (el 53%) como término exacto o desde la familia de palabras en tres casos ($n^\circ 38$, *producir*, $n^\circ 56$, *elaborado* y $n^\circ 64$, *solucionar*), de los cuales únicamente cinco lo hacen más de dos veces. Luego, no resultó apropiada la categorización de Kobayashi (2002) en tres niveles: 1) la palabra no aparece en otro lugar del texto; 2) aparece entre dos y cuatro veces y 3) aparece cinco veces o más. De allí que, como en Abraham y Chapelle (1992), se distinguieron dos grupos según se verificaban o no ocurrencias en el texto (31 vs 27 palabras).

Los procedimientos analíticos empleados fueron: prueba de dependencia / independencia entre distribuciones empíricas (chi cuadrado o V de Cramer si el porcentaje de casillas con un recuento esperado menor que cinco superara el 20%) a fin de analizar la distribución de los niveles funcionales de lectura; diferencia entre muestras independientes y emparejadas (t de Student) en el caso de las variables clase de palabra (lexicales o

funcionales), número de respuestas alternativas y otras ocurrencias del término en el texto, y coeficiente de correlación (ρ de Spearman) respecto de la cantidad de contexto requerido para completar cada blanco.

Los análisis estadísticos se llevaron a cabo mediante el programa SPSS 25.0 (IBM Corporation, 2017).

Resultados

Antes de abordar los resultados referidos a la incidencia de los factores intrínsecos en la dificultad de los ítems, objetivo principal de este trabajo, cabe señalar que: a) los blancos sin responder ($n = 183$) representaron el 2.8%, un porcentaje similar al que permiten Kleijn et al. (2019) para retener una resolución del cloze en su muestra, en tanto lagunas en las que es lícito asumir que el lector no encuentra una respuesta porque están dispersas a lo largo de la prueba; b) la confiabilidad del cloze aplicado (alfa de Cronbach) es de .85, un buen índice que habilita ulteriores análisis ya que, obviamente, los resultados de un estudio son confiables en la medida en que lo es el instrumento en el que se basan; c) el rango de las correlaciones ítem-total (coeficiente de homogeneidad) se distribuyó entre .05 (ítem $n^\circ 39$, *como*) y .59 ($n^\circ 43$, *Puesta*), con una media de .31.; d) ocho ítems no alcanzaron .20, el mínimo requerido de homogeneidad (Kline, 1998); sin embargo, se retuvieron dado que su supresión incrementaba el alfa en solo un centésimo (.86). Al respecto, Kobayashi (2002) trabaja con una distribución cercana: entre .08 y .53, con una media de .34, en estudiantes universitarios japoneses que completan textos cloze en inglés.

En la Tabla 2, que se divide en dos para facilitar la lectura, se presentan los resultados del análisis del nivel de dificultad e índice de descri-

Tabla 2

Nivel de dificultad e índice de discriminación de los ítems del cloze.

Ítem	ND	GS	GI	ID	Ítem	ND	GS	GI	ID
1	.95	.95	.95	0	21	.74	.95	.64	.31
2	.72	.82	.50	.32	22	.79	.95	.55	.40
3	.79	.95	.68	.27	23	.73	.95	.50	.45
4	.40	.59	.27	.32	24	.72	.86	.55	.31
5	.98	1	.95	.05	25	.91	1	.77	.23
6	.55	.73	.36	.37	26	.41	.64	.27	.37
7	.48	.64	.32	.32	27	.49	.68	.36	.32
8	.54	.77	.32	.45	28	.73	.91	.55	.36
9	.78	.91	.59	.32	29	.49	.82	.14	.68
10	.90	1	.77	.23	30	.80	.95	.68	.27
11	.73	.95	.50	.45	31	.93	1	.86	.14
12	.73	.91	.55	.36	32	.90	1	.77	.23
13	.94	.95	.95	0	33	.83	1	.68	.32
14	.32	.45	.14	.31	34	.26	.50	.09	.41
15	.48	.68	.36	.32	35	.54	.68	.36	.32
16	.80	1	.50	.50	36	.79	1	.68	.32
17	.90	1	.77	.23	37	.95	1	.95	.05
18	.96	1	.91	.09	38	.87	1	.68	.32
19	.82	.95	.64	.31	39	.37	.55	.32	.23
20	.84	1	.68	.32	40	.69	.82	.45	.37

Ítem	ND	GS	GI	ID	Ítem	ND	GS	GI	ID
41	.96	.95	.95	0	61	.32	.68	.04	.64
42	.80	1	.68	.32	62	.54	.82	.45	.37
43	.57	.95	.05	.91	63	.44	.64	.23	.41
44	.82	1	.64	.36	64	.71	.91	.59	.32
45	.18	.45	0	.45	65	.90	.95	.86	.09
46	.35	.68	.18	.50	66	.71	.91	.45	.46
47	.54	.73	.36	.37	67	.84	1	.55	.45
48	.98	.95	.95	0	68	.80	1	.68	.32
49	.73	.86	.50	.36	69	.61	.77	.41	.36
50	.89	1	.68	.32	70	.90	1	.77	.23
51	.40	.64	.18	.46	71	.57	.73	.41	.32
52	.60	.68	.36	.32	72	.77	.91	.55	.36
53	.46	.82	.14	.68	73	.73	.77	.45	.32
54	.78	1	.50	.50	74	.96	1	.91	.09
55	.72	.95	.45	.50	75	.46	.68	.27	.41
56	.67	.77	.55	.22	76	.61	.73	.27	.46
57	.37	.55	.23	.32	77	.73	.91	.50	.41
58	.90	1	.77	.23	78	.24	.50	.14	.36
59	.62	.73	.27	.46	79	.60	.73	.41	.32
60	.80	.95	.64	.31					

Nota. ND = Nivel de dificultad; GS = Grupo superior; GI = Grupo inferior; ID = Índice de discriminación.

minación de los ítems.

Diez ítems no alcanzaron el mínimo aceptable de discriminación porque los respondieron correctamente la mayoría de los estudiantes (en-

tre dos y seis errores por ítem); ellos son: n° 1 (*ordenamiento*), n° 5 (*tesis*), n° 13 (*investigación*), n° 18 (*como*), n° 31 (*mutuamente*), n° 37 (*para*), n° 41 (*Es*), n° 48 (*cabo*), n° 65 (*nivel*) y n° 74

(*problemas*). En consecuencia, el instrumento se redujo a 69 ítems, de los cuales 58 corresponden a palabras léxicas y 11 a palabras funcionales, cantidad que mantiene la relación proporcional inicialmente intentada (84% vs 16%).

También 10 son los ítems que evidenciaron discriminación marginal (destacados en letra itálica en la Tabla 2), el 14.5% del instrumento, de los cuales seis corresponden a ítems muy fáciles: n° 10 (*producción*), n° 17 (*aspectos*), n° 25 (*resolución*), n° 32 (*proceso*), n° 58 (*hasta*) y n° 70 (*sobre*). Los restantes se distribuyeron entre discriminación buena —35 ítems (el 50.7%)— y muy buena —24 ítems (el 34.8%)—; cuatro manifiestan excelente potencialidad discriminativa (superior a .60) y remiten a términos técnicos: n°

29 (*discurso*), n° 43 (*Puesta*), n° 53 (*revisión*) y n° 61 (*tarea* de escritura). El ítem n° 43 alcanzó discriminación prácticamente óptima (.91); recordemos que también manifiesta el índice más alto de correlación ítem-test.

En relación con el nivel de dificultad de los ítems, mediante la categorización de Backhoff-Escudero et al. (2000) desde ítems muy fáciles a muy difíciles, se hallaron los resultados que se presentan en la Tabla 3.

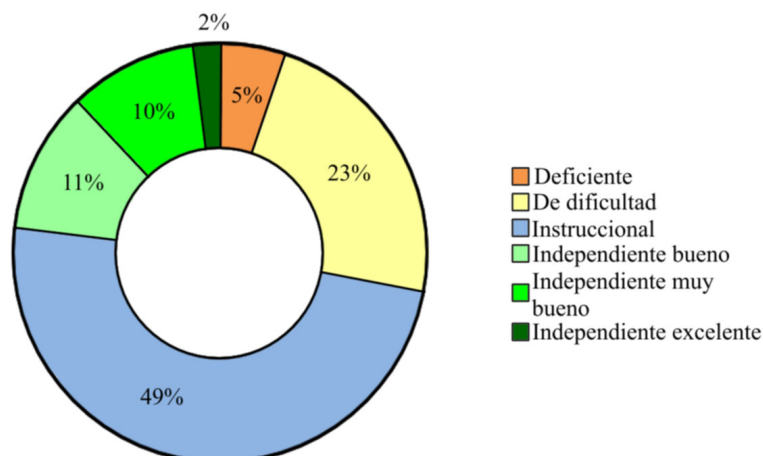
El índice medio de dificultad del cloze es .65, levemente superior al rango sugerido por la teoría estadística (entre .50 y .60; Brown, 1999).

El blanco más difícil es el n° 45 (*considerando*); solo 15 cursantes lo respondieron correctamente, con el gerundio esperado o una alternati-

Tabla 3

Categorización de los niveles de dificultad de los ítems del cloze.

Nivel de dificultad	Categorización de Backhoff-Escudero et al. (2000)	% aproximado ideal	% en este instrumento	Ítems N°
Muy fácil	.87 a 1	10%	10%	8 ítems: 10 (producción), 17 (aspectos), 25 (resolución), 32 (proceso), 38 (producir), 50 (por), 58 (hasta) y 70 (sobre)
Fácil	.74 a .86	20%	25%	17 ítems: 3 (escritura), 9 (estrategias), 16 (rol), 19 (dicho), 20 (concebirse), 21 (parte), 22 (cuya), 30 (se), 33 (implica), 36 (tiempo), 42 (etapa), 44 (expresar), 54 (evaluación), 60 (ajuste), 67 (operaciones), 68 (elementos) y 72 (procesos)
Moderado	.53 a .73	40%	38%	26 ítems: 2 (procesos), 6 (su), 8 (exige), 11 (aquellas, las), 12 (textos), 23 (elaboración), 24 (sentido), 28 (escrita), 35 (texto), 40 (lista), 43 (Puesta), 47 (detallado), 49 (conocer), 52 (requerido), 55 (verifica), 56 (elaborado), 59 (corrobora), 62 (adecuación), 64 (solucionar), 66 (edición), 69 (inserción), 71 (hacer), 73 (partida), 76 (además), 77 (reflexión) y 79 (razones)
Difícil	.33 a .52	20%	19%	13 ítems: 4 (resulta), 7 (dominio), 15 (posgrado), 26 (realizar), 27 (involucrados), 29 (discurso), 39 (como), 46 (escritura), 51 (lenguaje), 53 (revisión), 57 (su), 63 (ortográficas) y 75 (esta)
Muy difícil	.00 a .32	10%	7%	5 ítems: 14 (Así), 34 (comunicativa), 45 (considerando), 61 (tarea) y 78 (académicos)



va aceptable (por ejemplo, *utilizando*). El resto de los blancos de esta categoría remite a un adverbio (n° 14, *Así*), dos adjetivos (n° 34, *comunicativa*, y n° 78, *académicos*) y un sustantivo (n° 61, *tarea de escritura*); todos términos de contenido.

Los niveles funcionales de lectura hallados en la muestra pueden verse en el Gráfico 1.

En el nivel de frustración se ubicó el 28% de la muestra; por carrera, el porcentaje menor de cursantes en este nivel correspondió a los historiadores (4%), el siguiente (9%), a los maestrandos o doctorandos en Letras, Derecho, Filosofía y Psicología. El resto de los estudiantes se distribuyó de la siguiente manera: el 49% en el nivel instruccional y el 23% en el nivel independiente. Ninguna de estas diferencias por carrera alcanzó significatividad estadística (V de Cramer = .294, $p = .292$).

En relación con la incidencia de los factores intrínsecos en la dificultad de los ítems, en el primer análisis (palabras léxicas vs funcionales), la prueba t de Student (palabras funcionales: $X = 69.74$, $DE = 16.54$; palabras léxicas: $X = 63.95$, $DE = 13.61$) evidencia diferencias estadísticamente significativas ($t = 3.704$, $gl = 81$, $p < .001$), siendo superior el valor de la media en el caso de las palabras funcionales. Ya que la divergencia de medias en términos absolutos no es tan elocuente,

se realizó un análisis adicional. Se calcularon los niveles de lectura en la puntuación total en cada clase de palabras en las tres categorías principales: independiente, instruccional y de frustración. Luego, en una tabla cruzada (ver Tabla 4), se identificó el número de respuestas en las que: 1) se obtuvo el mismo nivel de rendimiento para ambas clases (43, el 53%); 2) fue superior el rendimiento en las palabras funcionales (28, el 34%) y 3) fue superior el desempeño en los términos léxicos (11, el 13%). Los resultados ($\chi^2 = 16.753$, $gl = 4$, $p < .01$) muestran que la diferencia entre 2) y 3) alcanza significatividad estadística.

Respecto de la cantidad de contexto requerido para completar cada blanco, el índice ($\rho = -.31$, $p < .05$) alcanzó significatividad estadística (es negativo por el modo de adscribir los números a las categorías de contexto), pero significa correlación débil (Hernández-Sampieri, Fernández-Collado, & Baptista-Lucio, 2014). Esto es, si bien se halló la progresión esperada en las cuatro primeras categorías (la media de ND de los ítems en la oración es .80; en oraciones consecutivas, .77; en el párrafo, .58; en el texto, .56), por un lado, son cercanos los índices de los pares primeros y de los segundos y, por otro, los ítems que se ubican en la categoría *fuera del texto*, en promedio, resultaron de dificultad moderada (.61).

Tabla 4

Niveles de lectura de las palabras léxicas y funcionales.

		Funcionales			Total
		De frustración	Instruccional	Independiente	
Léxicas	De frustración	10	10	4	24
	Instruccional	4	23	14	41
	Independiente	1	6	10	17
Total		15	39	28	82

A fin de analizar la incidencia del número de respuestas alternativas, esta variable se categorizó desde la mediana (rango = 0 a 15; mediana = 4) en dos grupos: 1) sin alternativa (términos técnicos) y hasta tres; y 2) cuatro o más alternativas. En el primer grupo, se ubicaron 24 palabras; en el segundo, las restantes 34.

La prueba *t* de Student manifestó que las diferencias ($X = .54$, $DE = .22$; $X = .70$, $DE = .13$) alcanzan significatividad estadística ($t = -3.514$, $gl = 56$, $p < .01$). En efecto, solo dos términos del grupo con más alternativas se asignaron a difíciles (n° 26, *realizar*, y n° 51, *lenguaje*); el resto se distribuyó desde ND moderado a muy fácil. Y viceversa: en el subgrupo sin / con menos alternativas se ubicaron seis ítems moderados (25%), ocho de los diez difíciles (n° 4, *resulta*; n° 7, *dominio*; n° 15, *posgrado*; n° 27, *involucrados*; n° 29, *discurso*; n° 46, plan de *escritura*; n° 53, *revisión*; y n° 63, *ortográficas*) y los cinco blancos muy difíciles.

Finalmente, en relación con otras ocurrencias de las palabras de contenido en el texto, las diferencias ($X = .66$, $DE = .19$; $X = .61$, $DE = .17$) no resultaron estadísticamente significativas ($t = 1.503$, $gl = 56$, $p = .297$).

En consecuencia, en el grupo con ocurrencias interesó analizar la asociación de esta variable con la cantidad de contexto requerido para completar cada blanco. Para ello, se diferenciaron dos niveles según se hallaren: 1) dos o más ocurrencias o una muy cercana (22 blancos) y 2) una lejana (9 blancos). La combinación de estos

criterios mostró que las diferencias ($X = .75$, $DE = .14$; $X = .46$, $DE = .16$) fueron estadísticamente significativas ($t = 5.136$, $gl = 29$, $p < .001$). En efecto, mientras que en el primer subgrupo se ubicó un único ítem difícil (n° 46, plan de *escritura*) y el resto se distribuyó desde ND moderado a muy fácil, en el segundo se posicionaron un blanco fácil (n° 42, *etapa*), uno moderado (n° 43, *Puesta*), cuatro difíciles (n° 15, *posgrado*, n° 29, *discurso*, n° 51, *lenguaje*, y n° 53, *revisión*) y tres muy difíciles (n° 14, *Así*, n° 34, *comunicativa*, y n° 61, *tarea* de escritura).

Discusión

El principal objetivo del presente trabajo consistió en identificar los factores intrínsecos al cloze que afectan el nivel diferencial de dificultad de los ítems: palabras de contenido o funcionales y, en el caso de las primeras, cantidad de contexto requerido para completar los blancos, número de respuestas alternativas y de otras ocurrencias del término en el texto. La relevancia de este análisis, a nuestro juicio, reside en que —a excepción de una investigación temprana y preliminar (Difabio de Anglat, 2008)— solo se dispone de este tipo de análisis respecto de textos en inglés y en portugués.

El instrumento aplicado resultó más sencillo que el ideal en cuanto aumenta el porcentaje de ítems fáciles y disminuye el correspondiente a muy difíciles. Sin embargo, a pesar del método

de calificación (palabra aceptable), en el nivel de frustración se ubicó una proporción alta de participantes, si se tiene en cuenta que son graduados universitarios que leen un cloze con un nivel promedio adecuado de dificultad (el cloze administrado incluso supera levemente dicho nivel promedio). Si bien podría pensarse que el tema del texto explica este resultado, también los doctorandos en Letras se posicionan en este nivel en porcentaje similar a los cursantes con formación en Derecho, Filosofía y Psicología.

Con relación a la incidencia de los factores intrínsecos en el nivel de dificultad (ND) de los ítems, los resultados parecen confirmar tres de los supuestos iniciales: tienden a ser más fáciles los ítems funcionales que los lexicales, los que requieren menor cantidad de contexto para restaurar las palabras suprimidas y los que registran alternativas conceptualmente aceptables. Sin embargo, no se verificó que el ND de los ítems fuera afectado por el número de ocurrencias de los términos en el texto.

Respecto del primer supuesto, en consonancia con trabajos en lengua inglesa (Abraham & Chapelle, 1992; Brown, 1989; Kobayashi, 2002; Trace et al., 2017), portuguesa (p. ej., Angeli Dos Santos et al., 2002; De Oliveira, 2011) y española (Difabio de Anglat, 2008), se confirma la dificultad superior de recuperación de las palabras léxicas en comparación con las funcionales (por ello, en pro de incrementar la entropía, el instrumento *ex profeso* incluyó un porcentaje considerablemente mayor de las primeras: 84%). Mientras que las palabras léxicas pertenecen a clases abiertas y contienen información semántica, las funcionales corresponden a clases sintácticas cerradas con un número restringido de ítems que son conocidos por prácticamente todos los hablantes. Es evidente, entonces, la probabilidad de su predicción acertada. Inclusive en las palabras funcionales dos estudiantes (sorprendentemente

doctorandos con formación en Educación, no en Letras) lograron resolución perfecta, desempeño que no se halló respecto de las léxicas. En suma, las primeras resultan más sencillas que las segundas, aun cuando tres blancos se ubicaron entre los ítems difíciles (n° 39, *como*, n° 57, *su*, y n° 75, *esta*); los dos últimos manifiestan la dificultad de la referencia anafórica, exigida por la precisión esperada en un texto académico.

Asimismo, el blanco más difícil corresponde al gerundio (*considerando*), una clase de palabra de contenido. La mayor dificultad del gerundio podría explicarse por su carácter de unión sin marca alguna de dos palabras (*que considera*); por otra parte, se trata de un gerundio adjunto o circunstancial, función que establece una relación “menos directa con el verbo principal” que otras funciones del gerundio (Real Academia Española, 2010, p. 513).

Respecto de la cantidad de contexto requerido para completar cada blanco, el índice de correlación en la muestra en estudio es cercano al que informan Abraham y Chapelle (1992) y levemente más alto que el de Kobayashi (2002); los tres coeficientes indican correlación débil. En este sentido, si bien algunos ítems que pueden completarse con información colindante (esto es, con información del contexto inmediato) resultan fáciles (p. ej., n° 25, *resolución*), otros también lo son (p. ej., n° 68, *elementos*) a pesar de requerir un término que no aparece en el texto. Este resultado podría asociarse con la frecuencia de uso de los vocablos desde el Corpus de Referencia del Español Actual —CREA— en cuanto, para cada término, presenta la frecuencia de uso absoluta y normalizada, un análisis de interés para un estudio futuro. Esta variable parece explicar la recuperación prácticamente perfecta, y consiguiente falta de discriminación, de expresiones muy familiares para graduados universitarios (por ejemplo, llevar a *cabo*).

En el caso de los blancos que mostraron rangos variables de ND desde moderado a muy difícil, cabe la observación de Kobayashi (2002) sobre la vinculación compleja de los términos que primordialmente transmiten significados con varias partes del texto; ello dificulta la localización fehaciente de las claves contextuales con las que dichas palabras se relacionan. Al respecto, solo Jonz (1990) proporciona criterios de identificación específicos, aunque priman pautas relativas a palabras funcionales —conjunciones, preposiciones, pronombres, marcadores del discurso, entre otras— (ver Jonz, 1990, *Appendix A*, pp. 76-77). De hecho, es el propio lector quien puede *brindar certeza* sobre qué información ha empleado para restaurar la palabra omitida. En un trabajo futuro, entonces, convendrá aplicar el cloze en entrevista (en línea, en nuestro caso, esto es, en el taller virtual de escritura que implementamos desde 2015, en el cual se administra el texto cloze como parte del diagnóstico) con el objeto de que el participante indique por qué eligió una respuesta en particular.

En cambio, sí es clara la relación entre número de respuestas alternativas y ND: en el subgrupo de más alternativas se ubica el 75% del total de ítems lexicales moderados y el 79% de los fáciles, mientras que en el subgrupo con pocas alternativas o una sola (términos técnicos) se halla el 80% de los difíciles y la totalidad de los muy difíciles. También Abraham y Chapelle (1992) encuentran una diferencia significativa de medias a favor de los blancos que admiten más de una respuesta. Por el contrario, Kobayashi (2002) informa medias cercanas pero en relación inversa (.36 vs .44), resultado que explica en función de la mayor demanda cognitiva que podrían exigir los términos que suponen un amplio rango de posibles respuestas. Si bien aparece como una argumentación plausible en el caso de la competencia semántica en una lengua extranjera (estudiantes

universitarios japoneses que completan textos cloze en inglés), para hablantes nativos instruidos (maestrandos y doctorandos) es razonable concluir que los blancos cuya resolución *verbatim* refiere a palabras ricas en sinónimos son más fáciles de completar.

A diferencia de los resultados de Kobayashi (2002), de Brown (1989) y de Trace et al. (2017), pero en consonancia con los de Abraham y Chapelle (1992), el nivel de dificultad de los ítems no es afectado por el número de ocurrencias de los términos en el texto. El primer autor encuentra una correlación débil que fundamenta en la asociación estrecha entre el tema del cloze y las palabras léxicas que se repiten. En Brown (1989) la combinación que alcanza la correlación múltiple mayor, con puntuación de la respuesta exacta, está conformada por cuatro variables: palabra de contenido o de función, frecuencia con la cual la respuesta correcta aparece en el cloze, más dos variables relativas a la comprensibilidad de un texto (número de caracteres de la palabra y de sílabas de la oración que contiene el blanco). Trace et al. (2017) informan que la cantidad de ocurrencias, conjuntamente con otras dos variables de la comprensibilidad (número de oraciones en un pasaje y de sílabas por palabra), explican el 24% de la varianza total en la dificultad de los ítems.

En nuestro caso, identificamos una asociación estadística entre ocurrencia y cantidad de contexto requerido para completar el blanco: los ítems con una ocurrencia lejana evidenciaron mayor dificultad que aquellos con más ocurrencias o una cercana. Esto es, como muestra la investigación previa (Abraham & Chapelle, 1992; Brown, 1989; Kobayashi, 2002; Trace, 2020; Trace et al., 2017) aunque con soluciones disímiles, el nivel diferencial de dificultad de las palabras de contenido no admite una explicación simple sino que exige el análisis conjunto de factores. Ya que transmiten significados y cada significa-

do individual opera en la conformación de la red semántica del texto (Kobayashi, 2002) y que obviamente dependen de la competencia del lector para aprovechar tanto las claves textuales como su conocimiento general, requieren una habilidad integradora de nivel superior. Al respecto, son los estudiantes con menor competencia lectora (esto es, aquellos que se posicionaron en el nivel de frustración en el cloze) quienes solo consideran el contexto inmediato para completar los blancos (el ejemplo paradigmático es el término técnico *situación “comunicativa”*, ítem cuya dificultad estriba en la necesidad de atender a la información entre párrafos).

En síntesis, el análisis efectuado parece apoyar el postulado de varios autores (Abraham & Chapelle, 1992; Brown, 2013; Gellert & Elbro, 2013; Shahnazari-Dorcheh et al., 2012; Trace, 2020, entre otros) sobre la validez del cloze (en dependencia de su método de construcción y calificación) para evaluar comprensión global, dado que la puntuación que obtiene el lector puede ser entendida como un indicador de su competencia para emplear las claves textuales. En este sentido, en la presente investigación la dificultad de los blancos referidos a lenguaje especializado varía de modo consistente según la *brecha* entre los ítems (Trace, 2020) —cantidad de contexto requerido para completarlos—, la que exige que se establezca la conexión entre proposiciones separadas en el texto (Abraham & Chapelle, 1992; Gellert & Elbro, 2013).

Por último, cabe enfatizar que la principal limitación de este estudio es el tamaño de la muestra que no posibilita análisis estadísticos más potentes; su incremento en un número considerable permitiría, por ejemplo, como en Brown (1989) y en Trace et al. (2017), un análisis de regresión a fin de determinar qué combinación de factores explica mejor la varianza de la dificultad de los ítems.

Referencias

- Abraham, R., & Chapelle, C. (1992). The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76(4), 468-479. doi: [10.1111/j.1540-4781.1992.tb05394.x](https://doi.org/10.1111/j.1540-4781.1992.tb05394.x)
- Álvarez, G. (2011). Principios para un final: Reflexiones en torno a la escritura de la tesis de Maestría. *Monográfico de la Maestría en Educación* (pp. 43-53). Medellín, Colombia: Universidad Pontificia Bolivariana.
- Angeli Dos Santos, A. A., Primi, R., Taxa, F. de O. S., & Vendramini, C. M. M. (2002). O teste de Cloze na avaliação da compreensão em leitura. *Psicologia: Reflexão e Crítica*, 15(3), 549-560. doi: [10.1590/S0102-79722002000300009](https://doi.org/10.1590/S0102-79722002000300009)
- Bachman, L. (1985). Performance on Cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19(3), 535-556. doi: [10.2307/3586277](https://doi.org/10.2307/3586277)
- Backhoff-Escudero, E., Larrazolo-Reyna, N., & Rosas-Morales, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1), 11-28. Recuperado de <https://redie.uabc.mx/redie>
- Brown, F. G. (1999). *Principios de la Medición en Psicología y Educación* (8ª ed.). México, DF: Manual Moderno.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64(3), 311-317. doi: [10.1111/j.1540-4781.1980.tb05198.x](https://doi.org/10.1111/j.1540-4781.1980.tb05198.x)
- Brown, J. D. (1988). Tailored cloze: Improved

- with classical item analysis techniques. *Language Testing*, 5(1), 19-31. doi: [10.1177/026553228800500102](https://doi.org/10.1177/026553228800500102)
- Brown, J. D. (1989). Cloze item difficulty. *Journal of the Japan Association of Language Teachers*, 11(1), 46-67. Recuperado de <http://jalt-publications.org/jj>
- Brown, J. D. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies*, 7(1), 1-32. Recuperado de <http://www.ijls.net>
- Cairney, T. H. (2018). *Enseñanza de la comprensión lectora* (6ª ed.). Madrid, España: Morata.
- Calderón-Ibáñez, A., & Quijano-Peñuela, J. (2010). Características de comprensión lectora en estudiantes universitarios. *Revista Estudios Socio-Jurídicos*, 12(1), 337-364. Recuperado de <https://revistas.urosario.edu.co/index.php/sociojuridicos>
- Carlisle, J., & Rice, M. (2004). Assessment of reading comprehension. En A. Stone, E. Silliman, B. Ehren & K. Apel (Eds.), *Handbook of Language and Literacy: Development and disorders* (pp. 521-555). New York, NY: Guilford.
- Condemarín, M., & Milicic, N. (1988). *Test de cloze: Aplicaciones psicopedagógicas*. Santiago de Chile, Chile: Andrés Bello.
- Cortada de Kohan, N. (2004). Teoría de respuesta al ítem: Supuestos básicos. *Revista Evaluar*, 4(1), 95-110. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Costa, P. (2006). *Hábito de leitura e compreensão de textos: Uma análise da realidade de pós-graduados em Administração*. (Tesis de Maestría). Recuperado de repositorio.ufsm.br/handle/1/4735
- Cubo de Severino, L., & Bosio, I. V. (2011). La tesis como clase textual y su proceso de escritura. En L. Cubo de Severino, H. Puiatti, & N. Lacon (Eds.), *Escribir una Tesis. Manual de Estrategias de Producción* (pp. 13-36). Córdoba, Argentina: Comunicarte.
- Chino-Vilca, B., & Zegarra-Valdivia, J. (2019). Implicancia de las funciones ejecutivas “frías y calientes” en la comprensión lectora, en una muestra de estudiantes universitarios peruanos. *Psicología, Conocimiento y Sociedad*, 9(1), 59-82. Recuperado de <https://revista.psico.edu.uy/index.php/revpsicologia>
- Datta, S., & Macdonald-Ross, M. (2002). Reading skills and reading habits: A study of new Open University undergraduate reserves. *Open Learning*, 17(1), 69-88. doi: [10.1080/02680510120110193](https://doi.org/10.1080/02680510120110193)
- De Oliveira, K. L. (2011). Considerações acerca da compreensão em leitura no ensino superior. *Psicologia: Ciência e Profissão*, 31(4), 690-701. doi: [10.1590/S1414-98932011000400003](https://doi.org/10.1590/S1414-98932011000400003)
- Difabio de Anglat, H. (2008). El test cloze en la evaluación de la comprensión del texto informativo de nivel universitario. *Revista de Lingüística Teórica y Aplicada*, 46(1), 121-137. Recuperado de <http://www2.udec.cl/rla/principal.htm>
- Esquivel-Gámez, I., Martínez-Olvera, W., Córdoba del Valle, R., & Reyes-Gutiérrez, C. (2016). Memoria operativa y lectura comprensiva: Medición con pruebas de amplitud lectora y tipo cloze en ámbitos pre y universitarios. *Apertura. Revista de Innovación Educativa*, 8(2), 38-53. Recuperado

de <http://www.udgvirtual.udg.mx/apertura/index.php/apertura/index>

- Gellert, A., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment, 31*(1), 16-28. doi: [10.1177/0734282912451971](https://doi.org/10.1177/0734282912451971)
- González-Moreyra, R. (1998). Comprensión lectora en estudiantes universitarios iniciales. *Persona, 1*, 43-65. Recuperado de <https://dialnet.unirioja.es/servlet/revista?codigo=12630>
- Green, B. (2001). Testing reading comprehension of theoretical discourse with cloze. *Journal of Research in Reading, 24*(1), 82-98. doi: [10.1111/1467-9817.00134](https://doi.org/10.1111/1467-9817.00134)
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la Investigación* (6ª ed.). México, DF: McGraw-Hill.
- IBM Corporation. (2017). IBM SPSS Statistics for Windows (Versión 25.0) [Software de cómputo]. Armonk, NY: IBM.
- Jonz, J. (1990). Another turn in the conversation: What does cloze measure? *TESOL Quarterly, 24*(1), 61-83. doi: [10.2307/3586852](https://doi.org/10.2307/3586852)
- Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Utrecht, Países Bajos: LOT. Recuperado de <https://dspace.library.uu.nl/handle/1874/363346>
- Kleijn, S., Pander-Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing, 36*(4), 553-572. doi: [10.1177/0265532219840382](https://doi.org/10.1177/0265532219840382)
- Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. Londres, Reino Unido: Routledge.
- Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal, 86*(4), 571-586. doi: [10.1111/1540-4781.00162](https://doi.org/10.1111/1540-4781.00162)
- León-Cascón, J. A. (2004). ¿Por qué las personas no comprenden lo que leen? *Psicología Educativa, 10*(2), 101-116. Recuperado de <https://journals.copmadrid.org/psed>
- Lozano-Fernández, L. M., & de la Fuente-Solana, E. I. (2015). Diseño y validación de cuestionarios. En E. Pantoja-Vallejo (Coord.), *Manual básico para la realización de tesis, tesis y trabajos de investigación* (2ª ed., pp. 247-272). Madrid, España: EOS.
- Márquez, C., Ibáñez, P., & Pérez, C. (2011). Concordancia entre aplicaciones de textos informativos y científicos elaborados mediante la técnica Cloze en alumnos universitarios de carreras del área de la Salud. *Revista de Educación en Ciencias de la Salud, 8*(2), 133-138. Recuperado de <http://www2.udec.cl/ofem/recs>
- Oller, J. W., & Jonz, J. (Eds.). (1994). *Cloze and coherence*. Lewisburg, PA: Bucknell University Press.
- O'Toole, J. M., & King, R. A. R. (2010). A matter of significance: Can sampling error invalidate cloze estimates of text readability? *Language Assessment Quarterly, 7*(4), 303-316. doi: [10.1080/15434303.2010.515699](https://doi.org/10.1080/15434303.2010.515699)
- O'Toole, J. M., & King, R. A. R. (2011). The deceptive mean: Conceptual scoring of cloze entries differentially advantages more able

- readers. *Language Testing*, 28(1), 127-144. doi: [10.1177/0265532210375687](https://doi.org/10.1177/0265532210375687)
- Parodi, G. (2001). Comprensión y producción lingüística: Una nueva mirada al procesamiento del discurso escrito. *Versión. Estudios de Comunicación y Política*, 11, 59-97. Recuperado de <http://version.xoc.uam.mx/index.php>
- Peña-Calvo, A. (1997). El Cloze como técnica para enseñar estrategias de comprensión lectora en una lengua extranjera (inglés). En Departamento de Filología Inglesa, Universidad de Murcia (Ed.), *Estudios de lingüística aplicada y literatura: Homenaje póstumo al profesor Juan Conesa Sánchez* (pp. 203-215). Murcia, España: Universidad de Murcia, Servicio de Publicaciones.
- Real Academia Española. (2010). *Nueva gramática de la lengua española*. Manual. Madrid, España: Espasa.
- Real Academia Española. (2019). *Corpus de Referencia del Español Actual -CREA-* [Base de datos]. Recuperado de <http://corpus.rae.es>
- Sadeghi, K. (2008). Measuring reading comprehension: The judgmental validity of cloze procedure. *Iranian Journal of Applied Linguistics*, 11(2), 115-132. Recuperado de https://ijal.khu.ac.ir/index.php?slc_lang=en&sid=1
- Shahnazari-Dorcheh, M., Roshan, S., & Hesabi, A. (2012). What is the optimum length of a cloze test? *International Journal of English Linguistics*, 2(5), 142-153. doi: [10.5539/ijel.v2n5p142](https://doi.org/10.5539/ijel.v2n5p142)
- Tabullo, A. J., Shalom, D., Sevilla, Y., Gattei, C. A., París, L., & Wainselboim, A. (2020). Reading comprehension and predictability effects on sentence processing: An event-related potential study. *Mind, Brain, and Education*, 14(1), 32-50. doi: [10.1111/mbe.12205](https://doi.org/10.1111/mbe.12205)
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism and Mass Communication Quarterly*, 30(4), 415-433. doi: [10.1177/107769905303000401](https://doi.org/10.1177/107769905303000401)
- Trace, J. (2020). Clozing the gap: How far do cloze items measure? *Language Testing*, 37(2), 235-253. doi: [10.1177/0265532219888617](https://doi.org/10.1177/0265532219888617)
- Trace, J., Brown, J. D., Janssen, G., & Kozhevnikova, L. (2017). Determining cloze item difficulty from item and passage characteristics across different learner backgrounds. *Language Testing*, 34(2), 151-174. doi: [10.1177/0265532215623581](https://doi.org/10.1177/0265532215623581)
- Watanabe, Y., & Koyama, D. (2008). A meta-analysis of second language cloze testing research. *Second Language Studies*, 26(2), 103-133. Recuperado de <https://scholarspace.manoa.hawaii.edu/handle/10125/40694?mode=full>

Anexo 1. Texto cloze

En el siguiente texto se han suprimido varias palabras para que las anticipes.

Debes colocar una palabra en cada blanco (destacándola con color); los términos omitidos, cualquiera sea su longitud, han sido reemplazados por una línea de extensión constante.

Nota: agregamos el número de cada blanco para facilitar la lectura de la clave de corrección; no se incluyen en el texto presentado a los participantes del taller.

Principios para un final: reflexiones en torno a la escritura de la tesis

La escritura, la presentación y la defensa de la tesis son las etapas conclusivas de los estudios de posgrado, que exigen al estudiante el ordenamiento del conocimiento que ha acumulado, analizado y desarrollado durante su carrera. Este 1. ----- lo compromete con dos 2. ----- interdependientes (de investigación y de 3. -----), de los cuales 4. ----- un producto final llamado 5. ----- . La escritura, por 6. ----- parte, junto con el 7. ----- del tema de investigación, 8. ----- el conocimiento de las 9. ----- discursivas necesarias para la 10. ----- de textos, en particular 11. ----- que permiten elaborar 12. ----- académicos (e.g. el proyecto de 13. -----, la monografía, la tesis). 14. -----, en el nivel de 15. -----, el estudiante, en su 16. ----- de escritor, debe resolver 17. ----- relativos tanto al contenido 18. ----- a la comunicación de 19. ----- contenido.

La escritura suele 20. ----- como una actividad que 21. ----- de un problema comunicativo 22. ----- resolución depende de la 23. ----- de un texto escrito. En este 24. -----, a fin de alcanzar con éxito la 25. ----- del problema, es necesario 26. ----- cada uno de los procesos 27. ----- en toda tarea de composición 28. -----: la planificación, la puesta en 29. ----- y la revisión, que 30. ----- alternan y se influyen 31. ----- durante todo el 32. -----.

La planificación 33. ----- una evaluación inicial de la situación 34. -----: tema, objetivos del 35. -----, destinatarios, formato o género, extensión, 36. ----- del que se dispone 37. ----- elaborarlo. Conduce a 38. ----- un plan de escritura, 39. ----- un índice comentado o una 40. ----- de tareas por hacer. 41. ----- evidente que detenerse en esta 42. ----- facilita la producción del texto. 43. ----- en discurso significa 44. ----- en palabras las ideas, 45. ----- como guía el plan de 46. ----- (más o menos 47. -----). Para llevar a 48. ----- esta tarea, es necesario 49. ----- las determinaciones impuestas 50. ----- la gramática de la lengua y el tipo de 51. ----- (e.g. vocabulario, sintaxis, frases frecuentes) 52. ----- por la situación comunicativa.

La 53. ----- comprende dos subprocesos: 54. ----- y edición.

En la evaluación, se 55. ----- la adecuación del texto 56. ----- a diferentes niveles: desde 57. ----- funcionamiento global 58. ----- la ortografía. Así, se 59. ----- que el texto se 60. ----- al problema planteado por la 61. ----- de escritura y al plan trazado, y su 62. ----- a las reglas gramaticales y 63. -----.

La edición busca 64. ----- los problemas del texto a 65. ----- local o global. Las operaciones de 66. ----- podrían describirse como 67. ----- de borrado o supresión de 68. -----, agregado o 69. -----, desplazamiento a lo largo del sintagma, etc.

Reflexionar 70. ----- estos procesos, 71. ----- conscientes nuestros propios 72. ----- mentales, es un punto de 73. ----- confiable para aprender a resolver los 74. ----- planteados por la escritura. A 75. ----- reflexión sobre la cognición, 76. -----, puede sumarse una 77. ----- sobre los usos 78. ----- del lenguaje y las 79. ----- epistemológicas de esos usos, es decir, la reflexión metalingüística. Ambas son instrumentos útiles para mejorar nuestra tarea como escritores académico-científicos.

Anexo 2. Clave de corrección

En el siguiente listado, la primera palabra (en letra itálica) corresponde al ítem *verbatim*; el resto, a las alternativas correctas ordenadas por frecuencia en la muestra; las casillas en color gris, a los ítems no discriminativos.

1. <i>ordenamiento</i> , proceso, trabajo, itinerario, quehacer
2. <i>procesos</i> , tareas, etapas/fases, actividades/acciones, objetivos/propósitos, instancias
3. <i>escritura</i> , redacción, comunicación
4. <i>resulta/resultará</i> , deviene
5. <i>tesis</i>
6. <i>su</i> , otra
7. <i>dominio</i> , desarrollo, estudio, conocimiento
8. <i>exige</i> , requiere, supone, necesita, implica, comprende, demanda, involucra
9. <i>estrategias</i> , herramientas, técnicas, formas, reglas, normas, estructuras, modalidades, prácticas
10. <i>producción</i> , redacción, elaboración, escritura, construcción, composición, estructuración
11. <i>aquellas</i> , las
12. <i>textos</i> , escritos, géneros
13. <i>investigación</i> , tesis
14. <i>Así</i> , Entonces, Ciertamente, Especialmente
15. <i>posgrado</i> , doctorado
16. <i>rol</i> , papel, carácter, función, condición
17. <i>aspectos</i> , problemas, asuntos, temas, tópicos, interrogantes
18. <i>como</i>
19. <i>dicho</i> , ese, este, tal
20. <i>concebirse</i> , presentarse, considerarse, entenderse, verse, definirse, aparecer, plantearse, pensarse, interpretarse, caracterizarse, proponerse, identificarse, comprenderse
21. <i>parte</i> , surge, resulta, procede, deriva, deviene
22. <i>cuya</i>
23. <i>elaboración</i> , producción, redacción, composición, construcción, formulación, realización, preparación
24. <i>sentido</i> , proceso, trayecto, camino, trabajo
25. <i>resolución</i> , solución
26. <i>realizar</i> , considerar, conocer, respetar, concretar, desarrollar, desplegar, entender, dominar, transitar, recorrer, atravesar, aplicar

27. <i>involucrados</i> , implicados, presentes, contenidos
28. <i>escrita</i> , textual
29. <i>discurso</i> , palabras, texto
30. <i>se</i>
31. <i>mutuamente</i> , dinámicamente, recíprocamente, respectivamente
32. <i>proceso</i>
33. <i>implica</i> , es, requiere, supone, constituye, conlleva, involucra, representa, proporciona, contiene, comprende, prevé
34. <i>comunicativa</i> , comunicacional, escritural
35. <i>texto</i> , trabajo, escrito, autor, investigador, discurso, escritor, comunicador, emisor
36. <i>tiempo</i> , material
37. <i>para</i>
38. <i>producir</i> , realizar, diseñar, establecer, desarrollar, generar, trazar, armar, formular, esbozar, pensar, confeccionar, concretar, definir, estructurar, avanzar
39. <i>como</i> , con, elaborando
40. <i>lista</i> , enumeración, guía, grilla, agenda, tabla, sistematización, descripción
41. <i>Es</i> , resulta, parece
42. <i>etapa</i> , tarea, instancia, actividad, fase
43. <i>Puesta/poner</i>
44. <i>expresar</i> , poner, plasmar, transformar, traducir, trasponer, presentar, comunicar, materializar
45. <i>considerando</i> , teniendo, siguiendo, utilizando
46. <i>escritura</i> , texto
47. <i>detallado</i> , definido, preciso, explícito/explicitado, elaborado, delimitado, desarrollado, planificado, programado, estructurado, especificado, esbozado, bosquejado, completo
48. <i>cabo</i>
49. <i>conocer</i> , respetar, considerar, seguir, cumplir, revisar, comprender
50. <i>por</i> , desde
51. <i>lenguaje</i> , discurso, registro, recursos, género, código
52. <i>requerido</i> , exigido, determinado, impuesto, establecido, definido, demandado, delimitado, condicionado
53. <i>revisión</i>
54. <i>evaluación</i>
55. <i>verifica</i> , analiza, observa, revisa, valora, examina, comprueba, considera, controla, determina, pondera, juzga

56. <i>elaborado</i> , escrito, producido, realizado, redactado, construido, final
57. <i>su</i>
58. <i>hasta</i> , a
59. <i>corroborar</i> , evalúa, asegura, comprueba, verifica, busca, controla, garantiza, confirma, establece, valora
60. <i>ajuste</i> , adecue, adapte, atenga, convenga
61. <i>tarea</i> , situación, actividad
62. <i>adecuación</i> , ajuste, sujeción, fidelidad
63. <i>ortográficas</i> , formales, lingüísticas
64. <i>solucionar</i> , resolver, corregir, subsanar, superar, acomodar, eliminar, despejar
65. <i>nivel</i> , escala
66. <i>edición</i>
67. <i>operaciones</i> , tareas, acción/es, proceso/s, actividades, las, procedimientos, técnicas, trabajo, ejercicio, labor, estrategias, rutinas, aplicación, aquellas
68. <i>elementos</i> , palabras, texto, errores, frases, contenido, información, términos, erratas, fragmentos, ideas
69. <i>inserción</i> , modificación/es, adición, reemplazo, sustitución, ampliación, cambio/s, aclaración, incorporación, mejora/s, corrección, expansión, reducción, explicitación, ajustes, inclusión
70. <i>sobre</i> , en
71. <i>hacer</i> , haciendo, volver, tornar, volviendo
72. <i>procesos</i> , esquemas, mecanismos, procedimientos, estados
73. <i>partida</i>
74. <i>problemas</i> , desafíos, retos, inconvenientes, obstáculos
75. <i>esta</i> , nuestra, esa
76. <i>además</i> , también, asimismo, finalmente, conjuntamente, ciertamente
77. <i>reflexión</i> , consideración, mirada, más, deliberación, discusión, concientización, evaluación, valoración, exploración, similar
78. <i>académicos</i> , científicos, especializados
79. <i>razones</i> , bases, consecuencias, reglas, implicancias, normas, raíces, derivaciones, fundamentaciones, condiciones, especificaciones, fuentes, pautas, categorías

Análisis psicométrico del inventario de orientación suicida ISO-19, en adolescentes cordobeses escolarizados

Psychometric analysis of the Inventory of Suicide Orientation ISO-19, in scholarized adolescents

Gerardo Ariel Vecco ¹ *, Pablo Ezequiel Flores-Kanter ^{2,3}, Leticia Elizabeth Luque ¹

1 - Facultad de Psicología, Universidad Nacional de Córdoba, Córdoba, Argentina.

2 - Secretaría de Investigación, Universidad Siglo 21, Córdoba, Argentina.

3 - Centro de Bioética, Universidad Católica de Córdoba.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 22/09/2020 Revisado: 05/11/2020 Aceptado: 15/11/2020

Resumen

Contar con instrumentos válidos y confiables para detectar el riesgo de suicidio es deseable en materia de prevención. En este estudio se explorarán algunas propiedades psicométricas de la versión reducida del Inventario ISO-30 en un grupo de 656 adolescentes escolarizados, de 14-20 años de edad, en Córdoba (Argentina). Para evaluar el ajuste de los datos empíricos al diseño identificado por Galarza, Fernández-Liporace, Castañeiras y Freiberg-Hoffmann (2019) para el ISO-19, se realizaron un análisis factorial confirmatorio (AFC) y un modelamiento exploratorio de ecuaciones estructurales (ESEM). Respecto del AFC, los índices de ajuste del modelo fueron óptimos, levemente inferiores a los obtenidos mediante el modelo ESEM. La cuantificación de la simplicidad factorial identificó 17 ítems por encima del criterio establecido como aceptable. Se concluye que la eliminación de reactivos de la versión original (ISO-30) y los nuevos agrupamientos identificados deben ser considerados para el índice global de orientación suicida que se utiliza en nuestro país.

Palabras clave: *adolescencia, suicidio, escala ISO-19, escala ISO-30, propiedades psicométricas*

Abstract

In terms of prevention, it is desirable to have valid and reliable instruments to detect the risk of suicide. Some psychometric properties of the reduced version of ISO-30 are explored, in a group of 656 school-aged adolescents, 14-20 years old, from Córdoba (Argentina). To evaluate the fit of empirical data to the design identified by Galarza, Fernández-Liporace, Castañeiras, and Freiberg-Hoffmann (2019) for ISO-19, a confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM) were performed. Regarding the CFA, the indices for evaluating the model fit were optimal and slightly lower than those obtained using the ESEM model. The quantification of factorial simplicity identified 17 items above the criteria established as acceptable. The elimination of reagents from the original version (ISO-30) and the new identified groupings should be considered for the global suicide orientation index used in our country.

Keywords: *adolescence, suicide, ISO 19 scale, ISO 30 scale, psychometric properties*

*Correspondencia a: Gerardo A. Vecco. Dirección postal: José Ingenieros 70, dpto. 19, Villa Carlos Paz (CP 5152). Teléfono: +54-3541-516210. E-mail: gerave62@gmail.com

Cómo citar este artículo: Vecco, G. A., Flores-Kanter, P. E., & Luque, L. E. (2021). Análisis psicométrico del inventario de orientación suicida ISO-19, en adolescentes cordobeses escolarizados. *Revista Evaluar*, 21(1), 40-52. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

La Asociación Argentina de Prevención del Suicidio define al *suicidio* “como un proceso complejo multideterminado. Una manera de vivir que va construyendo un enigma mortal, por condensación, sobre un marco representacional existente. Un proceso que va más allá del acto, rompiendo la concepción del hecho consumado” (Martínez, 2007, p. 19). Sin negar la evolución continua de los términos y el uso de diferentes expresiones en otros ámbitos del sector, la OMS (2014) entiende por *suicidio* al acto de matarse deliberadamente y al *intento suicida* como todo comportamiento que no causa muerte e incluye intoxicación autoinfligida, lesiones y autoagresiones intencionales. El comportamiento suicida incluye la ideación, la planificación, el intento suicida y la comisión del acto. La OMS (2014) aclara que la inclusión de la ideación —como parte del comportamiento— supone un diálogo académico en el que hay diversidad de posiciones no coincidentes.

Desde el punto de vista epidemiológico, el suicidio es considerado un problema grave de salud pública. Por año, alrededor del mundo, se suicidan más de 800.000 personas, lo cual convierte al suicidio en la segunda causa principal de muerte en el grupo etario de entre 15 y 29 años. Asimismo, se estiman 20 intentos por cada suicidio consumado (OMS, 2014). La Dirección de Estadísticas e Información de Salud, del Ministerio de Salud de la República Argentina (DEIS, 2016), informó que entre los años 2014 y 2016 se registraron 1.378 defunciones por suicidio y una tasa trienal de mortalidad adolescente por suicidio de 6.5 por cada 100000 habitantes, entre las que se incluyen las categorías de lesiones autoinfligidas y secuelas de lesiones autoinfligidas de la clasificación CIE-10 de enfermedades y problemas relacionados con la salud. Desde una tasa de 2.5 (n = 467), en el trienio 1990-1992, la

mortalidad de adolescentes por suicidio en nuestro país se elevó a una tasa de 7.2 para el trienio 2011-2013. En 2016, las defunciones adolescentes por suicidio fueron 419, de las cuales 375 correspondían al grupo de 15 a 19 años de edad y 282 correspondían a varones (DEIS, 2016).

Estos datos indican la relevancia que el tema tiene; sin embargo, la producción empírica sobre el suicidio en Argentina es baja (Flores-Kanter, 2017) y los modelos teóricos predictivos del suicidio evaluados en la región son escasos (Flores-Kanter, García-Batista, Moretti, & Medrano, 2019). Por este motivo, resulta imperante contar con instrumentos válidos y confiables para la medición de variables asociadas al riesgo suicida.

En algunos estudios se han indagado los instrumentos disponibles para la medición del riesgo suicida. Rangel-Garzón, Suárez-Beltrán y Escobar-Córdoba (2015), de Colombia efectuaron una revisión de numerosas escalas de riesgo suicida. Su análisis concluye que, si bien algunas escalas cuentan con buenas propiedades psicométricas, la mayoría de los instrumentos son de larga data, con escasos estudios actuales sobre dichas propiedades. Ninguno alcanza elevados índices de sensibilidad y especificidad (lo que implica escasa capacidad predictiva del riesgo suicida) y la complejidad de su modo de aplicación impide su uso en atención primaria.

Runeson et al. (2017) evalúan, mediante un meta-análisis, el riesgo de sesgo y la precisión diagnóstica de instrumentos desarrollados hasta 2014 para la evaluación de riesgo suicida. Concluyen que la mayoría de los instrumentos presentan escasos estudios que evalúen su precisión y, entre los instrumentos analizados, ninguno cumple con los criterios de precisión diagnóstica, a saber: una sensibilidad superior al 80% y especificidad mayor del 50%.

Abarca, Gheza, Coda y Elicer (2018) realizaron una revisión bibliográfica sistemática para

identificar escalas de evaluación de riesgo suicida en adultos usuarios de servicios de atención primaria de salud. Concluyen que ninguna de ellas es lo suficientemente precisa y predictiva para justificar las intervenciones que se realizan con respecto a sus resultados.

De lo expuesto surge la necesidad de estandarizar y revisar constantemente los instrumentos que se utilizan porque de los resultados surgen las propuestas de intervenciones preventivas y posventivas. Es importante tener presente que las pruebas psicométricas se utilizan no solo en investigación, sino también en la toma de decisiones acerca de las posibilidades y oportunidades de vida de las personas (Coolican, 2005).

En Argentina se aplica —tanto con fines investigativos como para la implementación de acciones preventivas—, el Inventario de Orientación Suicida (ISO-30) diseñado por King y Kowalchuk (1994) y adaptado localmente por Casullo (1997). Además, la Subsecretaría de Juventud de la Nación y la Fundación de Prevención del Suicidio lo integran en el Programa de Entrenadores de Vida (Galarza, Fernández-Liporace, Castañeiras, & Freiberg-Hoffmann, 2019).

Goldston (2000) realizó una revisión de los instrumentos de evaluación de riesgo suicida orientados a niños y adolescentes. Respecto de la escala ISO-30, comunica los índices psicométricos publicados por los autores del diseño original: a) índices de consistencia interna alfa de Cronbach de .92 y .90 para la puntuación total y de .79 y .78 para los ítems críticos, en muestras clínicas y de estudiantes respectivamente; b) fiabilidad test-retest de .80 para la puntuación total y .70 para los ítems críticos, considerando un período de 3-4 días; c) índices de correlación de .64 y .52 para la puntuación total de la escala con el Cuestionario de Ideación Suicida SIQ de Reynolds, y de .55 y .78 con el SIQ-JR de King y Kowalchuk (1994), en muestras clínicas y de estudiantes respectiva-

mente. Se informa también que no se encontraron datos publicados respecto a la dimensionalidad o validez predictiva del instrumento, señalando que los métodos de validación se enfocaron solo en la dimensión *ideación suicida* (Goldston, 2000).

Osman et al. (2005) realizaron un análisis factorial confirmatorio (AFC) sobre el diseño pentafactorial de la escala original (ISO-30; King & Kowalchuk, 1994), con 202 adolescentes pacientes psiquiátricos hospitalizados, de ambos sexos, de entre 14 y 17 años de edad. Los resultados respaldan el modelo oblicuo original de cinco factores, la consistencia interna del instrumento y sus dimensiones; se informa evidencia de validez discriminante, convergente e incremental. Sin embargo, los autores advierten limitaciones del estudio porque los datos solo respaldan el uso apropiado en adolescentes pacientes psiquiátricos.

Paniagua-Suárez, González-Posada y Rueda-Ramírez (2016) realizaron un estudio psicométrico sobre el ISO-30, en una población de 604 estudiantes adolescentes ($M = 14$ años, $DE = 1.9$), 48% hombres y 52% mujeres, de Colombia. Se efectuó un AFC, que muestra que los resultados de los índices de bondad de ajuste no están dentro de los parámetros de referencia ideales, pero se considera que la estructura pentafactorial es mejor que cualquier estructura aleatoria alternativa. Se determinó la validez concurrente con la escala RFL-A validada para población colombiana. El índice alfa de Cronbach para la escala completa es de .867. Los autores concluyen que la escala es una herramienta útil para diagnosticar el riesgo de orientación suicida en adolescentes escolarizados de la ciudad de Medellín.

En Argentina se aplica el ISO-30 según la versión adaptada en 1997 por Casullo y, si bien en algunas investigaciones se informa que se utiliza la solución identificada por Fernández-Liporace y Casullo (2006), se sostienen los puntos de corte definidos por la adaptación inicial. Casullo (1997)

adaptó la escala ISO-30 para ser empleada en adolescentes de Buenos Aires (Argentina). La validez concurrente se obtuvo a partir de una muestra de 30 adolescentes en tratamiento psiquiátrico que presentaban indicadores de potencial suicida (habiendo sido evaluados con técnicas proyectivas); el 80% obtuvo un puntaje que lo ubicaba como de riesgo suicida alto y el 20% restante en el nivel moderado. El índice de consistencia interna alfa de Cronbach, para la escala completa, es de .92 en la muestra clínica y de .72 para la muestra de los estudiantes que participaron voluntariamente. El estudio test-retest se efectuó en un lapso de dos semanas entre ambas tomas, con 35 adolescentes estudiantes de una escuela pública de la ciudad de Buenos Aires, y obtuvo una correlación de .86 (Casullo, 1997).

Fernández-Liporace y Casullo (2006) realizaron un estudio psicométrico sobre la versión adaptada del ISO-30 con una muestra de 642 adolescentes de entre 13 y 20 años, escolarizados, de una escuela pública de la ciudad de Buenos Aires. Las autoras analizan la capacidad discriminante de los reactivos y la validación factorial utilizando el método de los componentes principales, con rotación varimax y criterio Kaiser con extracción por contraste de caída. Informan que se eliminan 4 reactivos de la escala original (ítems 10, 19, 28 y 29) y que la solución factorial identificada no responde exactamente a las especificaciones previstas por el diseño, pero que es coherente con la propuesta teórica de base; en virtud de esto, los 5 factores aislados no replican exactamente las dimensiones subyacentes originales y los índices alfa de Cronbach informados son: .88 para el Factor 1 (*ideación suicida*— ítems 15, 20, 25 y 30), .74 para el Factor 2 (*desesperanza*— ítems 1, 6, 8, 9, 12, 16, 18, 21, 23 y 24), .67 el Factor 3 (*esperanza*— ítems 2, 3, 4, 11, 17, 27), .43 para el Factor 4 (*pertenencia interpersonal*, que solo incluye dos ítems: 5 y 26) y .50 para el Factor 5 (*capacidad*

de autocontrol y relación con los otros— ítems 7, 13, 14 y 22). La escala, con 26 ítems, presenta un índice alfa de Cronbach de .87. Se puede apreciar que los factores 4 y 5 presentan índices de consistencia interna por debajo de lo considerado como aceptable por Campo-Arias y Oviedo (2008).

Galarza et al. (2019) realizaron el análisis de las propiedades psicométricas del instrumento en adolescentes y jóvenes universitarios de la ciudad de Mar del Plata (Buenos Aires). Con 821 casos llevaron a cabo un AFE, utilizando análisis paralelo, *minimum rank factor analysis* robusto para variables ordinales, con rotación promax y normalización Kaiser. Eliminaron 11 ítems (3, 6, 9, 10, 13, 16, 17, 19, 21, 23 y 28) y obtuvieron una estructura tetrafactorial: Factor 1 denominado *ideaciones suicidas* (ítems 15, 20, 25 y 30), Factor 2 como *soledad y aislamiento social* (ítems 4, 5, 14, 26, 29), Factor 3 como *desesperanza* (ítems 2, 7, 11, 22, 27) y Factor 4 como *baja autoeficacia* (ítems 1, 8, 12, 18, 24). Luego efectuaron un AFC con 297 casos, utilizando como método de estimación ML robusto, comparando el modelo tetrafactorial derivado del procedimiento exploratorio y un modelo pentafactorial, en donde se retienen los mismos 19 ítems, ubicándolos en las siguientes dimensiones: *baja autoestima* (1, 11, 26), *desesperanza* (2, 7, 12, 22 y 27), *incapacidad para afrontar emociones* (8, 18), *soledad y aislamiento social* (4, 14, 24 y 29), *ideación suicida* (5, 15, 20, 25 y 30). Los índices RMSEA son inferiores en ambos modelos a .05, mientras que los índices de ajustes NFI, NNFI, CFI, IFI y MFI superan en ambos modelos .90, presentando el modelo tetrafactorial valores más elevados. Los índices de parsimonia considerados CAIC Independiente y CAIC Modelo indican que el modelo tetrafactorial presenta una mayor parsimonia que el pentafactorial. Los coeficientes alfa ordinales obtenidos para analizar la consistencia interna del modelo tetrafactorial son: .95, .65, .71

y .69 para los factores *ideaciones suicidas, soledad y aislamiento social, desesperanza y baja autoeficacia*, respectivamente.

En virtud de lo anterior, el objetivo de la presente investigación fue analizar algunas propiedades psicométricas del Inventario de Orientación Suicida ISO-19 propuesto por [Galarza et al. \(2019\)](#), considerando que no se encontraron antecedentes de análisis factoriales exploratorios o confirmatorios para la provincia de Córdoba (Argentina), teniendo en cuenta las diferencias observadas entre las soluciones factoriales identificadas por [Fernández-Liporace y Casullo \(2006\)](#) y [Galarza et al. \(2019\)](#) y el diseño original ([King & Kowalchuk, 1994](#)), y el hecho de que el instrumento propone medir varios factores asociados a la conducta suicida en lugar de tomar uno o algunos, como ocurre con otras escalas, tales como la CTS (ideas suicidas), PHQ-9 (depresión), el BHS (soledad), entre otras.

Método

Participantes

Mediante un muestreo no probabilístico accidental ([Bologna, 2011](#)) con participación voluntaria, se seleccionaron 656 adolescentes escolarizados, de ambos sexos (52.7% mujeres y 47.3% hombres), provenientes de tres instituciones de educación secundaria de la provincia de Córdoba, cuyas edades oscilan entre los 14 y 20 años, con una media de edad de 15.82 (DE = 1.342) y 15.83 (DE = 1.348). respectivamente.

Instrumento

El Inventario de Orientación Suicida (ISO-30) es una escala que, según [King y Kowalchuk \(1994\)](#), permite evaluar cinco factores asociados

al riesgo suicida: *baja autoestima, desesperanza, incapacidad para afrontar problemas emocionales, soledad y aislamiento social e ideación suicida*. Consta de 30 ítems formulados en forma positiva y negativa (directos e inversos) que se responden mediante una escala Likert de cuatro posiciones: 0 = *Totalmente en desacuerdo*, 1 = *En parte en desacuerdo*, 2 = *En parte de acuerdo*, 3 = *Totalmente de acuerdo*. La suma de las respuestas numéricas se expresa en un índice global de la Orientación Suicida que determina la clasificación del riesgo suicida en tres niveles: bajo ≤ 30 ; 31 \leq moderado ≤ 44 ; alto ≥ 45 . Con independencia de la puntuación obtenida, seis reactivos son considerados ítems críticos (5, 10, 15, 20, 25, 30) ya que aluden directamente a la ideación suicida; la presencia de tres o más ítems respondidos con dos o tres puntos indica la existencia de riesgo elevado.

La ISO-19 ([Galarza et al., 2019](#)) es una versión reducida de la ISO-30 que fue adaptada por [Casullo \(1997\)](#) para Argentina. Consta de 19 ítems formulados en forma positiva y negativa (directos e inversos), que se responden mediante una escala Likert de cuatro posiciones: 0 = *Totalmente en desacuerdo*, 1 = *En parte en desacuerdo*, 2 = *En parte de acuerdo*, 3 = *Totalmente de acuerdo*. Según los autores permite evaluar cuatro factores asociados al riesgo suicida: *ideaciones suicidas, soledad y aislamiento social, desesperanza y baja autoeficacia*.

El instrumento se administró de forma colectiva a estudiantes de diferentes cursos provenientes de tres instituciones de nivel educativo medio (dos de gestión pública y una de gestión privada), en los horarios normales de clases. La participación fue voluntaria, y se especificó que los datos obtenidos serían tratados de forma confidencial y utilizados exclusivamente con fines investigativos. Previamente, se solicitaron las autorizaciones institucionales y el consentimiento

informado de padres o tutores; las primeras fueron gestionadas por la dirección de “La Casa del Joven”.

Análisis de datos

Para evaluar el ajuste de los datos empíricos al diseño identificado por [Galarza et al. \(2019\)](#) se realizaron el AFC y el ESEM ([Asparouhov & Muthén, 2009](#); [Ferrando & Anguiano-Carrasco, 2010](#); [Lloret-Segura, Ferreres-Traver, Hernández-Baeza, & Tomás-Marco, 2014](#)), utilizando los paquetes estadísticos IBM SPSS 23.0 ([IBM Corp., 2015](#)), Mplus 6.12 ([Muthén & Muthén, 1998-2011](#)), y FACTOR 10.8.04 ([Lorenzo-Seva & Ferrando, 2018](#)).

Corresponde aclarar que se complementó el AFC con la aplicación de ESEM, por ser este un método menos restrictivo al no especificar como nulas las cargas secundarias y porque, en ocasiones, las restricciones que impone el AFC podrían no reflejar la complejidad del constructo que se investiga. ESEM presenta una integración de las principales características del AFC, añadiendo la flexibilidad del AFE al estimar en sus análisis el valor de las cargas secundarias. Las investigaciones precedentes que aplicaron ESEM han demostrado la potencia estadística del método en cuanto a los resultados identificados, contribuyendo además en una recuperación rigurosa de las correlaciones interfactoriales informadas e incidiendo en las conclusiones que de ellas se derivan ([Trógolo et al., 2020](#); [Dominguez-Lara & Merino-Soto, 2018](#); [Marsh, Morin, Parker, & Kaur, 2014](#); [Marsh et al., 2010](#)).

Luego de realizar un análisis de exploración inicial de los datos para identificar casos atípicos ([Pérez & Medrano, 2010](#)), se investigó el grado de asimetría y curtosis considerando como indicadores de alejamiento significativo de la norma-

lidad a los valores superiores a tres y diez, respectivamente ([Kline, 2005](#)).

Se contempló una matriz de correlaciones policóricas teniendo presente que la escala posee 4 alternativas de respuesta; el índice de Mardia señaló un valor de curtosis superior a 70 cuando se verificó el cumplimiento de los supuestos de la normalidad multivariada, teniendo en cuenta que el índice RMSEA informado fue inferior a .05 ([Baglin, 2014](#); [Ferrando & Lorenzo-Seva, 2014](#); [Lloret-Segura et al., 2014](#); [Freiberg-Hoffmann, Stover, de la Iglesia, & Fernández-Liporace, 2013](#); [Gadermann, Guhn, & Zumbo, 2012](#)). La adecuación de la matriz de correlaciones se investigó a partir de la aplicación del test de esfericidad de Bartlett y de la medida de adecuación muestral KMO.

Mediante AFC y ESEM se evaluó el ajuste del modelo utilizando el índice de ajuste comparado (CFI), el índice de Tucker-Lewis (TLI), el error de aproximación cuadrático medio (RMSEA), y el residuo cuadrático medio ponderado (WRMR). Los índices de ajuste obtenidos se analizaron teniendo en cuenta los criterios definidos por [Hu y Bentler \(1999\)](#), considerando como indicadores de un ajuste aceptable los valores superiores a .90 en los índices CFI y TLI e inferiores a .08 y a 1.00 en RMSEA y WRMR, respectivamente. Se aplicó como método de estimación a los mínimos cuadrados ponderados robustos (WLSMV), cuyo uso resulta recomendado cuando se trabaja con una matriz de correlaciones policóricas y los datos se alejan de la normalidad ([Lloret-Segura et al., 2014](#)). Para el modelamiento de ecuaciones estructurales exploratorias (ESEM) se especificó una rotación target.

Respecto de las magnitudes de las cargas principales, se consideró un análisis preliminar desde los criterios que proponen [Tabachnick y Fidell \(2013\)](#), para luego investigar su impacto en la fiabilidad del constructo ([Dominguez-Lara,](#)

2018a; Dominguez- Lara, 2018b). Por otra parte, en el modelo ESEM, se evaluó el grado de simplicidad factorial de un ítem mediante el índice ISF (Fleming & Merino-Soto, 2005; Fleming, 2003), estableciendo como criterio de indicadores aceptables $ISF > .70$ (Dominguez-Lara & Merino-Soto, 2018; Dominguez-Lara & Navarro-Loli, 2018).

Para obtener los índices de consistencia interna se calculó el índice omega (Ventura-León & Caycho-Rodríguez, 2017), considerando aceptables magnitudes que se ubicaran entre [.70; .90] (Campo-Arias & Oviedo, 2008; Ventura-León & Caycho-Rodríguez, 2017).

Resultados

Se llevó a cabo un análisis de exploración inicial de los datos con el propósito de detectar la existencia de casos atípicos uni y multivariados. La adecuación de los ítems se determinó a través del análisis de la matriz de correlaciones anti-imagen.

Se calcularon los estadísticos descriptivos (Tabla 1). Los índices de asimetría y curtosis de los ítems analizados presentan magnitudes adecuadas, por debajo de 3 y 10 respectivamente. El índice de Mardia señaló un valor > 70.00 , mientras que RMSEA informó un valor inferior a .05. Los resultados de aplicar el test de esfericidad de Bartlett ($\chi^2 = 2526.2$; $gl = 171$; $p < .001$ y de la medida de adecuación muestral Kaiser-Meyer-Olkin (.839) indican la adecuación de la matriz de correlaciones.

Teniendo en cuenta estos elementos, así como el tamaño de la muestra y que se trabaja con una escala Likert de 4 posiciones, el análisis factorial se realizó a partir de una matriz de correlaciones policóricas.

Tabla 1

Estadísticos descriptivos de los ítems ISO 19.

	M	DE	g_1	g_2
Ítem 1	1.569	1.005	-.220	-1.030
Ítem 2	.529	.691	1.241	1.329
Ítem 4	.697	.925	1.147	.249
Ítem 5	.922	.907	.597	-.635
Ítem 7	1.384	1.028	.208	-1.088
Ítem 8	2.299	.941	-1.165	.250
Ítem 11	.742	.838	.949	.185
Ítem 12	1.261	1.023	.240	-1.098
Ítem 14	.924	.944	.763	-.367
Ítem 15	.255	.663	2.742	6.810
Ítem 18	1.738	.920	-.423	-.598
Ítem 20	.369	.795	2.120	3.368
Ítem 22	1.063	.991	.674	-.557
Ítem 24	1.427	1.133	.034	-1.400
Ítem 25	.235	.658	2.927	7.762
Ítem 26	.700	.943	1.089	-.022
Ítem 27	.691	.801	1.133	.924
Ítem 29	1.070	1.132	.531	-1.188
Ítem 30	.264	.667	2.736	6.942

Nota. M: media; DE: desviación estándar g_1 : asimetría; g_2 : curtosis $p < .05$.

Los índices de ajuste absolutos e incrementales se presentan en la Tabla 2, en tanto los resultados del análisis factorial aplicando AFC y ESEM se informan en la Tabla 3.

En AFC se identificaron índices de ajuste óptimos (CFI = .966 y TLI = .960; RMSEA [90%] = .044[.038-.051]), ubicándose levemente por encima del valor establecido como criterio, el índice WRMR = 1.092 > 1.00 . Las cargas factoriales fueron significativas ($p \leq .05$), y en su mayoría aceptables, con valores que oscilaron entre

Tabla 2
Índices de ajuste modelo tetrafactorial ESEM y AFC.

	χ^2	<i>gl</i>	CFI	TLI	WRMR	RMSEA [IC]
ESEM	5733.37***	171	.979	.965	.734	.042 [.034-.050]
AFC	5733.37***	171	.966	.960	1.092	.044 [.038-.051]

Nota. *** $p < .001$

.862 y .925 en la dimensión *ideaciones suicidas*, de .408 a .745 en *soledad y aislamiento social*, de .457 a .706 en *desesperanza*, y de .340 a .784 en la dimensión *baja autoeficacia*. El índice de confiabilidad ω fue de .939 en la dimensión *ideaciones suicidas*, de .620 en *soledad y aislamiento social*, de .696 en *desesperanza*, y .655 en *baja autoeficacia* (Tabla 3).

Por otra parte, la salida del AFC informó índices de modificación significativos ($\chi^2 > 10$), que implican a 4 de los 5 elementos que agrupa la dimensión *soledad y aislamiento social* (ítems 4, 14, 26 y 29), y un reactivo de la dimensión *baja autoeficacia* (ítem 1), y señalan potenciales cargas cruzadas de los ítems 4, 14 y 29 en la dimensión *desesperanza*, del ítem 26 en las dimensiones *baja autoeficacia* e *ideaciones suicidas* y del ítem 1 con esta última dimensión (Dominguez-Lara & Merino-Soto, 2018).

Respecto del modelo ESEM, se obtuvieron índices de ajuste óptimos (CFI = .979; TLI = .965; RMSEA [90%] = .042 [.034, .050]; WRMR = .734); las cargas factoriales fueron significativas ($p \leq .05$) y en su mayoría aceptables, con magnitudes que oscilaron entre .749 y .992 en la dimensión *ideaciones suicidas*, de .322 a .643 en la dimensión *soledad y aislamiento social*, de .432 a .558 en *desesperanza* y entre .480 y .556 en *baja autoeficacia*. En cuanto a la confiabilidad, la dimensión *ideaciones suicidas* presentó una magnitud de .922 en el coeficiente ω seleccionado, y en las dimensiones restantes, *soledad*

y *aislamiento social, desesperanza y baja autoeficacia* se obtuvieron magnitudes de .588, .619 y .622 respectivamente.

La correlación interfactorial promedio en AFC ($\phi_{AFC} = .595$) fue mayor respecto de la obtenida en ESEM ($\phi_{ESEM} = .418$).

El análisis de simplicidad factorial de las filas (ISF; Kaiser, 1974), a partir del valor de las cargas estandarizadas de los ítems que informa el modelo ESEM, indicó que 15 ítems presentan $ISF > .80$ y otros 2, $ISF > .70$, señalando en la dimensión *soledad y aislamiento social* dos reactivos con índices por debajo del valor establecido como criterio para ser considerado aceptable (Dominguez-Lara & Merino-Soto, 2018): el ítem 4 ($ISF_4 = .576$; *Confío en que los que se preocupan por mí no me van a abandonar si tengo un fracaso*) y el ítem 14 ($ISF_4 = .556$; *Las personas que son importantes para mí comprenden mis sentimientos*; Fleming, 2003; Fleming & Merino-Soto, 2005).

Discusión

El suicidio es considerado un problema grave de salud pública a consecuencia del incremento de la tasa de mortalidad relacionada con esta conducta. Según informa la DEIS (2016), en nuestro país la tasa se incrementó trienalmente desde 1990. Dado el carácter relevante que adquieren los instrumentos de evaluación psicológica en el

Tabla 3

Estructura interna del Inventario de Orientación Suicida ISO 19 (AFC y ESEM).

		AFC	ESEM				
			F ₁	F ₂	F ₃	F ₄	ISF
<i>F₁: Ideaciones suicidas</i>							
	Ítem 15	.862	.861	.020	.052	-.051	.990
	Ítem 20	.880	.840	.044	.048	-.014	.992
	Ítem 25	.925	.992	-.092	-.034	.041	.985
	Ítem 30	.893	.749	.097	.064	.084	.953
<i>F₂: Soledad y aislamiento social</i>							
	Ítem 4	.418	-.103	.405	.275	-.078	.576
	Ítem 5	.408	.001	.322	.138	.000	.806
	Ítem 14	.446	-.07	.327	.300	-.048	.556
	Ítem 26	.745	.145	.643	-.142	.124	.840
	Ítem 29	.442	-.034	.635	-.234	.082	.827
<i>F₃: Desesperanza</i>							
	Ítem 2	.574	.123	.165	.464	-.106	.736
	Ítem 7	.457	-.089	.018	.470	.156	.832
	Ítem 11	.706	.126	-.042	.558	.172	.827
	Ítem 22	.460	.087	-.100	.432	.113	.814
	Ítem 27	.591	.057	.124	.550	-.051	.913
<i>F₄: Baja autoeficacia</i>							
	Ítem 1	.340	-.157	.006	.062	.504	.869
	Ítem 8	.360	-.024	-.040	-.007	.480	.987
	Ítem 12	.610	.045	.102	.004	.556	.949
	Ítem 18	.494	.045	-.014	.047	.487	.976
	Ítem 24	.784	.173	.129	.104	.460	.718
			F ₁	.597	.574	.648	
			F ₂	.508	.588	.648	
			F ₃	.386	.393	.517	
			F ₄	.519	.464	.237	
ESEM	ω		.922	.588	.619	.622	
AFC	Ω		.939	.620	.696	.655	

Nota. Sobre la diagonal: correlación interfactorial-método AFC; bajo la diagonal: correlación interfactorial-método ESEM; ω: coeficiente omega. $p < .05$.

contexto de la prevención y la detección temprana del sujeto en riesgo suicida, esta investigación tiene por objetivo el estudio de algunas propiedades psicométricas de la versión reducida de 19 ítems de la Escala ISO 30 (Galarza et al., 2019).

El Inventario de Orientación Suicida ISO-30 originalmente se organiza en 5 dimensiones: *baja autoestima, desesperanza, incapacidad para afrontar emociones, soledad y aislamiento social e ideación suicida*. Fernández-Liporace y Casullo (2006) realizan un análisis de las propiedades psicométricas del instrumento, según el cual se eliminan 4 reactivos, y los restantes ítems se reagrupan en 5 componentes que no replican exactamente los agrupamientos del diseño propuesto por King y Kowalchuk (1994). Sin embargo, en Argentina se sigue aplicando la adaptación realizada por Casullo en 1997 o se informa, en algunos casos, que se consideró la solución identificada por Fernández-Liporace y Casullo (2006), pero sosteniendo los puntos de corte definidos para el diseño original de 30 ítems de King y Kowalchuk (1994) adaptado por Casullo (1997).

Galarza et al. (2019) realizan un estudio psicométrico identificando una solución tetrafactorial de 19 ítems que presentan diferencias con la solución factorial identificada por Fernández-Liporace y Casullo (2006), y la adaptación realizada por Casullo (1997).

En vista de estas consideraciones, se procedió a realizar los análisis factoriales restrictivos, aplicando AFC y ESEM, con el propósito de evaluar el ajuste de los datos empíricos al modelo tetrafactorial propuesto por Galarza et al. (2019), en una muestra de 656 adolescentes escolarizados de tres instituciones educativas de la provincia de Córdoba.

Como se observó, el ESEM permitió una aproximación rigurosa a la estructura interna hipotetizada; se alcanzaron índices de ajuste óptimos (Hu & Bentler, 1999). Los ítems en gene-

ral estarían representando aceptablemente a las dimensiones que les asignó el diseño. Los parámetros identificados (índices de ajuste, valor de las cargas factoriales, evidencia de diferenciación empírica entre los factores), estarían apoyando en general el diseño propuesto por Galarza et al. (2019).

En cuanto a la cuantificación de la complejidad factorial de las variables observables, estimada a partir de las cargas estandarizadas de los ítems que informa el modelo ESEM, los resultados recolectaron evidencia de su relativa independencia y simplicidad, y señalaron un ajuste empírico adecuado a la estructura factorial que se propone, con excepción de los ítems 4 y 14 que presentan índices de simplicidad factorial por debajo del criterio especificado ($ISF > .70$; Dominguez-Lara & Merino-Soto, 2018). Estos elementos aportan evidencias aceptables de que los mismos estarían siendo influidos simultáneamente por más de un factor, lo cual afecta su interpretación e incide en el grado de simplicidad factorial que les supone el diseño teórico (Fleming, 2003; Fleming & Merino-Soto, 2005; Dominguez-Lara & Merino-Soto, 2018).

Respecto del AFC, los índices para evaluar el ajuste del modelo fueron óptimos y levemente inferiores a los obtenidos mediante el modelo ESEM, mientras que el índice WRMR indicó un valor de 1.092, ligeramente superior al valor establecido como criterio. Los ítems estarían representando aceptablemente, en general, a sus respectivas dimensiones. Las magnitudes de las correlaciones interfactoriales se incrementaron y eso incide en los resultados y conclusiones que se derivan. Este incremento ha sido investigado, y en su explicación se contempla el efecto de las cargas factoriales secundarias que informa el modelo exploratorio de ecuaciones estructurales (ESEM), y que en AFC se fijaron a 0 (Asparouhov & Muthén, 2009; Marsh et al., 2014).

Los índices de modificación que informa la salida del AFC implican a cuatro de los cinco elementos de la dimensión *soledad y aislamiento social* (ítems 4, 14, 26 y 29) y al ítem 1 del factor *baja autoeficacia*, en función de lo cual se informan potenciales cargas cruzadas. Además, los ítems 4 y 14 también fueron señalados en la cuantificación de la simplicidad factorial de las filas por presentar índices ISF por debajo del criterio establecido.

La valoración individual de las cargas factoriales es relevante y, al mismo tiempo, objeto de disensos; se inscribe en el debate que se plantea al momento de definir los puntos de corte que se sugieren para retener un ítem y el problema inherente del impacto sobre la fiabilidad del constructo (Dominguez-Lara, 2018a). Los índices de consistencia interna informados en la muestra considerada, que quedan por fuera del rango esperado [.70; .90], se vinculan con lo señalado, y aportan evidencia en este sentido (como se observa en la Tabla 3), de modo que sería recomendable que continúen siendo objeto de investigación (Campo-Arias & Oviedo, 2008; Ventura-León & Caycho-Rodríguez, 2017).

La modificación en el número de reactivos de la solución tetrafactorial de 19 ítems tendrá implicancias al momento de considerar los puntos de corte que propone el diseño original de King y Kowalchuk (1994), que permiten clasificar al sujeto en tres categorías de riesgo suicida: bajo, moderado o alto y, en particular, la pérdida de dos elementos en la dimensión *ideación suicida* tendrá incidencias en relación a la otra forma de determinar el nivel de riesgo, ya que el diseño originalmente previsto proponía que tres o más de estos ítems críticos respondidos con valores de dos o tres, indican alto riesgo suicida; por ende, correspondería revisar el indicador total del riesgo, y verificar la capacidad discriminativa del instrumento mediante la definición de curvas ROC,

entre adolescentes diagnosticados con y sin intento suicida, con y sin ideaciones suicidas activas.

Como expresa Galarza et al. (2019), la solución tetrafactorial de 19 ítems, con algunas diferencias, sostiene una versión reducida de la escala ISO 30 que debería complementarse con otros estudios que incorporen una muestra más heterogénea en cuanto a variables tales como edad, género, religión, región, condición clínica (a fin de contrastar adolescentes con y sin intento de suicidio previo), propiciando análisis que determinen el nivel de sensibilidad y especificidad del instrumento para establecer puntos de corte; y la recolección de evidencia de validez convergente y discriminante adoptando un criterio externo.

Siguiendo la misma línea de análisis, la eliminación de reactivos y los nuevos agrupamientos identificados en los trabajos presentados por Fernández-Liporace y Casullo (2006) y Galarza et al. (2019), deben ser contemplados cuando se analiza el comportamiento de la distribución de las puntuaciones en una escala y de los factores subyacentes en las muestras consideradas, porque hacen a la fiabilidad y estructura interna de los instrumentos, además de incidir en las conclusiones que de ellos se derivan.

Dada la relevancia del tema, y ubicando el acento en la prevención y detección temprana, se evidencia la importancia que adquiere el análisis de los instrumentos que se utilizan para evaluar el riesgo suicida en adolescentes con el objetivo de mejorar la identificación de sujetos en riesgo suicida y permitirle al profesional psicólogo contar con herramientas más fiables para trabajar con esta problemática.

Por lo tanto, es necesario avanzar en el conocimiento de la problemática del suicidio mediante nuevas investigaciones empíricas que contribuyan tanto en el accionar de políticas en salud pública, como en los tratamientos psicológicos (Flores-Kanter, 2017).

Referencias

- Abarca, C., Gheza, C., Coda, C., & Elicer, B. (2018). Revisión de la literatura para identificar escalas estandarizadas de evaluación del riesgo suicida en adultos atendidos en atención primaria de salud. *Medwave*, 18(05), e7246-e7246. doi: [10.5867/medwave.2018.05.7246](https://doi.org/10.5867/medwave.2018.05.7246)
- Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438. doi: [10.1080/10705510903008204](https://doi.org/10.1080/10705510903008204)
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research, and Evaluation*, 19(5). Recuperado de <https://scholarworks.umass.edu/pare>
- Bologna, E. (2011). *Estadística para psicología y educación*. Córdoba: Brujas.
- Campo-Arias, A., & Oviedo, H. C. (2008). Propiedades psicométricas de una escala: La consistencia interna. *Revista de Salud Pública*, 10(5), 831-839. Recuperado de <http://www.scielo.unal.edu.co>
- Casullo, M. M. (1997). Riesgo suicida en estudiantes universitarios. *Epidemiología y prevención. Desarrollos en Psiquiatría Argentina*, 2(4), 33-38.
- Coolican, H. (2005). *Métodos de investigación y estadística en psicología*. México D. F.: Manual Moderno.
- Dirección de Estadísticas e Información en Salud, Ministerio de Salud. (2016). Estadísticas vitales, Información básica, Argentina año 2016. Recuperado de <http://www.deis.msal.gov.ar/index.php/tabulados-2>
- Dominguez-Lara, S. (2018a). Propuesta de puntos de corte para cargas factoriales: Una perspectiva de fiabilidad de constructo. *Enfermería Clínica*, 28(6), 401-402. doi: [10.1016/j.enfcli.2018.06.002](https://doi.org/10.1016/j.enfcli.2018.06.002)
- Dominguez-Lara, S. (2018b). Fiabilidad y alfa ordinal. *Actas Urológicas Españolas*, 42(2), 140-141. doi: [10.1016/j.acuro.2017.07.002](https://doi.org/10.1016/j.acuro.2017.07.002)
- Dominguez-Lara, S., & Navarro-Loli, J. S. (2018). Revisión de metodologías empleadas en los estudios factoriales de la Escala de Bienestar Psicológico de Ryff (versión en español). *Revista Evaluar*, 18(2). doi: [10.35670/1667-4545.v18.n2.20800](https://doi.org/10.35670/1667-4545.v18.n2.20800)
- Dominguez-Lara, S., & Merino-Soto, C. (2018). Estructura interna del BFI-10P y BFI-15P: Un estudio complementario con enfoque CFA y ESEM. *Revista Argentina de Ciencias del Comportamiento*, 10(3), 22-34. Recuperado de <https://revistas.unc.edu.ar/index.php/racc>
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33. Recuperado de <http://www.papelesdelpsicologo.es>
- Fernández-Liporace, M., & Casullo, M. M. (2006). Validación factorial de una escala para evaluar riesgo suicida. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 1(21), 9-22. Recuperado de <https://www.redalyc.org>
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: Algunas consideraciones adicionales. *Anales de Psicología*, 30(3). doi: [10.6018/analesps.30.3.199991](https://doi.org/10.6018/analesps.30.3.199991)
- Fleming, J. S. (2003). Computing measures of simplicity of fit for loadings in factor-analytically derived scales. *Behavior Research Methods, Instruments, & Computers*, 35(4), 520-524. doi: [10.3758/bf03195531](https://doi.org/10.3758/bf03195531)
- Fleming, J. S., & Merino-Soto, C. (2005). Medidas de simplicidad y de ajuste factorial: Un enfoque para la evaluación de escalas construidas factorialmente. *Revista de Psicología*, 23(2), 250-266. doi: [10.18800/psico.200502.002](https://doi.org/10.18800/psico.200502.002)
- Flores-Kanter, P. E. (2017). El lugar de la psicología en las investigaciones empíricas del suicidio en Argentina: Un estudio bibliométrico. *Interdisciplinaria: Revista de Psicología y Ciencias Afines*, 34(1). doi: [10.16888/interd.2017.34.1.2](https://doi.org/10.16888/interd.2017.34.1.2)
- Flores-Kanter, P. E., García-Batista, Z. E., Moretti, L. S., & Medrano, L. A. (2019). Towards an explanatory model of suicidal ideation: The effects of cognitive emotional regulation strategies, affectivity and hope-

- lessness. *The Spanish Journal of Psychology*, 22. doi: [10.1017/sjp.2019.45](https://doi.org/10.1017/sjp.2019.45)
- Freiberg-Hoffmann, A., Stover, J. B., de la Iglesia, G., & Fernández-Liporace, M. (2013). Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios. *Ciencias Psicológicas*, 7(2), 151-164. Recuperado de <https://revistas.ucu.edu.uy>
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3). doi: [10.7275/n560-j767](https://doi.org/10.7275/n560-j767)
- Galarza, A., Fernández-Liporace, M., Castañeiras, C., & Freiberg-Hoffmann, A. (2019). Análisis psicométrico del Inventario de Orientaciones Suicidas ISO-30 en adolescentes escolarizados (14-18 años) y jóvenes universitarios marplatenses. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 51(2). doi: [10.21865/ridep51.2.10](https://doi.org/10.21865/ridep51.2.10)
- Goldston, D. (2000). Assessment of suicidal behaviors and risk among children and adolescents. Technical report submitted to NIMH under Contract No. 263-MD-909995. Recuperado de <https://vitaalere.com.br/download/assessment-of-suicidal-behaviors-and-risk-among-children-and-adolescents.pdf>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- IBM Corp. (2015). IBM SPSS Statistics for Windows (Version 23.0). [Software de cómputo]. Armonk, NY: IBM Corp.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36. doi: [10.1007/bf02291575](https://doi.org/10.1007/bf02291575)
- King, J. D., & Kowalchuk, B. (1994). *Manual for ISO-30 Adolescent: Inventory of Suicide Orientation-30*. Minneapolis: National Computer Systems.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2ª ed.). New York: Guilford.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Annals of Psychology*, 30(3), 1151-1169. doi: [10.6018/analesps.30.3.199361](https://doi.org/10.6018/analesps.30.3.199361)
- Lorenzo-Seva, U., & Ferrando, P. (2018). FACTOR (Versión 10.8.04). [Software de cómputo]. Recuperado de <http://psico.fcep.urv.es/utilitats/factor/index.html>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471-491. doi: [10.1037/a0019227](https://doi.org/10.1037/a0019227)
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85-110. doi: [10.1146/annurev-clinpsy-032813-153700](https://doi.org/10.1146/annurev-clinpsy-032813-153700)
- Martínez, C. (2007). *Introducción a la Suicidología*. Buenos Aires: Lugar.
- Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide. Sixth edition*. Los Angeles, CA: Muthén & Muthén.
- Organización Mundial de la Salud. (2014). Prevención del suicidio: Un imperativo global. Recuperado de <https://apps.who.int/iris>
- Osman, A., Gutierrez, P. M., Barrios, F. X., Bagge, C. L., Kopper, B. A., & Linden, S. (2005). The Inventory of Suicide Orientation-30: Further validation with adolescent psychiatric inpatients. *Journal of Clinical Psychology*, 61(4), 481-497. doi: [10.1002/jclp.20086](https://doi.org/10.1002/jclp.20086)
- Paniagua-Suárez, R. E., González-Posada, C. M., & Rueda-Ramírez, S. M. (2016). Validación de la versión en español del Inventario de Orientación del Suicidio - ISO 30 en estudiantes adolescentes de instituciones educativas en Medellín - Colombia. *World Journal of Education*, 6(4). doi: [10.5430/wje.v6n4p22](https://doi.org/10.5430/wje.v6n4p22)
- Pérez, E., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista*

Argentina de Ciencias del Comportamiento, 2(1), 58-66. Recuperado de <https://revistas.unc.edu.ar>

Rangel-Garzón, C. X., Suárez-Beltrán, M. F., & Escobar-Córdoba, F. (2015). Escalas de evaluación de riesgo suicida en atención primaria. *Revista de la Facultad de Medicina*, 63(4), 707-716. doi: [10.15446/revfac-med.v63.n4.50849](https://doi.org/10.15446/revfac-med.v63.n4.50849)

Runeson, B., Odeberg, J., Pettersson, A., Edbom, T., Jildevik-Adamsson, I., & Waern, M. (2017). Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. *PLOS ONE*, 12(7), e0180292. doi: [10.1371/journal.pone.0180292](https://doi.org/10.1371/journal.pone.0180292)

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6^a ed). New York: Allyn and Bacon.

Trógolo, M. A., Tosi, J. D., Poó, F. M., Ledesma, R. D., Medrano, L. A., & Dominguez-Lara, S. (2020). Factor structure and measurement invariance of the Multidimensional Driving Style Inventory across gender and age: An ESEM approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, 71, 23-30. doi: [10.1016/j.trf.2020.04.001](https://doi.org/10.1016/j.trf.2020.04.001)

Ventura-León, J. L., & Caycho-Rodríguez, T. (2017). El coeficiente omega: Un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627. Recuperado de <https://www.redalyc.org/pdf/773/77349627039.pdf>

Relación entre medidas neuropsicológicas de ejecución y autoinforme de las funciones ejecutivas

Relationship between Execution Neuropsychological Measures and Self-report of Executive Functions

Kelvin L. Morales-Millán *¹, Yashira Arroyo-Pérez¹, Maribella González-Viruet¹,
Israel Sánchez-Cardona²

¹ - Universidad Carlos Albizu, San Juan, Puerto Rico.

² - Kennesaw State University, Georgia, USA.

Introducción
Metodología
Resultados
Discusión
Referencias

Recibido: 15/09/2020 Revisado: 12/10/2020 Aceptado: 15/11/2020

Resumen

En este estudio se evaluó la relación entre las puntuaciones de funciones ejecutivas (FE) al utilizar medidas de autoinforme versus medidas neuropsicológicas de ejecución. Se examinaron diversos predictores que han sido asociados con las FE para evaluar la convergencia entre las puntuaciones de estas medidas. La muestra consistió en 60 adolescentes varones de entre 14 y 18 años con y sin antecedentes legales. Se utilizaron las puntuaciones del Comprehensive Executive Function Inventory y del Delis-Kaplan Executive Function System. Los resultados de los análisis de correlación indicaron una relación muy baja y en su mayoría no significativa entre las puntuaciones de ambas medidas. Los resultados del análisis de regresión múltiple indicaron que los predictores de las FE se relacionan de manera disímil con las medidas. No obstante, ambas formas de medir las FE no son excluyentes y pueden ser utilizadas para medir diferentes aspectos de las FE.

Palabras clave: funciones ejecutivas, planificación y organización, flexibilidad cognitiva, control inhibitorio, medidas de autoinforme, medidas neuropsicológicas de ejecución

Abstract

This study evaluated the relationship between executive function (EF) scores using self-report measures versus performance based neuropsychological measures. Various predictors that have been associated with EFs were examined to assess the convergence of scores for these measures. The sample consisted of 60 male adolescents between 14 and 18 years old with and without legal history. The scores from the Comprehensive Executive Function Inventory and Delis-Kaplan Executive Function System were used. The correlation analysis results indicated a small and mostly non-significant relationship between the scores of both measures. The multiple regression analysis results indicated that the predictors of EF are dissimilarly related to the measures. However, both ways of measuring EF are not exclusive, and it can be useful to use them to measure different EF aspects.

Keywords: executive functions, planning and organization, cognitive flexibility, inhibitory control, self-report measures, neuropsychological measures

*Correspondencia a: Kelvin Morales. E-mail: kmorales690@sju.albizu.edu

Introducción

Tradicionalmente la evaluación de las *funciones ejecutivas* (FE) se realiza a través de pruebas neuropsicológicas. Sin embargo, actualmente se han desarrollado instrumentos de autoinforme que exploran el funcionamiento ejecutivo conductual de las personas en la vida cotidiana. Ante las diferentes medidas de evaluación de las FE, se han originado debates sobre cuál de los métodos de evaluación es mejor para la detección de deterioro en el funcionamiento ejecutivo de una persona. Esto representa un problema ya que no existe un consenso acerca de la capacidad de estas medidas para medir con precisión los dominios de las FE. Maldonado-Belmonte (2016) propone que la evaluación de las funciones ejecutivas en la población de niños y adolescentes es compleja por el hecho de que los dominios de las funciones ejecutivas se desarrollan en diferentes momentos y a distinto ritmo.

Este estudio pretende evaluar si existe alguna relación entre las puntuaciones obtenidas en las pruebas de desempeño o ejecución y el autoinforme que miden las funciones ejecutivas (planificación, flexibilidad cognitiva y organización) de un grupo de jóvenes varones puertorriqueños intervenidos por el Tribunal de Menores de Puerto Rico y un grupo de jóvenes varones puertorriqueños sin antecedentes legales.

Funciones ejecutivas

A pesar de no haber un consenso en la definición, se ha planteado que el término *funciones ejecutivas* (FE) hace referencia a un conjunto de habilidades cognitivas superiores y acciones autodirigidas que involucran áreas como la atención selectiva, la inhibición, la flexibilidad cognitiva, la toma de decisiones, la organización y la regula-

ción emocional, entre otras (Barkley, 2012).

Para propósitos de este estudio se utilizará el modelo de resolución de problemas desarrollado por Zelazo, Carter, Reznick y Frye (1997). En dicho modelo, se parte de que las FE son un macroconstructo que se extiende a cuatro fases principales de solución de problemas. La primera fase es la *presentación del problema*, en la que se debe representar el problema de tal manera que sea posible resolverlo. Problemas en esta fase podrían implicar dificultades en habilidades atencionales y en el control inhibitorio. La segunda fase es la *planificación*, en la que el individuo evalúa su contexto al igual que las opciones que tiene disponibles o que cree lograr para solucionar el problema. La tercera fase es la *ejecución* e implica tener un plan y lograr traducir el mismo en acción. Por último, se encuentra la fase de *evaluación*, donde se espera que se evalúe si la solución ejecutada para el problema fue efectiva. Las variables evaluadas en este estudio (planificación, organización, flexibilidad y control inhibitorio) están involucradas en cada una de las cuatro fases para la solución de un problema.

Planificación y organización. La *planificación* se define como el proceso de identificar posibles opciones, pasos y resultados necesarios para lograr un objetivo específico (Patros, Tarle, Alderson, Lea, & Arrington, 2019). Por otro lado, el dominio de *organización* se define como la habilidad de organizar y lograr reconocer la información requerida en un proceso de aprendizaje (Roberts & Apaza, 2011).

Algunas de las pruebas que se utilizan para medir la planificación y organización son: Tower of London (Shallice, 1982; Krikorian, Bartok, & Gay, 1994), D-KEFS-prueba de clasificación (Delis, Kaplan, & Kramer, 2001), Behavioral Rating Inventory of Executive Functioning (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000), Com-

prehensive Executive Function Inventory (CEFI; Naglieri & Golstein, 2013), entre otras.

Flexibilidad cognitiva. La *flexibilidad cognitiva* se define como la habilidad del individuo de cambiar su curso de acción para adaptarse efectivamente a las demandas del ambiente, de una tarea o situación (Rueda, Cómbita, & Pozuelos, 2016). Algunas pruebas que pueden ser utilizadas para evaluar la flexibilidad cognitiva son: D-KE-FS-prueba del trazado (Delis et al., 2001), D-KE-FS-prueba de interferencia de color - palabra (Delis et al., 2001), Trail Making Test (Reitan, 1958), Wisconsin Card Sorting Test (Berg, 1948), Test de 5 dígitos (Sedó, 2004), BRIEF (Gioia et al., 2000), CEFI (Naglieri & Goldstein, 2013), entre otros.

Control inhibitorio. El *control inhibitorio* es la función que se encarga de suprimir una tendencia automática en una situación dada (Thierry, Glowinski, Goldman-Rakic, & Christen, 1994). La capacidad de inhibición es necesaria para lograr la negociación de objetivos que compiten entre sí y el retraso de la gratificación inmediata para culminar con los objetivos elegidos a largo plazo (Carlson, Zelazo, & Faja, 2013). Algunos instrumentos para medir la capacidad de inhibición son: Stroop (Stroop, 1935), Go/No Go (Donders, 1969), Stop signal (Logan & Cowan, 1984), BRIEF (Gioia et al., 2000), CEFI (Naglieri & Goldstein, 2013), entre otros.

Problema con la evaluación de las FE

La falta de unanimidad en la definición de las FE trae consigo problemas para su evaluación. Esto representa una dificultad para comparar las puntuaciones o ejecución del individuo en distintas pruebas, dado que cada una se basa en

distintos modelos de las FE. Esta dificultad se complejiza aún más dadas las distintas medidas de evaluación utilizadas para las FE, lo que tiene como consecuencia que los hallazgos de las investigaciones sean inconsistentes entre sí.

En la evaluación neuropsicológica de las funciones ejecutivas existen dos estrategias: a) la aplicación de pruebas de desempeño o ejecución, como son las tareas experimentales creadas para evaluar una función ejecutiva específica en un contexto controlado, y b) las pruebas de autoinforme, que consisten en la perspectiva y observación de las funciones ejecutivas conductuales en las actividades de la vida diaria.

Las medidas de ejecución se utilizan para evaluar el funcionamiento cognitivo (Harvey, 2012). Este ha sido el método utilizado típicamente para evaluar las FE (Comesaña, Stelzer, & Introzzi, 2017; Vélez-Pastrana et al., 2016). Su administración está estandarizada, por lo que los estímulos se presentan de la misma forma y orden a todos los examinados. La ventaja de una prueba estandarizada es que disminuye la posibilidad de errores durante la administración y corrección, facilita la comparación del evaluado con otros grupos y expone a los evaluados a las mismas condiciones y estímulos. La puntuación está compuesta por el tiempo de las respuestas del individuo y cuán correctas o precisas son las mismas. El problema con este tipo de evaluación es que carece de validez ecológica (Bakar, Taner, Soy-sal, Karakas, & Turgay, 2011). La validez ecológica en neuropsicología se refiere a la capacidad del instrumento de predecir comportamientos o problemas en el funcionamiento diario a partir del desempeño en la prueba (Spooner & Pachana, 2006). Esto sugiere que la evaluación con pruebas de desempeño no refleja la capacidad del paciente para hacer uso de sus habilidades para la solución de problemas reales.

Cabe señalar que muchas de las pruebas

neuropsicológicas que se utilizan actualmente para medir FE presentan niveles moderados de validez ecológica ya que se utilizan para conocer o predecir la funcionalidad del individuo (García-Molina, Tirapu-Ustárroz, & Roig-Rovira, 2007). Además, este tipo de instrumentos analizan solamente algunos aspectos de la complejidad de las funciones ejecutivas, obviando otros como lo son los procesos metacognitivos en el contexto del diario vivir o cognición social (Suchy, 2015).

Por otro lado, la administración del autoinforme surge por la necesidad de aumentar la validez ecológica y de valorar el funcionamiento ejecutivo del sujeto evaluando su comportamiento en la vida diaria (Pérez-Salas, Ramos, Oliva, & Ortega, 2016). En ellas se pretende evaluar la percepción del individuo respecto a destrezas del diario vivir relacionadas con las FE, como lo son el manejo de emociones, el poder planificar y organizar las tareas a realizar durante el día, entre otras. Este tipo de medida permite predecir algún deterioro en las actividades cotidianas, así como el funcionamiento ocupacional (Barkley & Fischer, 2011). Algunas de sus ventajas son su fácil administración e interpretación y la riqueza de la información recopilada al provenir de primeras manos (Mitchell & Miller, 2008).

Sin embargo, Paulhus y Vazire (2007) enumeran las siguientes desventajas de los autoinformes: a) la precisión de la información provista se ve afectada por la deseabilidad social y/o respuestas extremas, b) los resultados dependen del conocimiento del individuo y c) la medición puede ser afectada por factores culturales. Hoskin (2012) agregó que dichas medidas tienen una gran dependencia de la honestidad del individuo y pueden ser afectadas por la falta de destrezas de introspección, de entendimiento de la prueba y por el uso de diferentes interpretaciones de las puntuaciones de la escala de valoración. Además, en personas con deterioro cognitivo los autoinfor-

mes tienden a perder su utilidad.

Se han realizado varias investigaciones en las cuales se evalúa a las FE en distintos contextos utilizando una o ambas medidas y se ha encontrado que las puntuaciones son inconsistentes entre sí. Por ejemplo, en un estudio realizado por Mitchell y Miller (2008) se buscó evaluar la relación entre las medidas de las funciones ejecutivas y las habilidades funcionales tanto observadas como autoinformadas. Para evaluar las FE, los autores administraron el Delis-Kaplan Executive Function System (D-KEFS), específicamente las pruebas: Trail Making Test 4, Verbal Fluency Test y el Design Fluency Test. En los resultados de dicho estudio se observó que las pruebas mencionadas son más sensibles para detectar deterioro en las funciones ejecutivas que las pruebas de autoinforme de las habilidades de funcionamiento. Esto coincide con lo que han mencionado varios autores sobre la inconsistencia de los resultados para establecer relación entre las medidas de autoinforme y las neuropsicológicas para la evaluación de las FE (Loewenstein et al., 2001; Zanetti, Frisoni, Rozzini, Bianchetti, & Trabucchi, 1998).

Toplak, West y Stanovich (2013) concluyeron en su estudio que ambas medidas miden distintos aspectos de las FE. Explican que son diferentes en sus procesos de administración, las exigencias al evaluado y las puntuaciones por dominios. Las pruebas de ejecución implican un nivel de estructura más alto de parte del examinador a diferencia de las pruebas de autoinforme, que no requieren de dicho factor. De igual manera, los autores encontraron que existe una asociación de baja a moderada entre los dominios de ambas medidas. Por último, mencionan que, desde una perspectiva teórica de la ciencia cognitiva, ambos constructos evalúan distintos niveles de las FE: las medidas de ejecución miden la eficiencia del procesamiento, mientras que las de autoinforme miden la búsqueda de objetivos personales.

Estas discrepancias en los artículos mencionados pueden deberse a diferentes factores. Uno de ellos puede ser la diferencia entre las teorías en que se basa cada prueba, es decir, el modelo teórico. Otro, es el hecho de que las pruebas de autoinforme pueden verse influenciadas por distintos factores (como la discapacidad social, factores culturales, conocimiento del evaluado, entre otros). Todo esto hace relucir la necesidad de seguir evaluando, por un lado, si existe alguna relación entre ambas medidas, y por otro, la capacidad de evaluación de cada una con respecto a las funciones ejecutivas.

Varios autores recomiendan la integración de las pruebas de ejecución o desempeño con autoinformes del funcionamiento ejecutivo para obtener una imagen más completa de las habilidades ejecutivas del individuo (Nęcka, Lech, Sobczyk, & Śmieja, 2012; Nordvall, Jonsson, & Stigsdotter-Neely, 2017). Además, la integración de las medidas permite evaluar la asociación entre las dos metodologías de evaluación (Nęcka et al., 2012; Nordvall et al., 2017). Una mejor comprensión de la relación entre ambos instrumentos puede favorecer el desarrollo de procedimientos más efectivos para la evaluación de las funciones ejecutivas. Sin embargo, hasta el momento de revisión solo se encontró un estudio que mide la relación entre las pruebas de ejecución y las medidas autoinforme en adolescentes con problemas legales (Nordvall et al., 2017). No obstante, en Puerto Rico no existen investigaciones de este tipo en jóvenes transgresores. En los estudios antes mencionados los autores se centran en la comparación de las puntuaciones de cada prueba sin tomar en consideración otros de sus factores. Uno de estos factores se relaciona con evaluar si estas discrepancias existen en la relación de cada prueba con los predictores de un buen funcionamiento ejecutivo, tomando en cuenta que ambas medidas son construidas y miden aspectos distintos den-

tro de las FE. Es por esto que en este estudio se pretende examinar utilizando ambas medidas y tomando en cuenta los predictores sociodemográficos de las FE. Se considera que, si las puntuaciones de las FE medidas utilizando autoinformes y pruebas de ejecución se correlacionan entre sí, esto significaría una concordancia sobre el constructo que miden.

Funcionamiento ejecutivo y conductas transgresoras/antisociales

La adolescencia, que comprende el periodo entre la infancia y la edad adulta, está asociada a cambios biológicos, psicológicos, cognitivos y sociales (Blakemore, Burnett, & Dahl, 2010; Sowell et al., 1999). Durante la adolescencia continúan desarrollándose las regiones de la corteza prefrontal, como la corteza prefrontal dorsomedial y la prefrontal ventral, a causa de la poda sináptica y la mielinización (Nelson, Leibenluft, McClure, & Pine, 2005). El procesamiento neuronal inmaduro en la corteza frontal y otras regiones corticales y subcorticales conduce a conductas de alto riesgo, la gratificación inmediata y la desregulación emocional (Steinberg, 2010). Mundialmente la transgresión juvenil ha sido un gran desafío, y se ha convertido en un problema mundial de salud pública. La literatura señala que se han intensificado e incrementado las conductas consideradas como faltas por parte de los jóvenes en Puerto Rico y a nivel mundial, comúnmente en varones (Dorna-Pesquera & Caballero-González, 2010).

Las FE son las capacidades necesarias para la autorregulación, incluida la regulación emocional y social. Cada uno de estos dominios, principalmente la habilidad de tomar decisiones tiene un gran impacto en la conducta juvenil. Los déficits en estas capacidades dan lugar a una infle-

xibilidad cognitiva que puede resultar en comportamientos inapropiados, incapacidad para planificar y resolver problemas, incapacidad de ver el punto de vista de otra persona, distracción, agresividad, comportamientos impulsivos y juicio social deficiente. Estas alteraciones de las funciones ejecutivas se asemejan a ciertas características de las conductas antisociales y transgresoras. Esto quiere decir que los procesos de las funciones ejecutivas son importantes en el desarrollo de las conductas antisociales y transgresoras (Fuster, 2000; Mesulam, 2002).

Raine (2002) argumentó que existe una teoría de la disfunción prefrontal basada en el crecimiento tardío de la corteza prefrontal que puede ayudar a explicar la transgresión juvenil. Raine (2002) explica que en la adolescencia surge una interacción entre las demandas sociales y el desarrollo de las funciones ejecutivas. Estas demandas sociales producen una sobrecarga en el desarrollo de la corteza prefrontal que, a su vez, podría ocasionar disfunción prefrontal y una falta de control inhibitorio sobre conductas antisociales y violentas durante la adolescencia (Raine, 2002). Según Raine (2002), habrá mayor incidencia de transgresión juvenil en ambientes psicosociales menos estructurados y estables.

Predictores de funcionamiento ejecutivo

Existen varios predictores de las FE que se encargan de propiciar que dicho desarrollo se continúe dando de manera esperada para la edad del individuo (Zysset et al., 2018). Se ha encontrado que la exposición a eventos traumáticos familiares está relacionada con bajos niveles de funcionamiento ejecutivo (DePrince, Weinzierl, & Combs, 2009). Con respecto a los factores sociales, la literatura menciona que la exposición a eventos traumáticos impacta en las FE al afectar

estructuras y regiones cerebrales (Teicher & Samson, 2016). Existen diferentes tipos de eventos traumáticos, tales como el abuso físico o sexual, la violencia doméstica, la violencia en la comunidad, los desastres naturales, la muerte repentina de un familiar, entre otros. Las experiencias traumáticas pueden causar una anomalía en la organización y función del sistema neural del cerebro, lo cual puede impactar algunas capacidades tales como la memoria, el aprendizaje, el autocontrol, la concentración y las FE (Vasterling & Brailey, 2005; Op den Kelder, Van den Akker, Geurts, Lindauer, & Overbeek, 2018).

Por último, Rathinabalan y Naaraayan (2018) coinciden en que el sistema familiar es el centro primario de educación y socialización infantil y juvenil. Por tal razón, las dinámicas intrafamiliares juegan un rol esencial en el desarrollo cognitivo, psicológico, físico y social del adolescente (Lugo-Gil & Tamis-LeMonda, 2008). Algunos estudios mencionan que las relaciones afectivas seguras, los buenos tratos parentales y la estimulación desempeñan un rol fundamental en la organización, el desarrollo y el funcionamiento cerebral temprano (Schore, 2001).

Objetivos

El propósito de este estudio es evaluar la relación entre las puntuaciones de funciones ejecutivas al utilizar medidas de autoinforme versus medidas neuropsicológicas de ejecución. De igual modo, se examinan diversos predictores como la violencia comunitaria, cantidad de trauma informado y el apoyo familiar para examinar sus relaciones con las FE y la convergencia entre las puntuaciones de las medidas de ejecución y de autoinforme.

Hipótesis

H1: Las puntuaciones de FE evaluadas a través de autoinforme y las obtenidas de medidas neuropsicológicas de ejecución se relacionan significativamente.

H2a: La exposición a la violencia comunitaria se relaciona con las FE medidas a través de pruebas neuropsicológicas de ejecución.

H2b: La exposición a la violencia comunitaria se relaciona con las FE evaluadas a través de medidas de autoinforme.

H3a: La cantidad de trauma se relaciona con las FE medidas a través de pruebas neuropsicológicas de ejecución.

H3b: La cantidad de trauma se relaciona con las FE evaluadas a través de medidas de autoinforme.

H4a: El apoyo familiar se relaciona con las FE medidas a través de pruebas neuropsicológicas de ejecución.

H4b: El apoyo familiar se relaciona con las FE evaluadas a través de medidas de autoinforme.

Método

Participantes

Este estudio consiste en un análisis de datos secundarios recopilados dentro del proyecto titulado *Perfil del funcionamiento ejecutivo y aspectos psicosociales de adolescentes varones intervenidos por el Tribunal de Menores de Puerto Rico* (Arroyo-Pérez, 2019). La muestra consistió en 60 adolescentes varones de entre 14 y 18 años, reclutados en dos grupos: (a) un grupo de 30 adolescentes varones intervenidos por el Tribunal de Menores, según lo establecido por la Ley de Menores de Puerto Rico (Ley N° 88 de 9 de julio de 1986), que se encontraban en libertad condicional o en un programa de desvío y (b) un grupo de comparación compuesto por 30 menores de dife-

rentes comunidades que no poseían antecedentes legales. La investigadora principal del estudio autorizó el uso de los datos. El estudio primario fue aprobado por el Institutional Review Board (IRB) de la Universidad Carlos Albizu en San Juan, Puerto Rico. La Tabla 1 muestra las características de los participantes.

Procedimiento

La muestra del grupo de adolescentes intervenidos por el Tribunal de Menores fue reclutada en una clínica psicológica para niños y adolescentes en Puerto Rico, donde se ofrecían servicios a adolescentes que se encontraban en un programa de desvío o bajo libertad condicional. Los criterios de inclusión para este grupo fueron: 1) ser varón de entre 14 y 18 años; 2) encontrarse bajo libertad condicional o en algún programa de desvío, y 3) poseer un nivel académico de por lo menos sexto grado. La muestra del grupo de jóvenes sin antecedentes legales se reclutó en comunidades de diferentes zonas a lo largo de Puerto Rico. Los criterios de inclusión fueron: 1) ser varón y tener entre 14 y 18 años; 2) no tener antecedente legal previo, y 3) tener un nivel académico de por lo menos sexto grado. Para ambos grupos se consideraron los siguientes criterios de exclusión: 1) presentar una condición mental severa relacionada con esquizofrenia, epilepsia o disfunción intelectual, y 2) haber sufrido de algún trauma en la cabeza que haya terminado en pérdida del conocimiento.

Los jóvenes, que participaron de manera voluntaria, firmaron la *hoja de asentimiento*, luego de solicitar el consentimiento informado de su encargado o tutor legal. Posteriormente, se procedió a completar la *hoja de datos personales* y a la administración de la batería de pruebas. Las pruebas fueron administradas en la Clínica de Ni-

Tabla 1
Características sociodemográficas de la muestra.

	Tribunal	No Tribunal
<i>Participantes (n)</i>	30	30
<i>Edad</i>		
Media ± DE	17.1 ± .91	17.1 ± .91
<i>Servicios de salud mental</i>		
Psicológicos	19 (63%)	11 (36%)
Psiquiátricos	12 (40%)	
Psicológicos después del tribunal o en la actualidad	30 (100%)	3 (10%)
Psiquiátricos después del tribunal	21 (70%)	
Diagnóstico salud mental	30 (100%)	11 (36%)
<i>Diagnóstico salud mental</i>		
Relacionados a sustancias	26 (86%)	
Déficit de atención con hiperactividad	17 (56%)	5 (45%)
Específico del aprendizaje	15 (50%)	9 (81%)
Destructivos, del control de los impulsos y de la conducta no especificado	12 (40%)	
Historial autolesión	7 (23%)	
Ansiedad	7 (23%)	
<i>Aspectos académicos</i>		
Escuela superior	15 (50%)	9 (30%)
No encontrarse estudiando	17 (56%)	3 (10%)
Antecedente de fracaso escolar	25 (83%)	10 (33%)
Antecedente de abandono escolar	19 (63%)	2 (6.7%)
<i>Antecedente legal</i>		
Primera intervención por tribunal	21 (70%)	
Varias intervenciones por el tribunal	9 (30%)	
<i>Tipo de faltas</i>		
Uso de sustancias	27 (90%)	
Venta de sustancias controladas	9 (30%)	
Intimidación y amenaza	6 (20%)	
Robo con arma	5 (16%)	

Tabla 1 (Cont.)

Características sociodemográficas de la muestra.

	Tribunal	No Tribunal
<i>Tipo de trauma</i>		
Desastres naturales	30 (100%)	29 (96%)
Removidos del hogar	26 (86%)	
Pérdida de un amigo o familiar	23 (76%)	18 (63%)
Violencia comunitaria	22 (73%)	11 (36%)
Divorcio de los padres	20 (66%)	15 (50%)
Acoso escolar		11 (36%)

Nota. DE: desviación estándar.

ños y Adolescentes especializada en adolescentes con uso de sustancias y problemas legales y en las oficinas privadas de una profesional que trabajaba con adolescentes sin problemas legales. La participación tomó aproximadamente tres horas, con 15 minutos de tiempo libre para una merienda nutritiva. El ambiente de la administración de los instrumentos fue confidencial, con buena iluminación, temperatura adecuada, asientos cómodos, libre de otros estímulos y de distractores (ruidos e interrupciones) que pudieran afectar la participación.

Instrumentos

Comprehensive Executive Function Inventory (CEFI, Naglieri & Goldstein, 2013). El Inventario Comprensivo de Funciones Ejecutivas (CEFI, por sus siglas en inglés) es una prueba para evaluar las FE en individuos desde los 5 hasta los 18 años. El CEFI pretende evaluar el comportamiento del individuo en las últimas cuatro semanas. El mismo contiene formularios para ser contestados por padres, y maestros o el mismo individuo. La prueba está compuesta por 100 ítems en una escala tipo Likert de seis puntos, divididos en 9 escalas: atención, regulación emocional, organización, flexibilidad, planificación, automo-

nitoreo, iniciación, memoria de trabajo y control inhibitorio. La administración tiene una duración de aproximadamente 15 minutos. El coeficiente de confiabilidad interna de la prueba fluctuó entre .97 y .99 (Naglieri & Goldstein, 2013). Por otro lado, la confiabilidad interna entre las nueve escalas fue .77 a .86. Las escalas que se utilizarán en este estudio son las siguientes: control inhibitorio (10 ítems), flexibilidad cognitiva (7 ítems), planificación y organización (10 ítems).

Delis-Kaplan Executive Function System (Delis et al., 2001). El Delis-Kaplan Executive Function System (D-KEFS) es una prueba estandarizada compuesta por nueve subpruebas para evaluar las FE desde los 8 hasta los 89 años de edad. Las subpruebas son: Prueba del Trazado, Prueba de Fluidez Verbal, Prueba de Interferencia de Color-Palabra, Prueba de Clasificación, Prueba de las 20 Preguntas, Prueba del Contexto de la Palabra, Prueba de la Torre, Prueba de fluidez de Diseño y Prueba de Proverbio. Cada subprueba puede ser administrada como única medida de evaluación de manera individual o grupal. En este estudio se utilizaron las siguientes pruebas:

1. DKEFS- Prueba del Trazado: Esta prueba mide las funciones de flexibilidad y control inhibitorio. Se utilizaron las siguientes condiciones: secuencia de números,

secuencia de letras e intercambio de número y letra. La consistencia interna de la prueba se encuentra entre .69 y .79. (Delis et al., 2001).

2. D-KEFS-Prueba de Interferencia de Color-Palabra: Consiste en evaluar la flexibilidad cognitiva y el control inhibitorio. Se utilizó la subprueba de alternar entre mencionar el color de la tinta y leer la palabra incongruente. La confiabilidad interna entre las subtareas fue de entre .69 y .75.
3. D-KEFS-Prueba de Clasificación: evalúa la capacidad de planificación y organización. Dicha prueba consiste en dos tareas: a) clasificación libre (clasificar las cartas en dos grupos de forma espontánea en estímulos verbales y no verbales); y b) reconocimiento de categorías (organizar las cartas en dos grupos a partir de ocho posibles criterios para que luego el evaluado describa e identifique los criterios utilizados en la organización). La tarea de clasificación libre obtuvo una consistencia interna de entre .55 y .82; mientras que para la tarea de reconocimiento de categorías se encuentra entre .72 a .74.

Violencia comunitaria. La violencia comunitaria se examinó a través de la siguiente pregunta: *¿Has experimentado violencia comunitaria en algún momento de tu vida (por ejemplo, tiroteo, asalto, venta de drogas)?* Los participantes contestaron utilizando dos alternativas de respuesta (0 = *No*; 1 = *Sí*).

Cantidad de Trauma. Para examinar cuán expuestos habían estado los menores a experiencias extremadamente estresantes, se utilizó una lista de 20 eventos adversos. Algunos de los eventos adversos que se preguntaron fueron los siguientes: maltrato físico, negligencia, maltrato emocional,

abandono, acoso escolar, abuso sexual, remoción del hogar, entre otras experiencias traumáticas. Los participantes utilizaron una escala dicotómica (0 = *No*; 1 = *Sí*) para informar si habían experimentado alguna de estas situaciones a lo largo de su vida. Para calcular la cantidad de trauma a la que habían sido expuestos se sumaron todas las respuestas otorgadas a los 20 eventos adversos.

Apoyo familiar. Para examinar la percepción de apoyo familiar se utilizó un cuestionario de 13 ítems desarrollado para la investigación primaria (Arroyo-Pérez, 2019). Algunos de los ítems incluidos en el cuestionario son: *Mi familia es unida*, *Mi familia me escucha* y *Me siento amado por mi familia*. Todos los ítems se responden por medio de una escala Likert de 7 puntos (de 0 = *Nunca* a 6 = *Siempre*). A mayor puntuación, mayor percepción de apoyo familiar. Este instrumento mostró buena consistencia interna ($\alpha = .87$).

Análisis de datos

Para evaluar si existe relación entre las puntuaciones de las medidas de ejecución y de autoinforme de FE se utilizó un análisis de correlación producto-momento de Pearson. Además, para examinar la relación de algunos de los predictores de las FE definidos por la literatura (p. ej., violencia comunitaria, cantidad de trauma, apoyo familiar, etc.) se realizó un análisis de regresión lineal múltiple. Se utilizó el coeficiente estandarizado (β) para identificar la relación de cada predictor con las puntuaciones de las medidas de ejecución y de autoinforme de FE. Además, se calculó el coeficiente de determinación (R^2) de cada modelo como medida del tamaño del efecto. Todos los análisis se realizaron con el programa SPSS v.26 (IBM Corp., 2019).

Resultados

La Tabla 2 muestra los resultados de las correlaciones entre las puntuaciones de las subpruebas que miden las FE de *planificación y organización, flexibilidad cognitiva y control inhibitorio*, evaluadas a través de pruebas neuropsicológicas de ejecución y de autoinforme. Se encontraron relaciones fuertes (coeficiente r de .49 a .65) entre las puntuaciones de las distintas áreas de la prueba de Delis-Kaplan (prueba neuropsico-

lógica de ejecución). De igual modo, las distintas subpruebas del CEFI (medida de autoinforme) se relacionaron significativamente (coeficiente r de .59 a .63). Al examinar la relación entre las puntuaciones de las medidas de neuropsicológicas de ejecución y las de autoinforme, se encontraron algunas relaciones significativas entre medidas de distintas dimensiones de las funciones ejecutivas. Por ejemplo, se encontró que las puntuaciones de las subescalas del CEFI para *planificación y organización* se relacionaban significativamente y de

Tabla 2

Correlaciones entre las puntuaciones de las funciones ejecutivas medidas a través de pruebas neuropsicológicas de ejecución y de autoinforme.

	M	DE	1	2	3	4	5	6	7	8	9	10	11	12	13
1. D-KEFS Sorting Test: Condition 1: Confirmed Correct Sort	6.42	2.75													
2. D-KEFS Sorting Test: Condition 1: Free Sorting Description Score	6.63	3.09	.96**												
3. D-KEFS Sorting Test: Condition 2: Sort Recognition	7.38	3.32	.89**	.90**											
4. D-KEFS -Trail Making Test: Number Sequencing	9.53	2.69	.48**	.51**	.51**										
5. D-KEFS -Trail Making Test: Letter Sequencing	6.75	4.13	.68**	.68**	.69**	.72**									
6. D-KEFS -Trail Making Test: Number-Letter Switching	7.35	3.17	.60**	.58**	.63**	.67**	.75**								
7. D-KEFS-Color Test: Inhibition/Switching	8.13	3.79	.54**	.53**	.51**	.49**	.57**	.65**							
8. CEFI_Organización	95	13.7	.21	.17	.15	.29*	.33**	.33**	.19						
9. CEFI_Planificación	90.2	15.4	.20	.19	.23	.37**	.40**	.42**	.21	.77**					
10. CEFI_Flexibilidad Cognitiva	100	14.5	.11	.09	.07	.24	.19	.18	.09	.57**	.60**				
11. CEFI_Control Inhibitorio	90.4	19.9	.29*	.27*	.27*	.44**	.33**	.31*	.21	.59**	.61**	.63**			
12. Violencia Comunitaria	-	-	-.21	-.25	-.28*	-.10	-.35**	-.27*	-.11	-.17	-.14	-.1	-.04		
13. Cantidad de Trauma	6.82	3.78	-.34**	-.34**	-.38**	-.36**	-.34**	-.35**	-.22	-.08	-.17	-.16	-.15	.63**	
14. Apoyo Familiar	53.3	15	.10	.05	.01	.12	0	.01	0	.15	.24	.36**	.22	-.03	-.50**

Nota. ** $p < .01$ (2 colas); * $p < .05$ (2 colas).

Tabla 3

Resultados de regresión para los indicadores de Planificación y Organización con medidas de neuropsicológicas de ejecución y de autoinforme.

Planificación y Organización												
	Medidas neuropsicológicas de ejecución											
	D-KEFS Sorting Test: Condition 1: Confirmed Correct Sort				D-KEFS Sorting Test: Condition 1: Free Sorting Description Score				D-KEFS Sorting Test: Condition 2: Sort Recognition			
	β	SE	p	[95% IC]	β	SE	p	[95% IC]	β	SE	p	[95% IC]
Violencia comunitaria	.05	.98	.76	[-1.68, 2.28]	.03	1.09	.85	[-2.11, 2.41]	.06	1.14	.73	[-1.86, 2.65]
Cantidad de Trauma	-.43	.15	.04	[-.59, -.02]	-.45	.16	.03	[-.66, -.06]	-.55	.17	.01	[-.81, -.134]
Apoyo Familiar	-.11	.02	.47	[-.08, .05]	-.17	.03	.28	[-.10, .03]	-.25	.03	.10	[-.12, .018]
Adjusted R ²	.08				.09				.14			
R ²	.13				.14				.18			
	F (3,56) = 2.67, p = .05				F (3,56) = 3.038, p = .036				F (3,56) = 4.34, p = .008			

Planificación y Organización												
	Medidas de autoinforme											
	CEFI_Organización				CEFI_Planificación							
	β	SE	p	[95% IC]	β	SE	p	[95% IC]	β	SE	p	[95% IC]
Violencia comunitaria	-.34	5.00	.06	[-18.63, 1.30]	-.18	5.63	.31	[-16.88, 6.69]				
Cantidad de Trauma	.27	.76	.20	[-.78, 2.85]	.08	.86	.68	[-1.19, 1.94]				
Apoyo Familiar	.27	.15	.09	[-.01, .567]	.28	.16	.08	[-.011, .603]				
Adjusted R ²	.03				.03							
R ²	.08				.08							
	F (3,56) = 1.63, p = .193				F (3,56) = 1.68, p = .182							

forma moderada (coeficiente r de .29 a .42) con las puntuaciones del D-KEFS-Trail Making Test: Number Sequencing (flexibilidad cognitiva), D-KEFS-Trail Making Test: Letter Sequencing (flexibilidad cognitiva), D-KEFS-Trail Making Test: Number-Letter Switching y D-KEFS-Color Test: Inhibition/Switching (control inhibitorio). De igual forma, la subprueba que examina el *control inhibitorio* en el CEFI tuvo una relación significativa con las puntuaciones de las subpruebas del Delis Kaplan que examinan *planificación y organización* (D-KEFS Sorting Test: Condition 1: Confirmed Correct Sort; D-KEFS Sorting Test: Condition 1: Free Sorting Description Score;

D-KEFS Sorting Test: Condition 2: Sort Recognition), *flexibilidad cognitiva* (D-KEFS -Trail Making Test: Number Sequencing; D-KEFS -Trail Making Test: Letter Sequencing) y *control inhibitorio* (D-KEFS -Trail Making Test: Number-Letter Switching). Estas correlaciones van de .27 a .44.

Se realizó un modelo de regresión múltiple para examinar si existe relación entre la exposición a la violencia comunitaria, la cantidad de eventos traumáticos informados y el apoyo familiar, y las puntuaciones de las subpruebas que miden las funciones ejecutivas de *planificación y organización*, *flexibilidad cognitiva* y *control in-*

hibitorio. En la Tabla 3 se presentan los resultados del modelo de regresión múltiple para las medidas de *planificación y organización* en pruebas neuropsicológicas de ejecución y de autoinforme. Los resultados indican que solo la cantidad de trauma tuvo un efecto negativo y significativo con respecto a las puntuaciones de las subpruebas de la prueba Delis-Kaplan que examinan planificación y organización: D-KEFS Sorting Test: Condition 1: Confirmed Correct Sort ($\beta = -.43, p < .05$), Sorting Test: Condition 1: Free Sorting

Description Score ($\beta = -.45, p < .05$), D-KEFS Sorting Test: Condition 2: Sort Recognition ($\beta = -.55, p < .05$). No se encontró ninguna relación significativa entre la exposición a la violencia, la cantidad de eventos traumáticos reportados y el apoyo familiar por una parte y las puntuaciones de las FE de *planificación y organización* medidas con pruebas de autoinforme por otra parte.

En la Tabla 4 se presentan los resultados relacionados con las medidas de *flexibilidad cognitiva*. Los resultados indican que la cantidad de

Tabla 4

Resultados de regresión para los indicadores de Flexibilidad Cognitiva y Control Inhibitorio con medidas neuropsicológicas de ejecución y de autoinforme.

	Flexibilidad cognitiva											
	D-KEFS -Trail Making Test: Number Sequencing				D-KEFS -Trail Making Test: Letter Sequencing				CEFI_Flexibilidad Cognitiva			
	β	SE	<i>p</i>	[95% IC]	β	SE	<i>p</i>	[95% IC]	β	SE	<i>p</i>	[95% IC]
Violencia comunitaria	.30	.92	.08	[.13, 3.07]	-.13	1.44	.44	[-.50, 1.58]	-.33	5.00	.06	[-19.17, -.70]
Cantidad de Trauma	-.65	.14	.00	[-.75, -.15]	-.34	.22	.09	[-.79, .01]	.29	.76	.14	[-.39, 2.51]
Apoyo Familiar	-.19	.02	.21	[-.08, .02]	-.17	.04	.27	[-.14, .06]	.50	.15	.00	[.18, .78]
Adjusted R ²	.13				.12				.14			
R ²	.18				.16				.185			
	F (3, 56) = 4.183, <i>p</i> = .010				F (3, 56) = 3.72, <i>p</i> = .016				F (3, 56) = 4.247, <i>p</i> = .009			

Tabla 5

Resultados de regresión para los indicadores de Control Inhibitorio con medidas neuropsicológicas de ejecución y de autoinforme.

	Control Inhibitorio											
	D-KEFS -Trail Making Test: Number-Letter Switching				D-KEFS-Color Test: Inhibition/Switching				CEFI_Control Inhibitorio			
	β	SE	<i>p</i>	[95% IC]	β	SE	<i>p</i>	[95% IC]	β	SE	<i>p</i>	[95% IC]
Violencia comunitaria	.03	1.12	.86	[-2.03, 2.58]	.13	1.32	.45	[-1.68, 4.06]	-.01	7.38	.97	[-14.38, 15.30]
Cantidad de Trauma	-.48	.17	.02	[-.72, -.06]	-.41	.213	.05	[-.83, .01]	-.04	1.13	.82	[-.25, 2.09]
Apoyo Familiar	-.23	.03	.60	[-.11, .04]	-.20	.04	.21	[-.12, .04]	.19	.22	.60	[-.17, .69]
Adjusted R ²	.11				.028				.000			
R ²	.16				.077				.051			
	F (3, 56) = 3.56, <i>p</i> = .020				F (3, 56) = 1.557, <i>p</i> = .210				F (3, 56) = 1.00, <i>p</i> = .400			

eventos traumáticos se relaciona con las puntuaciones de la subprueba de D-KEFS -Trail Making Test: Number Sequencing para flexibilidad cognitiva ($\beta = -.65, p < .05$). En el caso de la subprueba D-KEFS -Trail Making Test: Letter Sequencing, aunque el resultado no fue significativo, se encontró una relación moderada ($\beta = -.34, p = .09$). Un resultado similar se observa al examinar el efecto de la violencia comunitaria en la flexibilidad cognitiva evaluada a través del D-KEFS-Trail Making Test: Number Sequencing. Se encontró un efecto moderado, aunque no estadísticamente significativo ($\beta = .30, p = .08$). Al considerar las relaciones con las puntuaciones de las pruebas de autoinforme encontramos que la violencia comunitaria tiene un efecto moderado, aunque no estadísticamente significativo con la flexibilidad cognitiva ($\beta = -.33, p = .06$). Además, se encontró que el apoyo familiar se relaciona significativamente con la flexibilidad cognitiva evaluada a través de autoinforme ($\beta = .50, p < .05$).

Por último, la Tabla 5 presenta los resultados relacionados con las medidas de control inhibitorio. Al examinar las relaciones con las puntuaciones de las pruebas neuropsicológicas de ejecución encontramos que solo la cantidad de trauma mostró una relación significativa al evaluarse con la subprueba D-KEFS -Trail Making Test: Number-Letter Switching ($\beta = -.48, p < .05$). El resultado con las puntuaciones de la subprueba D-KEFS-Color Test: Inhibition/Switching, aunque no fue significativo, refleja un efecto moderado ($\beta = -.41, p = .05$). No se encontró ninguna relación significativa entre la exposición a la violencia, la cantidad de eventos traumáticos informada y el apoyo familiar por una parte y las puntuaciones de control inhibitorio evaluado a través de pruebas de autoinforme por la otra.

La prueba Shapiro-Wilk para el análisis de normalidad indica que las puntuaciones de algunas variables no se distribuyen normalmente

(p. ej., violencia comunitaria, cantidad de trauma, violencia comunitaria, D-KEFS Sorting Test: Condition 1: Confirmed Correct Sort, CEFI_Organización, D-KEFS -Trail Making Test: Number Sequencing, D-KEFS -Trail Making Test: Letter Sequencing, D-KEFS -Trail Making Test: Number-Letter Switching, D-KEFS-Color Test: Inhibition/Switching). Por otro lado, los indicadores de tolerancia (violencia comunitaria = .48; cantidad de trauma = .36; apoyo familiar = .60) son mayores a .20, mientras que los indicadores de factor de inflación de varianza (Variance Inflation factor [VIF], en inglés) fueron todos menores de 10 (violencia comunitaria = 2.05; cantidad de trauma = 2.75; apoyo familiar = 1.65) lo que indica que no existen problemas de multicolinealidad en los análisis de regresión. Con el objetivo de ofrecer una mayor precisión en los estimados debido a la limitación de la muestra y la no normalidad de algunas variables, se calcularon intervalos de confianza al 95% con 1000 muestras *bootstrapped* en todos los análisis (Field, 2018). Estos intervalos permiten una comprensión más clara y comparable de la proporción de la varianza explicada (Bologna, 2014). Todos los análisis se realizaron con el programa SPSS v.26 (IBM Corp., 2019).

Discusión

Este estudio tenía como propósito explorar si existe una relación entre las puntuaciones de pruebas de autoinforme y de ejecución para la evaluación de las FE, así como conocer las relaciones con diversos predictores del funcionamiento ejecutivo (p. ej., la cantidad de eventos traumáticos, el apoyo familiar y la violencia comunitaria). De la literatura revisada se desprende que existen ciertas dificultades u obstáculos al evaluar las FE utilizando estas dos medidas, es-

pecíficamente al comparar las puntuaciones. Algunos autores han encontrado discrepancias entre estas puntuaciones (Nordvall et al., 2017; Toplak et al., 2013). Para verificar estos hallazgos se evaluó la relación entre ellas.

Las investigaciones señalan que la ausencia de correlación entre las medidas de autoinformes y pruebas neuropsicológicas se debe a que evalúan aspectos diferentes de las funciones ejecutivas mencionadas (Toplak et al., 2013; Suchy, 2015). Por lo tanto, los resultados de ambas medidas no deben ser interpretados como equivalentes o intercambiables. Esto puede implicar que ambas medidas son esenciales para realizar una evaluación completa, ya que permiten al evaluador desarrollar recomendaciones prescriptivas, específicas y holísticas (emocionales, conductuales y cognitivas) en relación con el funcionamiento ejecutivo.

Las únicas relaciones significativas encontradas fueron aquellas que involucran a las variables del mismo constructo. Se encontró relación entre las funciones ejecutivas *planificación y organización* y *flexibilidad cognitiva* tanto en las medidas de autoinforme como en las neuropsicológicas. Esto puede deberse a que la habilidad de planificación se compone de otros procesos como la memoria y la flexibilidad cognitiva. El control inhibitorio y la flexibilidad cognitiva son componentes importantes en las tareas relacionadas a la planificación, por ejemplo, para determinar si los participantes toman tiempo para pensar en lugar de actuar impulsivamente (Asato, Sweeney, & Luna, 2006). Cualquier tarea que requiera manejo del tiempo, memoria y flexibilidad en el pensamiento será un desafío para las personas con limitaciones en la capacidad de planificación.

La literatura menciona que la exposición a eventos traumáticos impacta en las FE al afectar estructuras y regiones cerebrales (Teicher & Samson, 2016). Por ejemplo, las regiones cere-

brales más afectadas en los casos de jóvenes que sufren maltrato son la corteza prefrontal y la corteza orbitofrontal, que están relacionadas con las FE (Teicher & Samson, 2016). Se encontró una relación estadísticamente significativa y negativa entre cantidad de eventos traumáticos y algunas funciones ejecutivas medidas por las pruebas neuropsicológicas. Estos resultados indican que, a mayor cantidad de eventos traumáticos, menor es el desempeño en las medidas neuropsicológicas que evalúan las funciones ejecutivas *control inhibitorio, planificación y organización* y *flexibilidad cognitiva*. Sin embargo, es importante tomar en cuenta que no podemos obtener conclusiones sólidas por la limitación de la muestra.

Conclusión

La utilidad de una de estas medidas sobre otra es un aspecto que se debe seguir investigando. Los resultados de esta investigación indican que el uso de un tipo de prueba no excluye el de la otra, por lo que parecería ser de mayor utilidad usar ambas medidas para medir las FE. Esto puede estar sugiriendo que ambas medidas examinan aspectos diferentes de las FE. El autoinforme mide el área conductual y emocional de la vida del sujeto y las pruebas neuropsicológicas evalúan los factores metacognitivos de las funciones en un ambiente con condiciones estructuradas. Estos resultados nos señalan que las FE son complejas y se componen de diferentes niveles de procesos metacognitivos, reguladores emocionales y reguladores conductuales.

No obstante, para futuras investigaciones se recomienda utilizar otro modelo de las funciones ejecutivas en donde se identifiquen los diferentes procesos que interactúan en las mismas, tanto las funciones de orden superior como los componentes emocionales y conductual. Por otro

lado, se sugiere que se identifiquen más variables predictoras para ser utilizadas en la evaluación de estas dos medidas. Aunque los resultados de este estudio comparten algunas semejanzas con lo hallado en la literatura, sirven para propiciar otros estudios con una muestra mayor, al igual que una mayor selección de pruebas de ambas medidas. Una vez incorporadas otras pruebas y variables se podrá tener una visión más amplia y clara sobre la utilidad de ambas medidas.

Durante el progreso de este estudio se pudieron identificar varias limitaciones. Al ser un análisis de datos secundarios, se analizaron los datos recopilados en la investigación primaria. Esto presenta varias limitaciones en cuanto a las medidas de ejecución y autoinforme de las FE disponibles para examinar. De igual modo, al considerar los predictores, solo se pudieron utilizar en los análisis algunos de los mencionados por la literatura. Cabe señalar que en el estudio principal se demostró que existen diferencias en las puntuaciones entre el grupo de jóvenes transgresores y el grupo de jóvenes de la comunidad sin problemas legales. Durante la revisión de literatura se encontraron otras variables que podrían ser consideradas, las cuales no se pudieron explorar en la presente investigación. Algunas de estas variables son el estatus socioeconómico y el tipo de crianza.

Es importante mencionar que, al ser un análisis de datos secundarios, la muestra de adolescentes varones de entre 14 y 18 años disponible fue pequeña, lo que limita la posibilidad de generalización de los resultados. Estudios posteriores deben examinar diferencias o convergencias en la medición de las funciones ejecutivas utilizando muestras más heterogéneas en género y edad. De igual modo, la no significancia de algunos predictores con relación a las medidas de funciones ejecutivas (de desempeño/ejecución y autoinforme) puede deberse a la muestra limitada con la que

contó el estudio. Futuras investigaciones deben integrar muestras más amplias y otras medidas de variables (p. ej., violencia comunitaria, apoyo familiar) para ofrecer apoyo a los resultados encontrados.

Referencias

- Arroyo-Pérez, Y. (2019). *Perfil del funcionamiento ejecutivo y aspecto psicosociales de adolescentes varones intervenidos por el tribunal de menores de Puerto Rico* (Tesis doctoral inédita). Universidad Carlos Albizu, Puerto Rico.
- Asato, M., Sweeney, J., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia*, 44(12), 2259-2269. doi: [10.1016/j.neuropsychologia.2006.05.010](https://doi.org/10.1016/j.neuropsychologia.2006.05.010)
- Bakar, E., Taner, Y., Soysal, A., Karakas, S., & Turgay, A. (2011). Behavioral rating inventory and laboratory tests measure different aspects of executive functioning in boys: A validity study. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology*, 21(4), 302-316. doi: [10.5455/bcp.20111004014003](https://doi.org/10.5455/bcp.20111004014003)
- Barkley, R. (2012). *Executive functions: What they are, how they work, and why they evolved*. Guilford Press.
- Barkley, R., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, 36(2), 137-161. doi: [10.1080/87565641.2010.549877](https://doi.org/10.1080/87565641.2010.549877)
- Berg, E. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39(1), 15-22. doi: [10.1080/00221309.1948.9918159](https://doi.org/10.1080/00221309.1948.9918159)
- Bologna, E. (2014). Estimación por intervalo del tamaño

- del efecto expresado como proporción de varianza explicada. *Revista Evaluar*, 14(1), 43-66. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Blakemore, S., Burnett, S., & Dahl, R. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping*, 31(6), 926-933. doi: [10.1002/hbm.21052](https://doi.org/10.1002/hbm.21052)
- Carlson, S., Zelazo, P., & Faja, S. (2013). Executive function. En P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology. Volumen I: Body and mind* (pp. 706-743). New York, NY: Oxford University Press.
- Comesaña, A., Stelzer, F., & Introzzi, I. (2017). Inhibición de borrado en adultos mayores: Aportes para la validación de una tarea. *Revista Evaluar*, 17(2), 143-153. doi: [10.35670/1667-4545.v17.n2.18727](https://doi.org/10.35670/1667-4545.v17.n2.18727)
- Delis, D., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan Executive Function System (D-KEFS): Examiner's manual*. San Antonio, TX: The Psychological Corporation.
- DePrince, A., Weinzierl, K., & Combs, M. (2009). Executive function performance and trauma exposure in a community sample of children. *Child Abuse & Neglect*, 33(6), 353-361. doi: [10.1016/j.chiabu.2008.08.002](https://doi.org/10.1016/j.chiabu.2008.08.002)
- Dorna-Pesquera, S., & Caballero-González, H. (2010). *Informe estadístico anual de las procuradurías de menores: Año Fiscal 2008 - 2009 comparativo con el 2007 -2008, departamento de justicia*. Recuperado del sitio de Estadística del Gobierno de Puerto Rico: <http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=VXTAUK13saA%-3D&tabid=186>
- Donders, F. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431. doi: [10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics (5ª ed.)*. London, UK: SAGE Publications.
- Fuster, J. (2000). Executive frontal functions. *Experimental Brain Research*, 133(1), 66-70. doi: [10.1007/s002210000401](https://doi.org/10.1007/s002210000401)
- García-Molina, A., Tirapu-Ustárroz, J., & Roig-Rovira, T. (2007). Validez ecológica en la exploración de las funciones ejecutivas. *Anales de Psicología*, 23(2), 289-299. Recuperado de <https://revistas.um.es/analesps/index>
- Gioia, G., Isquith, P., Guy, S., & Kenworthy, L. (2000). *Behavior rating inventory of executive function: BRIEF*. Odessa, FL: Psychological Assessment Resources.
- Harvey, P. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, 14(1), 91-99. doi: [10.31887/DCNS.2012.14.1/pharvey](https://doi.org/10.31887/DCNS.2012.14.1/pharvey)
- Hoskin, R. (2012). The dangers of self-report. *Science Brainwaves*.
- IBM Corp. (2019). IBM SPSS Statistics for Windows (Versión 26.0). [Software de cómputo]. Armonk, NY: IBM Corp.
- Krikorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: A standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology*, 16(6), 840-850. doi: [10.1080/01688639408402697](https://doi.org/10.1080/01688639408402697)
- Ley N° 88. (1986). *Ley de Menores*. Puerto Rico. Recuperado de www.lexjuris.com
- Loewenstein, D., Argüelles, S., Bravo, M., Freeman, R., Argüelles, T., Acevedo, A., & Eisdorfer, C. (2001). Caregivers' judgments of the functional abilities of the Alzheimer' disease patient: A comparison of proxy reports and objective measures. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 56(2), 78-84. doi: [10.1093/geronb/56.2.78](https://doi.org/10.1093/geronb/56.2.78)

10.1093/geronb/56.2.P78

- Logan, G., & Cowan, W. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295-327. doi: [10.1037/0033-295X.91.3.295](https://doi.org/10.1037/0033-295X.91.3.295)
- Lugo-Gil, J., & Tamis-LeMonda, C. (2008). Family resources and parenting quality: Links to children's cognitive development across the first 3 years. *Child Development*, *79*(4), 1065-1085. doi: [10.1111/j.1467-8624.2008.01176.x](https://doi.org/10.1111/j.1467-8624.2008.01176.x)
- Maldonado-Belmonte, M. (2016). *Adaptación del BRIEF (Behavior Rating Inventory of Executive Function) a población española y su utilidad para el diagnóstico del trastorno por déficit de atención-hiperactividad subtipos inatento y combinado*. (Tesis doctoral). Recuperado de <https://dialnet.unirioja.es>
- Mesulam, M. (2002). The human frontal lobes: Transcending the default mode through contingent encoding. En D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 8-30). New York, NY: Oxford University Press.
- Mitchell, M., & Miller, L. (2008). Prediction of functional status in older adults: The ecological validity of four Delis-Kaplan Executive Function System tests. *Journal of Clinical and Experimental Neuropsychology*, *30*(6), 683-690. doi: [10.1080/13803390701679893](https://doi.org/10.1080/13803390701679893)
- Naglieri, J., & Goldstein, S. (2013). *Comprehensive Executive Function Inventory*. North Tonawanda, NY: Multi-Health Systems.
- Nęcka, E., Lech, B., Sobczyk, N., & Śmieja, M. (2012). How much do we know about our own cognitive control? *European Journal of Psychological Assessment*, *28*(3), 240-247. doi: [10.1027/1015-5759/a000147](https://doi.org/10.1027/1015-5759/a000147)
- Nelson, E., Leibenluft, E., McClure, E., & Pine, D. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, *35*(2), 163-174. doi: [10.1017/S0033291704003915](https://doi.org/10.1017/S0033291704003915)
- Nordvall, O., Jonsson, B., & Stigsdotter-Neely, A. (2017). Self-reported and performance-based measures of executive functions in interned youth. *Psychology, Crime & Law*, *23*(3), 240-253. doi: [10.1080/1068316X.2016.1239725](https://doi.org/10.1080/1068316X.2016.1239725)
- Op den Kelder, R., Van den Akker, A., Geurts, H., Lindauer, R., & Overbeek, G. (2018). Executive functions in trauma-exposed youth: A meta-analysis. *European Journal of Psychotraumatology*, *9*(1), Artículo N° 1450595. doi: [10.1080/20008198.2018.1450595](https://doi.org/10.1080/20008198.2018.1450595)
- Patros, C., Tarle, S., Alderson, R., Lea, S., & Arrington, E. (2019). Planning deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review of tower task performance. *Neuropsychology*, *33*(3), 425-444. doi: [10.1037/neu0000531](https://doi.org/10.1037/neu0000531)
- Paulhus, D., & Vazire, S. (2007). The self-report method. En R.W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224-239). New York, NY: The Guildford Press.
- Pérez-Salas, C., Ramos, C., Oliva, K., & Ortega, A. (2016). Bifactor modeling of the Behavior Rating Inventory of Executive Function (BRIEF) in a Chilean sample. *Perceptual and Motor Skills*, *122*(3), 757-776. doi: [10.1177/0031512516650441](https://doi.org/10.1177/0031512516650441)
- Raine, A. (2002). Biosocial studies of antisocial and violent behavior in children and adults: A review. *Journal of Abnormal Child Psychology*, *30*(4), 311-326. Recuperado de <https://www.springer.com/journal/10802>
- Rathinabalan, I., & Naaraayan, S. (2018). Effect of personal and school factors on juvenile delinquency. *The Indian Journal of Pediatrics*, *85*(7), 569. doi: [10.1007/s12098-017-2566-z](https://doi.org/10.1007/s12098-017-2566-z)

- Reitan, R. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8(3), 271-276. Recuperado de <https://journals.sagepub.com/home/pms>
- Roberts, S., & Apaza, R. (2011). Funciones ejecutivas, atención y conducta. *Revista de Investigación UCSP*, 2(2), 7-28. Recuperado de <http://repositorio.unsa.edu.pe>
- Rueda, M., Cómbita, L., & Pozuelos, J. (2016). Childhood and adolescence. En T. Strobach & J. Karbach (Eds.), *Cognitive training: An overview of features and applications* (pp. 33-44). doi: 10.1007/978-3-319-42662-4_4
- Schore, A. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal Publication of The World Association for Infant Mental Health*, 22(1-2), 7-66. Recuperado de <https://onlinelibrary.wiley.com/journal/10970355>
- Sedó, M. (2004). '5 Digit Test': A multilingual non-reading alternative to the Stroop test. *Revista de Neurología*, 38(9), 824-828. doi: 10.33588/rn.3809.2003545
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Serie B: Biological Sciences*, 298(1089), 199-209. doi: 10.1098/rstb.1982.0082
- Sowell, E., Thompson, P., Holmes, C., Batth, R., Jernigan, T., & Toga, A. (1999). Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *Neuroimage*, 9(6), 587-597. doi: 10.1006/nimg.1999.0436
- Spooner, D., & Pachana, N. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, 21(4), 327-337. doi: 10.1016/j.acn.2006.04.004
- Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology*, 52(3), 216-224. doi: 10.1002/dev.20445
- Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662. doi: 10.1037/h0054651
- Suchy, Y. (2015). *Executive functioning: A comprehensive guide for clinical practice*. New York, NY: Oxford University Press.
- Teicher, M., & Samson, J. (2016). Annual research review: Enduring neurobiological effects of childhood abuse and neglect. *Journal of Child Psychology and Psychiatry*, 57(3), 241-266. doi: 10.1111/jcpp.12507
- Thierry, A., Glowinski, J., Goldman-Rakic, P., & Christen, Y. (Eds.). (1994). *Motor and cognitive function of the prefrontal cortex*. Berlin, Alemania: Springer.
- Toplak, M., West, R., & Stanovich, K. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, 54(2), 131-143. doi: 10.1111/jcpp.12001
- Vasterling, J., & Brailey, K. (2005). Neuropsychological findings in adults with PTSD. En J. Vasterling & C. Brewin (Eds.), *Neuropsychology of PTSD: Biological, Cognitive, and Clinical Perspectives* (pp. 178-207). New York, NY: The Guilford Press.
- Vélez-Pastrana, M., González, R., Rodríguez-Cardona, J., Purcell-Baerga, P., Alicea-Rodríguez, Á., & Levin, F. (2016). Psychometric properties of the Barkley Deficits in Executive Functioning Scale: A Spanish-language version in a community sample of Puerto Rican adults. *Psychological Assessment*, 28(5), 483-498. doi: 10.1037/pas0000171
- Zanetti, O., Frisoni, G., Rozzini, L., Bianchetti, A., & Tra-

bucchi, M. (1998). Validity of Direct Assessment of Functional Status as a tool for measuring Alzheimer's disease severity. *Age and Ageing*, 27(5), 615-622. doi: [10.1093/ageing/27.5.615](https://doi.org/10.1093/ageing/27.5.615)

Zelazo, P., Carter, A., Reznick, J., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1(2), 198-266. doi: [10.1037/1089-2680.1.2.198](https://doi.org/10.1037/1089-2680.1.2.198)

Zysset, A., Kakebeeke, T., Messerli-Bürgy, N., Meyer, A., Stülb, K., Leegee-Aschmann, C. S., & Jenni, O. (2018). Prediction of executive functions in preschoolers: Findings from the SPLASHY study. *Frontiers in Psychology*, 9. doi: [10.3389/fpsyg.2018.02060](https://doi.org/10.3389/fpsyg.2018.02060)

Propiedades Psicométricas de una prueba de admisión universitaria

Psychometric Properties of a Higher Education Admission Test

Tania Elena Moreira-Mora ¹ *

1 - Departamento de Orientación y Psicología, Comité Exámen de Admisión. Instituto Tecnológico de Costa Rica, Cartago.

Recibido: 30/09/2020 **Revisado:** 17/11/2020 **Aceptado:** 08/12/2020

Introducción
Método
Resultados
Discusión
Referencias

Resumen

Esta investigación pretende demostrar las propiedades psicométricas de la prueba de admisión del Instituto Tecnológico de Costa Rica y presentar las evidencias de contenido, constructo y criterio para el modelo de medida. En el estudio no experimental y correlacional participaron estudiantes que realizaron la prueba en 2019, de los cuales 57% eran hombres y 43% mujeres, 74% provenían de colegios públicos y el resto de colegios privados y subvencionados. Los hallazgos demostraron que la tabla de especificaciones proporciona las primeras evidencias asociadas al contenido, cuyos indicadores matemáticos y verbales se asociaron significativamente con el rasgo latente. También se comprobó el valor predictivo de estos indicadores y del promedio de la educación diversificada con respecto al rendimiento del estudiantado en el primer periodo universitario. Se concluyó que la validación debe ser un proceso permanente y riguroso para consolidar el diseño de una prueba y garantizar una interpretación y uso adecuado de sus resultados.

Palabras clave: *evidencias de validez, prueba estandarizada, aptitud académica, modelo de ecuaciones estructurales, modelo predictivo*

Abstract

The objective of this study is to demonstrate the psychometric properties of the Aptitude Academic Test of the Instituto Tecnológico of Costa Rica, and it presents evidences about the content, construct and criteria. On this non-experimental and correlational study participated the students who applied for this test on 2019, 57% of them were men and 43% women, 74% came from public high schools and the rest, from private and subsidized schools. The findings demonstrate that the specification table provides the first evidence associated with the test content, whose mathematical and verbal indicators are significantly associated with the trait of academic aptitude. Additionally, the predictive value of these test indicators and the grades of diversified education were verified with the performance of the students in the first period. It is concluded that validation should be a permanent and rigorous process to consolidate the design of a standardized test and guarantee the proper use and interpretation of its results.

Keywords: *evidence of validity, standardized test, academic aptitude, structural equation model, predictive model*

*Correspondencia a: tmoreira@tec.ac.cr

Introducción

Las instituciones de educación superior han venido utilizando sistemas, estrategias e instrumentos para la selección y admisión de sus estudiantes. En el caso de la universidad estatal, y específicamente en el del Instituto Tecnológico de Costa Rica (TEC), se ha utilizado la Prueba de Aptitud Académica (PAA) desde su apertura en 1973 para elegir a los candidatos con mayores probabilidades de éxito académico. Desde entonces, con algunas variantes en el diseño, el modelo de medida de la prueba se mantuvo enfocado tanto en la medición de habilidades matemáticas relacionadas con: operaciones básicas, ecuaciones, expresiones algebraicas, porcentajes, geometría, entre otras; como en habilidades verbales relativas a la comprensión lectora en textos cortos y largos, antónimos y el razonamiento mediante analogías y silogismos.

En el 2009, los responsables de la construcción de la prueba consideraron necesaria una revisión teórica y técnica del modelo de medida de la PAA con el objetivo de actualizarlo de acuerdo al perfil de entrada de los estudiantes, a las nuevas demandas cognoscitivas de las carreras y al compromiso ético de garantizar una medición equitativa acorde con las propiedades psicométricas requeridas para una prueba con altas consecuencias sociales. [Bringas-Benavídez y Pérez-Mejía \(2014\)](#) señalan que este tipo de pruebas, además de cumplir con los propósitos específicos de un examen de admisión y de seleccionar a los aspirantes a primer ingreso; pueden proporcionar un diagnóstico de las competencias desarrolladas en la formación previa, informar de las capacidades académicas de los aspirantes, y brindar insumos para generar acciones compensatorias con la finalidad de mejorar el rendimiento académico.

Gracias a esta revisión del modelo se logró establecer un conjunto de habilidades verbales y

matemáticas básicas que, para estudiantes y docentes, eran relevantes en el perfil de entrada de los estudiantes en el TEC. Estas han sido la base para iniciar una etapa de transición hacia un nuevo modelo de medición de la PAA a partir de un riguroso proceso de validación.

De esta manera, en el presente estudio se pretende demostrar las propiedades psicométricas de la Prueba de Aptitud Académica del Instituto Tecnológico de Costa Rica y analizar cuál es el grado de relación entre las variables observables de las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de aptitud académica, y cuál es el grado de validez predictiva de los componentes de la nota de admisión en el rendimiento académico de los estudiantes en su primer año de ingreso.

Por ello, en esta investigación se presentan los análisis más recientes de las propiedades psicométricas de la prueba, relativos a las evidencias de validez asociadas al contenido, criterio y constructo, desde la perspectiva de la teoría clásica de los test (TCT). Tales análisis se sustentan en la premisa de que cuantas más evidencias de validez se obtengan de una prueba, más sólido es el modelo de medida y más apropiado el uso de sus resultados.

Desde la perspectiva unificada propuesta por [Messick](#), la validez encierra

...un juicio integrado y evaluativo del grado en que la evidencia empírica y las razones teóricas apoyan lo adecuado y lo apropiado de las interpretaciones y las acciones basadas en las puntuaciones de las pruebas u otras formas de evaluación ([1995, p. 5](#)).

En los estándares para las pruebas educativas y psicológicas establecidos por la [American Educational Research Association \(AERA\)](#), la [American Psychological Association \(APA\)](#) y el [National Council on Measurement in Education](#)

(NCME), el proceso de validación es un asunto del grado en que la evidencia acumulada proporciona una base científica para la interpretación de las puntuaciones de la prueba y de su relevancia (2014). Además, señalan que existen cinco fuentes de evidencias de validez, basadas en el contenido, los procesos de respuesta, la estructura interna, la relación con otras variables y las consecuencias de los usos de la prueba.

La validez de contenido es la primera característica a analizar en una prueba y, como lo destacan Juárez-Hernández y Tobón (2018), es un componente importante de la validez de constructo, ya que se relaciona con el grado en que los ítems son relevantes y pertinentes al constructo de interés.

La construcción de los ítems de la PAA se fundamenta en las tareas cognitivas descritas en una tabla de especificaciones, lo que permite garantizar una adecuada conceptualización y operacionalización del constructo; es decir, especificar previamente las dimensiones a medir y sus indicadores, como lo destacan Escobar-Pérez y Cervo-Martínez (2008). Además, es el uso riguroso de esta tabla de especificaciones lo que garantiza las evidencias de validez basadas en el contenido (Rojas-Torres & Ordóñez-Gutiérrez, 2019).

También se pretende obtener evidencias del grado de relación entre las variables observables relativas a las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de *aptitud académica*. Como lo explica Cea-D'Ancona (2002), estas evidencias se pueden generar a partir de una teoría, de generalizaciones empíricas o de la estructura latente que el investigador espera encontrar en los datos.

Usualmente, la investigación en ciencias sociales ha usado los modelos de ecuaciones estructurales (SEM) para evaluar tanto las propiedades psicométricas del modelo de medida como para obtener evidencias de validez de constructo,

puesto que permiten relacionar las variables latentes con indicadores o variables observables.

Específicamente, en un análisis realizado sobre los indicadores del constructo de un modelo de admisión para predecir el rendimiento académico en postgrado, resultó que los predictores más potentes fueron las habilidades intelectuales como razonamiento verbal o numérico, dependiendo de la magnitud del coeficiente de validez predictiva, y de las características de los postgrados (Guglietta & Delgado-Álvarez, 2010).

En cuanto a las evidencias asociadas a los coeficientes de validez predictiva entre una medida criterio, usualmente el promedio del primer año universitario, y los predictores, como la puntuación de las variables de la batería de selección, se han usado modelos de regresiones mínimo cuadráticas (Contreras, Gallegos, & Meneses 2009; Young, 2001). Particularmente, se espera que el ordenamiento de los postulantes por nota de examen coincida con el ordenamiento que tendrían respecto de su potencial desempeño universitario (Manzi et al., 2010).

En estudios predictivos con pruebas de ingreso como predictores del rendimiento, se ha encontrado una correlación significativa entre los predictores y los resultados obtenidos por los estudiantes en los primeros semestres (García-Domínguez, 2016; López, Echazarreta, Pech, & Gómez, 2010).

Particularmente, Vergara-Díaz y Peredo-López (2017) encontraron que los puntajes de las pruebas de selección universitaria tienen una relación más débil con el rendimiento académico estudiantil en comparación con el promedio de nota de la educación media, que lo explica mejor al presentar una mejor correlación en el primer semestre y una más fuerte en el segundo. De acuerdo con Bleyaert (2010), esta tendencia también se ha encontrado en algunos estudios donde la calificación promedio (Grade Point Average,

GPA) de escuelas secundarias ha mostrado una mayor contribución predictiva que las puntuaciones de la prueba universitaria americana (American College Testing, ACT); aunque en otros estudios ACT ha sido un mejor predictor que el GPA universitario. Finalmente, la autora concluye que la combinación más eficaz para predecir el éxito universitario es el GPA de secundaria y la nota de ACT.

Una limitación en estos análisis predictivos es que los promedios de las calificaciones obtenidas en la universidad y secundaria no son buenos criterios, por la variación de un profesor a otro y de una institución a otra (Backhoff, Tirado, & Larrazolo, 2001). La otra limitación es el problema del sesgo en la selección de la muestra, conocido como restricción de rango, ya que los sujetos de evaluación son aquellos estudiantes que logran ingresar al sistema universitario (Contreras et al., 2009).

A modo de síntesis, destaca Moreira-Mora (2016) que los estudios de pruebas estandarizadas se han centrado en dos vertientes. Una es la psicométrica, fundamentada en los estándares para pruebas psicológicas y educativas establecidas por la AERA, la APA y el NCME, que competen a la calificación, confiabilidad y validez, procesos administrativos y de construcción (Kobrin & Kimmel, 2006; Lawrence, Rigor, Van Essen, & Jackson, 2003; Martínez-Arias, 2005; Young, 2001). La segunda está más orientada hacia los modelos cognitivos y la conceptualización del constructo, que se relacionan con las habilidades o dimensiones cognoscitivas abordadas por estas pruebas (Gierl, Tan, & Wang, 2005; Milewski, Johnsen, Glazer, & Kubota, 2005). No obstante, a pesar de que estos paradigmas se han desarrollado separadamente, en la actualidad existen propuestas que plantean posibles integraciones (Cliff & Montero, 2010).

En relación con el rasgo latente de *razona-*

miento general medido en la prueba de admisión del TEC, la teoría psicológica respalda teóricamente que un rasgo es evaluable en una única medición, puesto que los factores que determinan la capacidad general de una persona no deberían experimentar modificaciones notorias en un período relativamente breve, por lo tanto, esa estabilidad favorece las predicciones del desempeño (Donoso-Díaz & Hawes-Barrios, 2001). Al respecto, Contreras, Bravo y Sanhueza (2001) destacan que estas pruebas están referidas principalmente a rasgos estables a partir de los cuales pueden hacerse predicciones sobre desempeños futuros de los sujetos y que permiten clasificar y ordenar a los postulantes sobre una base científicamente sustentada.

Este rasgo de razonamiento general ha sido medido con dos factores: *razonamiento matemático* y *verbal*. El primero es una habilidad que involucra la resolución de problemas a partir del aprendizaje de conocimientos y del desarrollo de estrategias cognitivas y metacognitivas. Para la Organización para la Cooperación y Desarrollo Económico (OCDE), la *competencia matemática* es la capacidad del estudiantado para razonar y emplear conceptos, procedimientos, hechos y herramientas matemáticas para describir, explicar y predecir fenómenos de diverso tipo (Instituto Nacional de Evaluación Educativa, 2019).

Una de las categorías de este factor es la *resolución de problemas*, que implica el uso de conocimientos básicos, definir y ejecutar una estrategia de solución. La resolución de un problema, como señalan Marino y Rodríguez (2009), demanda “una complejidad cognitiva mayor, en tanto que el alumno debe elaborar su propio método de resolución, apelando a sus conocimientos previos, estableciendo nuevas relaciones entre ellos y, además, empleando diversos procedimientos, tanto algorítmicos como heurísticos” (p. 161).

La categoría de *razonamiento deductivo*

“implica procedimientos matemáticamente bien definidos para extraer las consecuencias que se siguen con certeza o necesidad de algún tipo de evidencia” (Lassiter & Goodman, 2015, p. 124). Mediante la deducción, se conduce de forma sistemática de un grupo de proposiciones a otro. En tanto que, mediante la inducción, se pueden crear nuevos conceptos analizando semejanzas o diferencias; este razonamiento se basa en “operaciones como clasificar, completar series, hacer analogías y comparaciones con diferentes tipos de símbolos (verbales, figuras, entre otros), que permiten llegar a hacer inferencias para definir esos nuevos conceptos y posteriormente aplicarlos y evaluarlos” (Iriarte-Díaz-Granados et al., 2010, p. 42). Finalmente, el *razonamiento con figuras* está más relacionado con la habilidad de percibir formas y transformarlas mentalmente, mediante la manipulación mental de formas y objetos percibidos.

En cuanto al *razonamiento verbal*, ha sido definido como la capacidad para utilizar el lenguaje verbal en el análisis semántico e inferencial en la lectura de diversos textos. Según el [College Board \(2014\)](#) las pruebas de aptitud académica se enfocan, por un lado, en la habilidad para reconocer las relaciones entre las distintas partes de una oración y reconocer el significado de las palabras en contexto y, por otro lado, en la habilidad para analizar, inferir y establecer relaciones analógicas, así como la habilidad para la síntesis de la información y la comparación entre las partes de un mismo texto o entre textos o lecturas diferentes.

En el caso de la PAA del TEC, estos indicadores matemáticos y verbales se han ido especificando en los últimos años, a partir de la inclusión de distintas tareas cognitivas que permiten medir de manera más precisa el rasgo latente de razonamiento general.

Método

El enfoque es cuantitativo por el interés en demostrar las propiedades psicométricas de la PAA, por tal razón, este estudio se diseñó como no experimental y correlacional y se realizó con los resultados de la prueba de admisión aplicada en el 2019.

Participantes

Para el análisis factorial confirmatorio se incluyó a los 16285 examinados de las tres convocatorias ordinarias del 2019, de los cuales 74.1% provenían de colegios públicos, 17.1% de establecimientos privados y 8.3% de subvencionados. De estos, 57.2% eran hombres y 42.8% mujeres. En cuanto al lugar de procedencia, la provincia con más alto porcentaje de población fue San José, a donde pertenecía el 28.8% de los examinados, seguida de Alajuela con 20.9%, Cartago con un 20.6% y Heredia con un 8.4%; en tanto que el 21.1% restante era de las provincias costeras de Guanacaste, Limón y Puntarenas. Finalmente, de acuerdo con el Índice de Desarrollo Social (IDS), 76.7% de los examinados provenía de distritos con un índice igual o superior a 60 y 23.1% menor a 60 (niveles más bajos de desarrollo). El IDS es calculado por el Ministerio de Planificación Nacional y Política Económica ([MIDEPLAN, 2017](#)) a partir de un conjunto de índices socioeconómicos distribuidos en cinco dimensiones: económica, participación social, salud, educación y seguridad, para clasificar los distritos de Costa Rica por su nivel de desarrollo. De esta manera, un valor de 100 indica que, en promedio, el distrito tiene los mejores indicadores en relación con los demás.

La muestra para los análisis predictivos es-

tuvo conformada por los estudiantes que se matricularon en el primer semestre del 2020: 66% hombres y 34% mujeres, de los cuales 62% estudiaba en colegios públicos, 23% en colegios privados y 15% en colegios subvencionados. Además, el 85.6% provenía de las provincias de San José, Cartago, Alajuela y Heredia, el grupo restante era de las provincias costeras de Guanacaste, Puntarenas y Limón.

Instrumentos

La fuente de las notas del componente matemático y verbal de la PAA fue el Comité Examen de Admisión. Este equipo técnico construye y valida cada año la PAA, la cual está estructurada con 80 preguntas de selección única: 50 de matemáticas y 30 de razonamiento verbal. El Departamento de Admisión y Registro (DAR) proporcionó la nota de la educación diversificada y el promedio simple del primer semestre del 2020.

La ponderación de la nota de admisión en el TEC corresponde en un 60% a los componentes de matemática y verbal de la PAA y en un 40% a la nota de la educación diversificada informada por el colegio al DAR. En el 2019, esta nota se calculó como un promedio de las calificaciones obtenidas en español, matemática, estudios sociales, educación cívica, lengua extranjera (inglés o francés) y ciencia (biología, química o física) en el último ciclo de la educación media.

Procedimiento

A partir de la identificación del conjunto de habilidades matemáticas y verbales que requieren los estudiantes de primer ingreso del TEC en el

estudio exploratorio-descriptivo realizado en el 2009 se logró rediseñar la tabla de especificaciones de la PAA. Desde ese año se han recolectado evidencias de los ítems, las tareas cognitivas y las categorías de habilidades para lograr una mejor precisión en la medición de los contenidos de acuerdo con el siguiente procedimiento:

1. Determinación de las categorías de razonamiento matemático y razonamiento verbal, con sus correspondientes tareas cognitivas, que son medidas en la prueba.
2. Diseño de la tabla de especificaciones como una matriz con los contenidos en las filas y los niveles de complejidad en las columnas.
3. Asignación de la proporción de ítems por cada habilidad, según el valor predictivo con respecto al criterio evaluado, la pertinencia con el perfil de entrada de los estudiantes de primer ingreso del TEC y su grado de relación con el constructo de la aptitud académica.
4. Una vez elaborada la tabla de especificaciones se procede con la construcción de la prueba.
5. Selección de los ítems por parte del equipo técnico, quien evalúa el grado en que cada ítem representa una categoría de razonamiento y sus tareas cognitivas.

Análisis de datos

En el análisis descriptivo de los datos fue utilizado el programa estadístico IBM SPSS Statistics 19 (IBM Corporation, 2010).

Con el propósito de obtener evidencias del grado de relación entre los indicadores de las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de razonamiento general, se realizó un análisis factorial confirmatorio (AFC), conforme al siguiente pro-

cedimiento:

1. Identificación del modelo: La regla de los grados de libertad (gl) es la más frecuente y se calcula con la fórmula $k * (k + 1) / 2$ (Cupani, 2012), donde k son los indicadores. Según el número de indicadores matemáticos (3) y verbales (2) incluidos en la ecuación, se tiene un total de 15 elementos conocidos en la matriz de covarianza y se ha especificado un total de cinco parámetros. Al restar este número, el modelo queda con 10 grados de libertad; por lo tanto, es un modelo sobreidentificado, lo que asegura que el modelo sea tan generalizable como sea posible.

2. Evaluar la identificación del modelo: Una recomendación para tratar de que todos los parámetros estén identificados es utilizar al menos tres indicadores por variable latente e igualar la métrica de cada variable latente con uno de sus indicadores, lo que se consigue fijando arbitrariamente al valor 1 el peso de uno de los indicadores (Cea-D'Ancona, 2002). En el caso de este modelo se incluyeron cinco indicadores para el constructo de *razonamiento general*.

3. Seleccionar las medidas, con base en la recolección, descripción y depuración de datos: las mediciones se basaron en las puntuaciones de los ítems de la PAA del 2019, las cuales se agruparon según la categorización de las habilidades matemáticas y verbales descritas en la tabla de especificaciones.

4. Estimar el modelo con el método de máxima verosimilitud, ya que es la técnica más empleada en estos análisis.

5. Comprobar el ajuste del modelo de acuerdo con los índices de ajuste global, de parsimonia y comparativos.

Una vez concluida la etapa de la estimación del modelo AFC, se interpretan las relaciones de los indicadores con la variable latente medida en

la PAA con la finalidad de comprobar el grado de relación entre ambos factores. Para una descripción más detallada de este procedimiento se recomienda consultar a Escobedo, Hernández, Estebané y Martínez (2016), quienes apuntan que el software EQS (abreviatura de *Equations*), creado por Bentler, ha tenido gran aceptación entre los investigadores que trabajan con modelos de ecuaciones estructurales. En suma, con este procedimiento se pretendió comprobar el siguiente modelo de un factor de razonamiento general, como se muestra en la Figura 1.

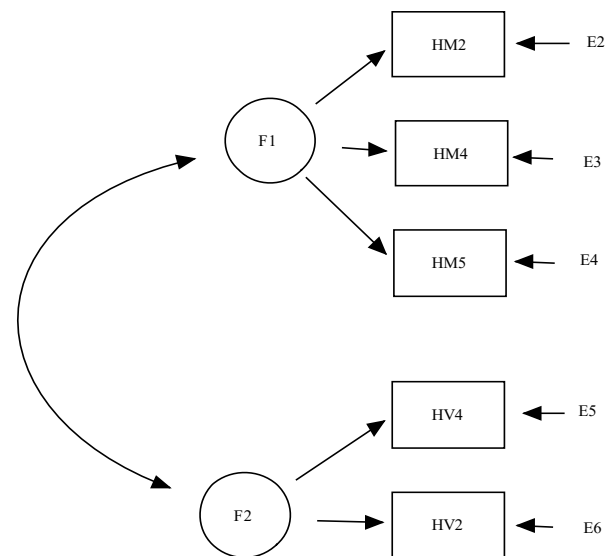


Figura 1

Modelo de medida de la Prueba de Aptitud Académica del 2019.

El análisis predictivo se realizó con base en un modelo de regresión múltiple lineal usando el programa IBM SPSS 19 (IBM Corporation, 2010), donde se incluyeron las variables predictoras de la nota de admisión matemática, verbal y nota de presentación de la educación diversificada, en tanto que el criterio fue promedio del primer semestre del 2020 de los estudiantes de primer ingreso al TEC. El modelo que se busca probar es el siguiente:

$$Y = \beta_0 + \beta_1 X_{\text{matemática}} + \beta_2 X_{\text{verbal}} + \beta_3 X_{\text{ciclo diversificado}} + \epsilon_i$$

Y: es la variable respuesta, el promedio del primer periodo universitario del 2020.

β_0 : es la intersección que determina el valor de Y cuando X es cero.

β : es la pendiente que determina la cantidad en la que cambia Y cuando X se incrementa en una unidad.

X: son las variables explicativas del modelo: notas de matemática, verbal y educación diversificada.

ϵ_i : errores aleatorios.

De acuerdo con Carrasquilla-Batista et al. (2016) este modelo es usado cuando la variable dependiente depende linealmente de cada una de las variables explicativas y un regresor no basta para explicar suficientemente la variabilidad de la variable respuesta.

Resultados

Validez de contenido

Como se indicó en la sección metodológica, el diseño de la tabla de especificaciones de la PAA se ha ajustado año tras año a partir de las evidencias psicométricas de los ítems, las tareas cognitivas y las categorías de habilidades. Como se muestra en la Tabla 1, para el 2019 se seleccionaron tres *habilidades matemáticas*, principalmente por sus aportes predictivos durante los últimos años.

En cuanto al *razonamiento deductivo*, medido tradicionalmente con silogismos, resulta necesario aclarar que antes de 2016 pertenecía al componente verbal de la prueba. No obstante, por su alta correlación con las habilidades matemáticas, se creó una nueva categoría que combinó los silogismos con ítems de *razonamiento inductivo*. A partir de este cambio, entre otros, en el 2019 se midieron solo dos habilidades verbales como se

Tabla 1

Especificaciones de las habilidades matemáticas medidas en la PAA del 2019.

Habilidades matemáticas	Peso	Tareas cognitivas
Razonamiento deductivo e inductivo	32%	A) Realizar inferencias o deducciones a partir de cierta información dada. B) Identificar premisas que permitan obtener una conclusión dada. C) Obtener la forma general de una sucesión o algunos de sus términos.
Resolución de problemas	34%	A) Realizar representaciones gráficas para resolver un problema. B) Interpretar la información dada en un problema para plantear la solución. C) Aplicar reglas básicas de conteo. D) Aplicar definiciones o teoremas nuevos en la resolución de problemas.
Razonamiento con figuras	34%	A) Identificar patrones geométricos. B) Analizar secuencias de figuras. C) Reconocer dos o más elementos que conforman un todo. D) Reconocer un objeto por medio de sus partes y combinar características de varios objetos para realizar representaciones.

resume en la Tabla 2.

Las categorías de *razonamiento verbal* también han mostrado una relación positiva estadísticamente significativa con el desempeño en el primer año universitario, a pesar de ser las más cambiantes en el diseño de la tabla de especificaciones. Por ejemplo, el razonamiento inferencial se medía en dos subcategorías, una con textos cortos y otra con textos largos, pero en 2014 se integró en una sola categoría. También, por unos años se incluyó la categoría de *razonamiento dis-*

Tabla 2

Especificaciones de las habilidades verbales medidas en la PAA del 2019.

Habilidades verbales	Peso	Tareas cognitivas
Razonamiento inferencial	50%	A) Establecer relaciones entre ideas de un texto. B) Sintetizar información de distintos textos. C) Comprender instrucciones. D) Comparar ideas de textos distintos. E) Extraer conclusiones.
Razonamiento semántico	50%	A) Reconocer las relaciones semánticas entre palabras en un texto. B) Identificar el significado de las palabras en contexto.

cursivo, sin embargo, por no mostrar valor predictivo fue excluida a partir de 2018.

Finalmente es necesario destacar que, en el proceso de juzgamiento de los ítems, es crucial el cumplimiento de los estándares de escritura, el criterio de expertos en el análisis de sensibilidad de los ítems para asegurar que no haya contenidos irrelevantes que favorezcan o perjudiquen a determinado grupo de examinados, la aplicación piloto de ítems experimentales y análisis psicométricos como el de funcionamiento diferencial de los ítems, dificultad y discriminación. Estos análisis permiten garantizar los estándares psicométricos de los ítems antes de ser seleccionados para el banco, de manera que no se altera el modelo de medida de la PAA al cumplir con los parámetros establecidos en el diseño de cada año.

Modelo logístico de dos parámetros

Con el propósito de ampliar las evidencias de la calidad psicométrica de los ítems de la PAA,

se estimó el modelo logístico de dos parámetros de la Teoría de Respuesta al Ítem (TRI) para cada uno de los factores de la PAA, con el uso de STATA 14 (StataCorp, 2015). También se utilizó la función de información (FI) para determinar en qué niveles del constructo se proporcionaban mediciones más precisas.

La medida de adecuación muestral KMO para el componente de *matemática* fue de .97 y la prueba de esfericidad de Bartlett arrojó un valor de 117741.9 ($gl = 946$; $p < .001$) y para el componente *verbal* fue de .93 y en la prueba de esfericidad de Bartlett se obtuvo un valor de 36518.5 ($gl = 300$; $p < .001$). Estos valores confirmaron la posibilidad de aplicar un análisis de componentes principales a la matriz de datos y tener una aproximación aceptable al supuesto de unidimensionalidad requerido para los modelos de la TRI, como se muestra en los gráficos de sedimentación de la Figura 2.

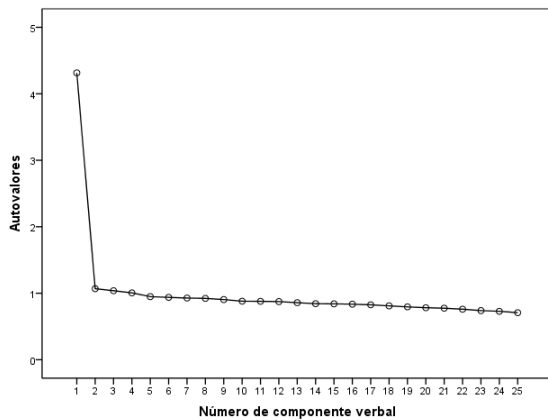
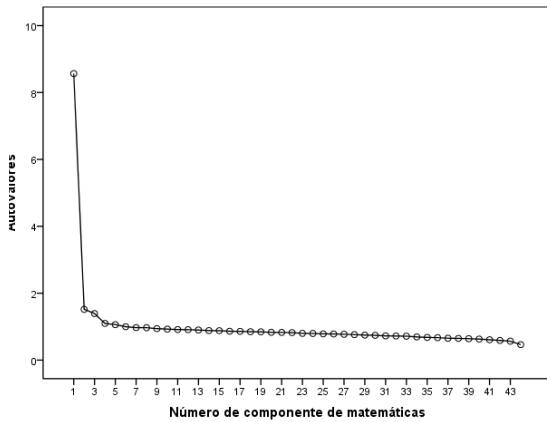
En cuanto a la función de información de cada componente, que se puede observar en la Figura 3, ambos aportan una medición más precisa en los niveles intermedios de θ , lo cual es consistente con el diseño de la PAA. En el caso del componente *verbal*, se presenta un ligero aumento en la precisión en niveles fáciles de habilidad.

Además de la FI relacionada con la confiabilidad de la medida, también se calculó el coeficiente alfa de Cronbach, que resultó en .92. Para completar el análisis de los ítems, se presentan los parámetros de dificultad y discriminación en las Tablas 3 y 4.

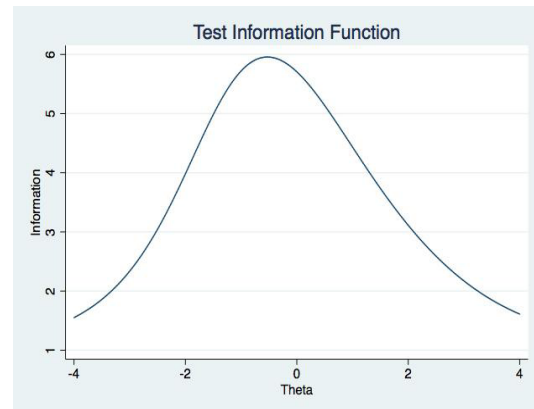
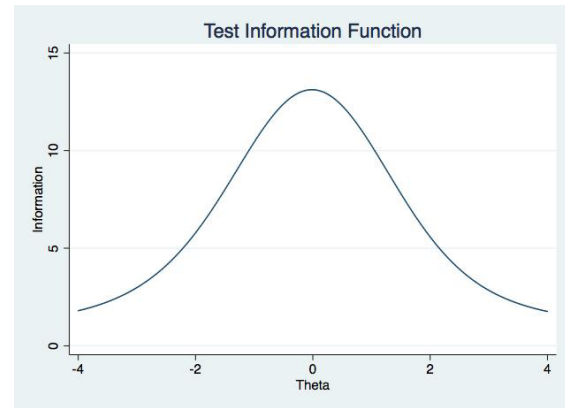
A partir de estos parámetros se puede observar que solo dos ítems de *matemática* (43 y 44) no discriminan entre los examinados de alto y bajo nivel de habilidad; en tanto que el ítem 74 de *razonamiento verbal* resultó con valores extremos en la dificultad y el error estándar. Estos resultados del modelo TRI complementan los análisis realizados con la TCT y demuestran el

Figura 2

Gráficos de sedimentación de los factores de matemática y verbal de la PAA 2019.

**Figura 3**

Función de información del factor de matemática y de razonamiento verbal de la PAA 2019.



cumplimiento de los parámetros de dificultad y discriminación por los ítems de la PAA del 2019, con excepción de los tres mencionados.

Evidencias de validez de constructo

En este análisis multivariado se utilizó el paquete estadístico EQS 6.2 para Windows ([Multivariate Software Inc, 2012](#)) para estimar el modelo SEM con una distribución normal. De acuerdo con [Pardo y Ruiz \(2002\)](#), con estos modelos se busca ajustar las covarianzas entre las variables y minimizar la diferencia entre las covarianzas ob-

servadas de la muestra y las covarianzas estimadas por el modelo estructural.

Los indicadores del factor *matemático* (F1) del modelo de medida de la PAA son los que más explican la varianza en el constructo: 67% razonamiento deductivo e inductivo (HM2), 74% resolución de problemas (HM4) y 66% razonamiento con figuras (HM5). Los indicadores del factor de *razonamiento verbal* (F2) denominados razonamiento inferencial (HV2) y semántico (HV4) también evidencian un buen ajuste con el modelo, con 61% y 62% de varianza explicada. En general, cuanto más alta sea la contribución de los indicadores, más alta es la correlación del

Tabla 3

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm3 Discrim	1.60	.03	46.93	0.001	1.53 1.67
Diff	-.74	.02	-44.45	0.001	-.77 -.71
rm4 Discrim	1.25	.03	44.07	0.001	1.20 1.31
Diff	-.78	.02	-39.14	0.001	-.82 -.74
rm5 Discrim	.93	.02	42.47	0.001	.89 .98
Diff	.22	.02	10.68	0.001	.18 .26
rm6 Discrim	1.17	.03	44.88	0.001	1.12 1.23
Diff	-.51	.02	-27.83	0.001	-.55 -.48
rm7 Discrim	1.42	.03	47.87	0.001	1.34 1.48
Diff	-.47	.02	-29.38	0.001	-.50 -.44
rm8 Discrim	1.12	.02	44.96	0.001	1.07 1.17
Diff	-.33	.02	-18.50	0.001	-.37 -.30
rm9 Discrim	1.27	.03	46.58	0.001	1.21 1.32
Diff	-.43	.02	-25.41	0.001	-.47 -.40
rm10 Discrim	.67	.02	32.80	0.001	.63 .71
Diff	-.68	.03	-21.87	0.001	-.74 -.62
rm11 Discrim	.83	.02	39.65	0.001	.79 .87
Diff	.29	.02	12.58	0.001	.24 .33
rm12 Discrim	.54	.02	29.28	0.001	.51 .58
Diff	.38	.03	11.30	0.001	.31 .44
rm13 Discrim	.94	.02	42.37	0.001	.89 .98
Diff	.67	.02	27.73	0.001	.62 .72
rm14 Discrim	.82	.02	39.43	0.001	.78 .86
Diff	.37	.02	15.81	0.001	.33 .42
rm15 Discrim	.41	.02	22.87	0.001	.37 .44
Diff	1.26	.07	19.11	0.001	1.14 1.39
rm16 Discrim	1.01	.02	40.29	0.001	.97 1.05
Diff	-.85	.02	-35.12	0.001	-.90 -.80
rm19 Discrim	1.25	.03	46.91	0.001	1.20 1.30
Diff	-.33	.02	-19.81	0.001	-.37 -.30
rm20 Discrim	.95	.02	42.85	0.001	.91 1.00
Diff	.10	.02	5.01	0.001	.06 .14
rm21 Discrim	1.11	.02	46.14	0.001	1.07 1.16
Diff	.14	.02	7.78	0.001	.10 .17
rm22 Discrim	1.03	.02	44.37	0.001	.99 1.08
Diff	-.04	.02	-1.97	0.049	-.07 -.00
rm23 Discrim	.91	.02	40.10	0.001	.86 .95
Diff	-.50	.02	-22.50	0.001	-.54 -.45
rm24 Discrim	1.52	.03	49.32	0.001	1.46 1.58
Diff	-.39	.01	-25.87	0.001	-.42 -.36

Tabla 3 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm25 Discrim	1.58	.03	51.27	0.001	1.52 1.64
Diff	-.08	.01	-5.63	0.001	-.11 -.05
rm26 Discrim	1.30	.03	48.71	0.001	1.25 1.35
Diff	-.05	.02	-3.02	0.003	-.08 -.02
rm27 Discrim	.90	.02	39.87	0.001	.86 .95
Diff	-.48	.02	-21.80	0.001	-.53 -.44
rm28 Discrim	.27	.02	16.00	0.001	.24 .31
Diff	.99	.08	11.78	0.001	.83 1.16
rm29 Discrim	1.35	.03	49.84	0.001	1.29 1.40
Diff	.19	.02	11.87	0.001	.16 .22
rm30 Discrim	1.10	.02	46.09	0.001	1.05 1.14
Diff	.31	.02	16.26	0.001	.27 .34
rm31 Discrim	.89	.02	40.62	0.001	.85 .94
Diff	-.16	.02	-7.87	0.001	-.20 -.12
rm32 Discrim	1.66	.03	52.46	0.001	1.60 1.72
Diff	.22	.01	15.28	0.001	.19 .25
rm33 Discrim	1.44	.03	50.78	0.001	1.38 1.49
Diff	.22	.02	13.78	0.001	.19 .25
rm36 Discrim	1.40	.03	40.62	0.001	1.33 1.46
Diff	-1.24	.02	-49.94	0.001	-1.29 -1.20
rm37 Discrim	.76	.02	35.91	0.001	.72 .80
Diff	-.58	.03	-22.01	0.001	-.64 -.53
rm38 Discrim	.83	.02	38.98	0.001	.79 .87
Diff	-.17	.02	-7.66	0.001	-.21 -.12
rm39 Discrim	.89	.02	37.88	0.001	.84 .94
Diff	-.89	.03	-32.56	0.001	-.94 -.83
rm40 Discrim	.65	.02	33.07	0.001	.61 .69
Diff	-.22	.03	-8.02	0.001	-.27 -.16
rm41 Discrim	.98	.02	43.45	0.001	.94 1.03
Diff	.09	.02	4.62	0.001	.05 .13
rm42 Discrim	.70	.02	34.14	0.001	.66 .74
Diff	-.53	.03	-19.09	0.001	-.59 -.48
rm43 Discrim	1.16	.02	46.86	0.001	1.11 1.21
Diff	.01	.02	0.32	0.750	-.03 .04
rm44 Discrim	.95	.02	42.65	0.001	.91 1.00
Diff	.02	.02	0.77	0.441	-.02 .05
rm45 Discrim	.59	.02	30.17	0.001	.55 .62
Diff	1.46	.05	27.66	0.001	1.36 1.56
rm46 Discrim	1.38	.03	50.26	0.001	1.32 1.43
Diff	.39	.02	22.95	0.001	.35 .42

Tabla 3 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm47 Discrim	.60	.02	30.22	0.001	.56 .63
Diff	1.64	.06	28.82	0.001	1.52 1.75
rm48 Discrim	1.78	.03	52.37	0.001	1.71 1.84
Diff	.60	.02	37.86	0.001	.57 .63
rm49 Discrim	1.42	.03	47.34	0.001	1.36 1.48
Diff	1.24	.02	51.85	0.001	1.19 1.29
rm50 Discrim	1.65	.03	47.71	0.001	1.58 1.72
Diff	1.26	.02	56.86	0.001	1.21 1.30

indicador con la aptitud académica. Como se observa en la Figura 4, el gráfico de la solución estandarizada muestra el grado de asociación y los errores de medición.

Particularmente, el diagrama facilita la interpretación de los resultados al basarse en unidades estandarizadas que favorecen la comparabilidad entre los coeficientes de correlación. Según el análisis, el indicador matemático con el mayor grado de asociación a la variable latente exógena fue *resolución de problemas* (HM4), fijado con valor 1, mientras que el indicador de *razonamiento semántico* (HV4), mostró la más alta correlación. Este modelo incluye los indicadores observables, los errores que afectan a las mediciones y las relaciones entre el constructo, específicamente, se consideran:

- Variables observadas exógenas (Lambda x): HM2, HM4, HM5, HV2 y HV4.

- Variable latente exógena (ξ): razonamiento general.

- Errores de medición en variables observadas exógenas (δ): De .510 a .621.

- Coeficientes de correlación entre la variable latente con sus respectivos indicadores (λ): Los valores oscilan entre .78 y .86.

Este modelo de ecuaciones estructurales

proporcionó evidencias de la existencia de factores comunes que explicaban la variabilidad de un conjunto de indicadores (HMx y HVx) y la de un factor general de *razonamiento* que explicaba la variabilidad de estos factores comunes, altamente correlacionados (.79) del modelo de medida de la PAA. Finalmente, se comprobó un buen ajuste del modelo de medida de la PAA, tal como se muestra en la Tabla 5, cuyos valores se analizan de acuerdo con los criterios descritos por Ruiz, Pardo y San Martín (2010).

A partir de estos datos, se puede afirmar que existe un ajuste óptimo entre el rasgo latente de *aptitud académica* y los indicadores, al obtenerse un valor aceptable en el índice RMSEA (inferior a .08), y también valores aceptables de ajuste global del modelo propuesto (GFI) de ajuste incremental (NFI) y de ajuste de parsimonia (AGFI). De acuerdo con Cupani (2012), los valores superiores a .95 sugieren un ajuste óptimo y, en lo que respecta al chi cuadrado, explica que debe ser no significativo para indicar un buen ajuste; sin embargo, al ser un estadístico muy sensible al tamaño muestral debe interpretarse con precaución. Por ello, usualmente se interpreta que la razón del χ^2 sobre los grados de libertad con valores inferiores a 2 indica un buen ajuste. Por otra parte, también debe resaltarse que todos los indicadores reflectivos de los componentes *matemático* y *verbal*

Tabla 4

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente *verbal* de la PAA 2019.

Verbal	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rv53 Discrim	1.34	.03	39.16	0.001	1.27 1.41
Diff	-1.33	.03	-48.34	0.001	-1.38 -1.27
rv54 Discrim	1.09	.03	41.00	0.001	1.03 1.14
Diff	-.65	.02	-30.42	0.001	-.69 -.61
rv55 Discrim	1.08	.03	42.07	0.001	1.03 1.13
Diff	-.09	.02	-4.88	0.001	-.12 -.05
rv56 Discrim	.82	.02	34.66	0.001	.78 .87
Diff	-1.09	.03	-32.42	0.001	-1.16 -1.03
rv57 Discrim	.91	.02	38.28	0.001	.86 .96
Diff	-.48	.02	-21.41	0.001	-.53 -.44
rv58 Discrim	.61	.02	29.81	0.001	.57 .65
Diff	-.38	.03	-12.67	0.001	-.44 -.32
rv59 Discrim	.75	.02	34.45	0.001	.71 .79
Diff	-.34	.02	-13.58	0.001	-.39 -.29
rv60 Discrim	1.21	.03	43.82	0.001	1.16 1.27
Diff	-.15	.02	-9.03	0.001	-.19 -.12
rv61 Discrim	.94	.02	39.02	0.001	.89 .99
Diff	.43	.02	19.87	0.001	.39 .47
rv62 Discrim	.81	.02	35.07	0.001	.76 .85
Diff	.83	.03	27.94	0.001	.77 .88
rv63 Discrim	.85	.03	33.53	0.001	.80 .90
Diff	1.50	.04	35.65	0.001	1.42 1.58
rv64 Discrim	.82	.02	32.97	0.001	.77 .87
Diff	1.48	.04	34.52	0.001	1.40 1.57
rv65 Discrim	1.58	.04	40.76	0.001	1.50 1.65
Diff	-1.18	.02	-52.24	0.001	-1.22 -1.13
rv69 Discrim	1.20	.03	41.29	0.001	1.14 1.26
Diff	-.89	.02	-39.41	0.001	-.93 -.84
rv70 Discrim	1.30	.03	43.82	0.001	1.25 1.36
Diff	-.56	.02	-30.92	0.001	-.59 -.52
rv71 Discrim	.58	.02	28.57	0.001	.54 .62
Diff	-.45	.03	-13.83	0.001	-.51 -.39
rv72 Discrim	.86	.02	37.23	0.001	.81 .90
Diff	-.43	.02	-18.81	0.001	-.48 -.39
rv73 Discrim	.67	.02	31.76	0.001	.62 .71
Diff	-.40	.03	-14.35	0.001	-.46 -.35
rv74 Discrim	.04	.02	2.00	0.046	.00 .08
Diff	24.07	12.05	2.00	0.046	.46 47.69
rv75 Discrim	.84	.02	37.00	0.001	.79 .88
Diff	-.11	.02	-4.85	0.001	-.15 -.06

Tabla 4 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente *verbal* de la PAA 2019.

Verbal	Coefficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rv76Discrim	1.23	.03	43.91	0.001	1.17 1.28
Diff	-.22	.02	-12.88	0.001	-.25 -.18
rv77Discrim	.96	.02	38.47	0.001	.91 1.01
Diff	.79	.03	31.37	0.001	.74 .84
rv78Discrim	.76	.02	32.62	0.001	.71 .80
Diff	1.26	.04	31.51	0.001	1.18 1.34
rv79Discrim	.57	.03	22.07	0.001	.52 .62
Diff	3.06	.13	23.77	0.001	2.81 3.31
rv80Discrim	.90	.02	37.94	0.001	.86 .95
Diff	.60	.02	25.17	0.001	.56 .65

resultaron coeficientes significativos; lo cual resulta muy positivo para efectos del modelamiento de la PAA y la confirmación de la hipótesis de que existe una relación entre el modelo de medida y el hipotetizado.

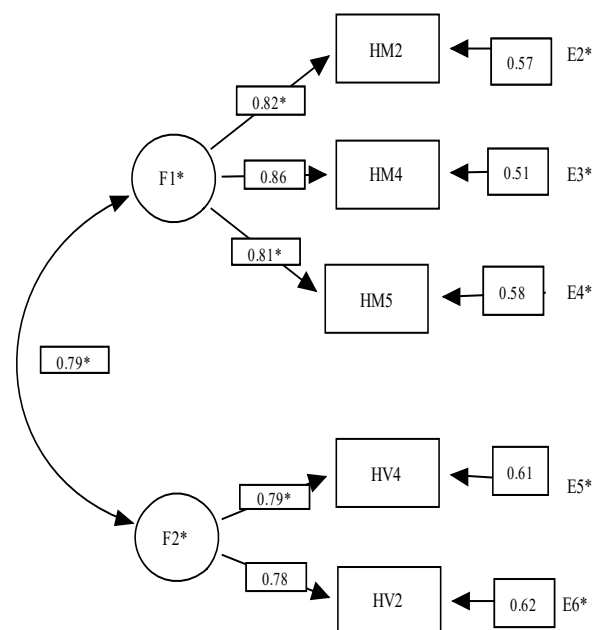
Evidencias de validez predictiva

El modelo de regresión lineal múltiple se ejecutó con el software SPSS Statistics versión 19 (IBM Corporation, 2010), con los examinados que realizaron la PAA en el 2019 y se matricularon en el primer semestre del curso lectivo del 2020. Específicamente, en el modelo clásico se incluyeron 1769 estudiantes de primer ingreso, cuyos resultados se presentan en la Tabla 6.

Con respecto a los supuestos del modelo de regresión lineal múltiple, no se encontraron correlaciones altas entre estos predictores, cuyos coeficientes Pearson oscilaron entre .231 y .303, por lo tanto, estas estimaciones no están afectadas por la colinealidad. El supuesto de normalidad no se cumplió, según la prueba de Kolmogorov-Smirnov; como tampoco el supuesto de homocedasticidad, comprobado con el estadístico Breusch-Pa-

gan; por tal razón, se estimó un segundo modelo de regresión robusta con R Core Team (2018), como se muestra en la Tabla 7.

Estas son las evidencias predictivas más recientes del modelo de medida de la prueba de admisión del TEC. Para el análisis correspondien-

**Figura 4**

EQS 6 modelo PAA 2019_Chi Sq.= 236.00 $p < 0.001$ CFI = 0.99 RMSEA = 0.06.

Tabla 5
Estadísticos de bondad de ajuste del modelo SEM de la PAA 2019.

Estadístico	Abreviatura	Criterio	Valores del modelo
Chi-cuadrado	χ^2	Significación > .05	.000
Razón chi cuadrado	χ^2/df		2.370
Bentler-Bonett Normed Fit Index	NFI	≥ .95	.994
Bentler-Bonett Non-Normed Fit Index	NNFI	≥ .95	.986
McDonald's fit index	MFI	≥ .90	.993
Bollen's fit index	IFI	≥ .90	.994
Índice de bondad de ajuste comparativo	CFI	≥ .95	.994
Joreskog-Sorbom's GFI	GFI	≥ .95	.994
Joreskog-Sorbom's AGFI	AGFI	≥ .95	.978
Raíz del residuo cuadrático promedio	RMR	Próximo a cero	.136
Raíz del residuo cuadrático promedio de aproximación	RMSEA	< .08	.060
90% Intervalo de confianza de RMSEA			(.053; .066)

te de cada uno de los predictores, se registraron los resultados de los coeficientes de regresión no estandarizados y el nivel de significancia ($\alpha = .05$). Estos coeficientes indican el cambio que corresponde a la nota del promedio simple por cada unidad de cambio en la puntuación de cada componente, todas medidas en una escala 0-100. De manera que la nota del ciclo de educación diversificada (EDciclo100) es la que muestra una mayor contribución en ambos modelos. Esto significa que el incremento en el promedio simple de .58 puntos (vg. modelo robusto) se produce independientemente de la puntuación de los otros dos componentes *matemático* y *verbal*.

A partir de los coeficientes estandarizados se puede comparar el poder explicativo de cada predictor en la ecuación. Al ser directamente comparables, se observa que en ambos modelos la nota de matemática y educación diversificada resulta con el más alto coeficiente; en tanto que el razonamiento verbal no resulta tan relevante; pero sí ha mostrado una relación estadísticamente significativa con el promedio simple en el primer año desde el 2016.

La proporción de la variabilidad de los promedios simples explicada por los tres predictores, esto es, el coeficiente de determinación, fue de .095; en tanto que en años anteriores para el

Tabla 6
Modelo de regresión clásico de los componentes de la nota de admisión con el promedio simple del primer semestre del 2020.

Modelo	Coeficientes no estandarizados		Coeficientes Beta	t	Sig.
	B	Error típ.			
(Constante)	-13.656	7.743		-1.764	.078
EDciclo100	.749	.080	.211	9.325	.000
Mate100	.360	.038	.228	9.526	.000
Verbal100	-.079	.037	-.051	-2.119	.034

Nota. Modelo 1 promedio simple del primer semestre del 2020.

Tabla 7

Modelo de regresión robusta de los componentes de la nota de admisión con el promedio simple primer semestre de 2020.

Predictores	Coefficientes	Error estándar	Beta	Error estándar	Estadístico	<i>p</i>
(Intercepto)	11.22	5.55	.22	.02	13.85	.043
EDciclo100	0.58	0.06	.17	.02	9.92	< .001
Mate100	0.24	0.03	.15	.02	9.12	< .001
Verbal100	-0.02	0.02	-.02	.02	-0.98	.327

Nota. Promedio simple del primer semestre de 2020.

mismo periodo fue de .174 y .124 con la PAA aplicada en el 2018 y 2017 respectivamente. Esta disminución en la varianza explicada, como en el valor predictivo del componente *verbal*, se presume que se relaciona con el cambio radical de la modalidad presencial hacia la virtualidad de todos los cursos universitarios durante el primer periodo del 2020, a causa de los efectos de la pandemia en el país. Por otra parte, tales coeficientes podrían mejorar si se incluyen otras variables independientes, tales como tipo de financiamiento del colegio, sexo y provincia de procedencia entre otros. Sin embargo, para los propósitos de esta investigación lo que interesa es encontrar evidencias del grado de validez predictiva de estos tres componentes de la nota de admisión con respecto al rendimiento de los estudiantes en su primer año de ingreso.

Discusión

La validación debe ser un proceso permanente: cuantas más evidencias de validez se obtengan de una prueba, más sólido es el diseño y más razonables los usos y las interpretaciones de sus resultados, especialmente por sus altas consecuencias sociales, como es el caso de la PAA para el TEC. Como lo resaltan [Muñiz y Fonseca-Pedrero \(2019\)](#), siempre habrá que aportar evidencias empíricas de la fiabilidad y la validez, para

garantizar que los instrumentos de medida evalúan de forma objetiva y rigurosa.

Por ello, el primer paso para validar el modelo de medida es analizar las evidencias de validez de contenido a partir de una herramienta básica: la tabla de especificaciones. Esta permite a los constructores de ítems tener una guía básica sobre el tipo de tareas que deben medir los ítems; así como garantizar la distribución estándar del nivel de dificultad en la prueba: 25% fáciles, 50% intermedias y 25% difíciles. En este sentido, [Sireci y Faulkner-Bond \(2014\)](#) consideran que las evidencias de validez de contenido son críticas para hacer inferencias acerca de los examinados con respecto al dominio de interés. Además, en esta etapa es muy importante el criterio de los expertos para analizar el grado de representatividad de los ítems de cada indicador o categoría de razonamiento y sus tareas cognitivas.

Una vez que se obtienen evidencias de contenido, como la proporcionada por la tabla de especificaciones, es fundamental comprobar el grado de asociación entre las variables observadas exógenas (x) de los indicadores matemáticos y verbales por una parte, y el constructo de aptitud académica, la variable latente exógena (ξ) por otra parte, para realizar los ajustes necesarios. Precisamente, cada año el análisis de ese conjunto de indicadores ha permitido tomar decisiones para mejorar la precisión de la medida, tales como la eliminación de algunos indicadores (analogías,

antónimos, análisis discursivo) y la integración de algunas habilidades, como el razonamiento inferencial con textos cortos y largos; o bien, realizar ajustes en las tareas cognitivas como en los silogismos, el conteo o la resolución de problemas con operaciones aritméticas, entre otros.

De esta manera, es muy importante apuntar que se estima un modelo SEM entre otros posibles para medir el razonamiento general. La evidencia obtenida de este modelo cumple satisfactoriamente con los distintos criterios de ajuste y significación estadística, por lo que permite afirmar que existe una relación significativa entre los indicadores matemáticos y verbales de la prueba y el rasgo latente de razonamiento general. Evidentemente, este es un rasgo complejo que, al ser medido por cinco indicadores, resulta un posible modelo entre otros. Al respecto, [Donoso-Díaz y Hawes-Barríos \(2001\)](#) señalan que el reducido ámbito de las dimensiones de inteligencia de las pruebas que consideran solo la aptitud verbal y matemática; si bien es relevante, es insuficiente para predecir acertadamente el éxito de un estudiante ante un proceso de formación, que en la actualidad demanda otras dimensiones de inteligencia (espacial, kinésica, interpersonal, intrapersonal, etc).

Finalmente, resulta de suma importancia la obtención de evidencias del valor predictivo de los componentes de la nota de admisión, puesto que el propósito de la PAA es seleccionar a los candidatos con mayor probabilidad de éxito en el TEC. De manera que las evidencias de asociación de las notas de matemática, verbal y de la educación diversificada con el rendimiento de los estudiantes en el primer año universitario permite afirmar que sí aportan evidencias predictivas. Como destacan [Guglietta y Delgado-Álvarez \(2010\)](#), las pruebas estandarizadas de habilidad verbal y numérica se comportan como predictores válidos del rendimiento. Por supuesto que

este modelo resulta limitado por considerar únicamente los predictores de la nota de admisión. Al respecto, existen múltiples estudios de factores asociados al rendimiento en el primer año, como lo señala [García-Domínguez \(2016\)](#), por lo que resultaría interesante incorporar otros datos de naturaleza personal y socioeconómica como variables que pudieran incidir en el rendimiento de los estudiantes.

Una limitación de estos análisis predictivos es la variabilidad de criterios para calificar los cursos universitarios y de las asignaturas de la educación diversificada. Como señalan [Contreras et al. \(2009\)](#), estos criterios reflejan distintas políticas de asignación de notas, niveles de exigencia y, eventualmente, ‘inflación’ de notas. No obstante, la inclusión de la nota de educación diversificada en el modelo de regresión se debe a que resulta un buen predictor del rendimiento de los estudiantes durante su primer año en el TEC.

Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Autor.
- Backhoff, E., Tirado, F., & Larrazolo, N. (2001). Ponderación diferencial de reactivos para mejorar la validez de una prueba de ingreso a la universidad. *Revista Electrónica de Investigación Educativa*, 3(1), 1-10. Recuperado de <https://redie.uabc.mx/redie>
- Bleyaert, B. (2010). ACT and college success. *Education Partnerships*. ERIC N° ED537914. Recuperado de <https://eric.ed.gov/?id=ED537914>
- Bringas-Benavides, M. R., & Pérez-Mejía, J. (2014). El examen de ingreso al nivel superior. ¿Admisión o de-

- cepción? *Ra Ximhai*, 10(5), 103-114. Recuperado de <https://www.redalyc.org/pdf/461/46132134007.pdf>
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde-Cerdas, J., & Guerrero-Barrantes, M. (2016). Regresión lineal simple y múltiple: Aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Tecnología en Marcha*, 29(8), 33-45. doi: 10.18845/tm.v29i8.2983
- Cea-D'Ancona, M. A. (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid, España: Síntesis.
- Cliff, A., & Montero, E. (2010). El balance entre excelencia y equidad en pruebas de admisión: Contribuciones de experiencias en Sudáfrica y Costa Rica. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 7-28. Recuperado de <https://revistas.uam.es/index.php/rie>
- College Board. (2014). *Guía de estudio para presentar la Prueba de Aptitud Académica*. Recuperado de <https://latam.collegeboard.org>
- Cupani, M. (2012). Análisis de ecuaciones estructurales: Conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*, 2(1), 186-199. Recuperado de <https://revistas.unc.edu.ar/index.php/tesis>
- Contreras, D., Bravo, D., & Sanhueza, C. (2001). PAA, ¿una prueba de inteligencia? *Revista Perspectivas*, 4(2), 233-247. Recuperado de <http://www.dii.uchile.cl/~revista>
- Contreras, D., Gallegos, S., & Meneses, F. (2009). Determinantes de desempeño universitario: ¿Importa la habilidad relativa? *Calidad en la Educación*, 30, 18-48. doi: 10.31619/caledu.n30.172
- Donoso-Díaz, S., & Hawes-Barrios, G. (2001). El constructo de inteligencia del sistema de selección de alumnos a las universidades chilenas. Discusión de sus fundamentos y lineamientos para el cambio. *Revista Cubana de Educación Superior*, 21(3), 57-80.
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36. Recuperado de <http://www.humanas.unal.edu.co/psicometria/revista/volumenes/volumen-6>
- Escobedo, M. T., Hernández, J. A., Estebané, V., & Martínez, G. (2016). Modelos de ecuaciones estructurales: Características, fases, construcción, aplicación y resultados. *Ciencia & Trabajo*, 18(55), 16-22. Recuperado de <https://scielo.conicyt.cl/pdf/cyt/v18n55/art04.pdf>
- García-Domínguez, L. A. (2016). Pruebas de selección como predictores del rendimiento académico de estudiantes de Medicina. *Investigación en Educación Médica*, 5(18), 88-92. Recuperado de <http://riem.facmed.unam.mx>
- Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the SAT*. College Board Research (Reporte N° 2005-11). Recuperado de <https://files.eric.ed.gov>
- Guglietta, L., & Delgado-Álvarez, C. (2010). Validez de constructo de un modelo de admisión a postgrado. Un análisis de ruta. *Revista Galego-Portuguesa de Psicoloxía e Educación*, 18(1), 227-237. Recuperado de <https://ruc.udc.es/dspace/handle/2183/6562>
- IBM Corporation. (2010). IBM SPSS Statistics for Windows (19.0). [Software de cómputo]. Armonk, NY: Autor.
- Instituto Nacional de Evaluación Educativa. (2019). *Pisa 2018. Informe español*. Madrid, España: Ministerio de Educación y Formación Profesional. Recuperado de <https://www.educacionyfp.gob.es/inee/portada.html>
- Iriarte-Díaz-Granados, F., Espeleta-Maya, A., Zapata-Zapata, E., Cortina-Peñaranda, L., Zambrano-Ojeda, E., & Fernández-Candama, F. (2010). El razonamiento

- lógico en estudiantes universitarios. *Zona Próxima*, 12. Recuperado de <http://rescientificas.uninorte.edu.co/index.php/zona>
- Juárez-Hernández, L. G., & Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Revista Espacios*, 39(53), 23. Recuperado de <http://www.revistaespacios.com>
- Kobrin, J. L., & Kimmel, E. W. (2006). Test development and technical information on the writing section of the SAT reasoning test. *Research Notes*, 25. Recuperado de <https://files.eric.ed.gov>
- Lassiter, D., & Goodman, N. (2015). How many kinds of reasoning? Inference, probability and natural language semantics. *Cognition*, 136, 123-134. doi: 10.1016/j.cognition.2014.10.016
- Lawrence, I. M., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2003). *A historical perspective on the content of the SAT*. College Board (Report N° 2003-3). doi: 10.1002/j.2333-8504.2003.tb01902.x
- López, I., Echazarreta, C., Pech, S., & Gómez, B. (2010). Selección y permanencia en la educación superior: El caso de la Universidad Autónoma de Yucatán. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 90-102. Recuperado de http://rinace.net/rie/rie_home.html
- Manzi, J., Bosch, A., Bravo, D., del Pino, G., Donoso, G., & Pizarro, R. (2010). Validez diferencial y sesgo en la predictividad de las pruebas de admisión a las universidades chilenas (PSU). *Revista Iberoamericana de Evaluación Educativa*, 3(2), 29-48. Recuperado de http://rinace.net/rie/rie_home.html
- Marino, T., & Rodríguez, M. (2009). Un estudio exploratorio sobre heurísticas en estudiantes de un curso de matemática de nivel pre-universitario. *Paradigma*, 30(2), 159-178. Recuperado de <http://revistaparadigma.online/ojs/index.php/paradigma>
- Martínez-Arias, M. R. (2005). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid, España: Síntesis.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8. doi: 10.1111/j.1745-3992.1995.tb00881.x
- MIDEPLAN. (2017). *Costa Rica Índice de Desarrollo Social (IDS)*. Recuperado de <https://www.mideplan.go.cr>
- Milewski, G. B., Johnsen, D., Glazer, N., & Kubota, M. (2005). *A survey to evaluate the alignment of the New SAT®. Writing and critical reading sections to curricula and instructional practices* (Research report N° 2005-1). College Board. doi: 10.1002/j.2333-8504.2005.tb01984.x
- Moreira-Mora, T. E. (2016). Eficacia de la prueba de aptitud académica para pronosticar el rendimiento académico: Un reto de las universidades públicas. *VIII Congreso del Consejo Superior Universitario Centroamericano (CSUCA)*, Ciudad de Panamá. Recuperado de <https://repositoriotec.tec.ac.cr>
- Multivariate Software Inc. (2012). EQS for Windows (6.2). [Software de cómputo]. Recuperado de <http://www.mvsoft.com/eqsdownload.htm>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. Recuperado de <http://www.psicothema.com>
- Pardo, A., & Ruiz, M. A. (2002). *SPSS II. Guía para el análisis de datos*. Madrid, España: McGraw-Hill.
- R Core Team. (2018). R: A language and environment for statistical computing. [Software de cómputo]. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <http://www.R-project.org>
- Rojas-Torres, L., & Ordóñez-Gutiérrez, G. (2019). Proceso de construcción de pruebas educativas: El caso de la

Prueba de Habilidades Cuantitativas. *Revista Evaluar*, 19(2). doi: [10.35670/1667-4545.v19.n2.25080](https://doi.org/10.35670/1667-4545.v19.n2.25080)

Ruiz, M. A., Pardo, A., & San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del Psicólogo*, 31(1), 34-45. Recuperado de <http://www.papeles-delpsicologo.es>

StataCorp. (2015). Stata Statistical Software: Release 14. [Software de cómputo]. College Station, TX: Autor.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100-107. Recuperado de <http://www.psicothema.com>

Vergara-Díaz, G., & Peredo-López, H. (2017). Relación del desempeño académico de estudiantes de primer año de universidad en Chile y los instrumentos de selección para su ingreso. *Revista Educación*, 41(2), 95-104. doi: [10.15517/revedu.v41i2.21514](https://doi.org/10.15517/revedu.v41i2.21514)

Young, J. W. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis*. College Board Research (Reporte N° 2001-6). Recuperado de <https://files.eric.ed.gov>

Measurement Invariance of the New Sexual Satisfaction Scale short form

Invarianza de la Nueva Escala de Satisfacción Sexual versión corta

Virginia Flores-Perez ¹, Esteban Jaime Camacho-Ruiz ¹*, María del Consuelo Escoto-Ponce de León ², Brenda Sarahi Cervantes-Luna ², Manuel Leonardo Ibarra-Espinosa ¹

1 - Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México, Nezahualcóyotl, México.

2 - Centro Universitario UAEM Ecatepec, Universidad Autónoma del Estado de México, Ecatepec, México.

Recibido: 30/09/2020 Revisado: 17/11/2020 Aceptado: 08/12/2020

Abstract

The objective was to assess the incremental validity, internal consistency and temporal stability of the New Sexual Satisfaction Scale short version (NSSS-S) in Spanish-speaking Latino adults. The sample included 401 university students (232 women and 169 men), sexually active, 18 to 38 years old ($M = 20.83$; $SD = 2.66$). A subsample of 76 participants answered the scale four weeks after the first application. Confirmatory factor analysis (CFA) identified a two-factor structure. The invariance analysis with multi-group confirmatory factor analysis (MGCFA) confirmed the invariance by sex of the scale. The internal consistency reliability of the scale and its subscales was excellent for both men and women ($\omega = .89 - .95$) and the temporal stability, four weeks after the first application, was adequate ($r = .74$; $ICC = .85$). Finally, men scored higher in overall sexual satisfaction and the Ego-Centered subscale, but not in the Partner -and Sexual Activity-Centered subscale, as compared to women. To conclude, the NSSS-S is a psychometrically appropriate scale to measure sexual satisfaction in Spanish-speaking Latino adults.

Keywords: *measurement invariance, factorial validity, confirmatory factor analysis, sexual satisfaction, university students*

Resumen

El objetivo fue evaluar la validez incremental, la consistencia interna y la estabilidad temporal de la Nueva Escala de Satisfacción Sexual versión corta. La muestra incluyó 401 universitarios (232 mujeres y 169 hombres), sexualmente activos, de 18 a 38 años de edad ($M = 20.83$; $DE = 2.66$). Una submuestra de 76 participantes respondió la escala cuatro semanas después de la primera aplicación. El análisis factorial confirmatorio identificó una estructura de dos factores. El análisis de invarianza, usando MGCFA confirmó la invarianza por sexo de la escala. La consistencia interna de la escala y sus subescalas fue excelente, tanto para hombres como para mujeres ($\omega = .89-.95$) y la estabilidad temporal, cuatro semanas después de la primera aplicación, fue adecuada ($r = .74$; $ICC = .85$). Finalmente, los hombres puntuaron más alto en satisfacción sexual global y en Ego-Centered, pero no en Partner -and Sexual Activity-Centered, en comparación con las mujeres. La NSSS-S es una medida de satisfacción sexual psicométricamente apropiada para la población hispano-latina.

Palabras clave: *medición de invarianza, validez factorial, análisis factorial confirmatorio, satisfacción sexual, universitarios*

Introduction
Method
Results
Discussion
References

Author's note

The first author thanks the National Council of Science and Technology (CONACYT) for granting scholarship 701590 for the completion of the master's studies in Sociology of Health.

***Correspondence to:** Esteban Camacho Ruiz. Address: Av. Bordo de Xochiaca, Col. Benito Juárez, Nezahualcóyotl, Estado de México, México. C.P. 57000. Telephone: (52) 55-51126372 ext.7912. E-mail: jaimo_camacho_ruiz@hotmail.com

How to cite: Flores-Perez, V., Camacho-Ruiz, E. J., Escoto-Ponce de León, M del C., Cervantes-Luna, B. S., & Ibarra-Espinosa, M. L. (2021). Measurement invariance of the New Sexual Satisfaction Scale short form. *Revista Evaluar*, 21(1), 94-103. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introduction

Sexual satisfaction (SS), the last phase of the human sexual response (Carrobes & Sanz-Yaque, 1991) refers to the subjective assessment of liking or disliking that people make about their sexual relationships and includes the pleasant sensations derived from the physical act that are related to the emotional satisfaction of the individual in order to find confidence and self-confidence (Ahumada, Luttes, Molina, & Torres, 2014; Lawrance & Byers, 1995). High levels of SS are associated with higher *global* quality of life and self-esteem, and less depression, anxiety, stress, physical health problems, and poor self-concept; furthermore, both, self-esteem and body image predict SS (for a review, see Sánchez-Fuentes, Santos-Iglesias, & Sierra, 2014).

Several instruments have been developed to measure SS: The Index of Sexual Satisfaction (ISS; Hudson, Harrison, & Crosscup, 1981), the Golombok-Rust Inventory of Sexual Satisfaction (GRISS; Rust & Golombok, 1986), the Pinney Sexual Satisfaction Inventory (PSSI; Pinney, Gerrard, & Denney, 1987), the Global Measure of Sexual Satisfaction (GMSEX; Lawrance & Byers, 1995), the Sexual Satisfaction Scale for Women (SSS-W; Meston & Trapnell, 2005), the Subjective Sexual Satisfaction Scale (ESSS; González-Rivera, Veray-Alicea, Santiago-Santos, Castro-Castro, & Quiñones-Soto, 2017) and its short form (ESSS-B; González-Rivera & Hernández-Gato, 2019), the Sexual Sensation Seeking Scale (Kalichman et al., 1994; Moral de la Rubia, 2018) and the New Sexual Satisfaction Scale (NSSS; Štulhofer, Buško, & Brouillard, 2010).

The PSSI (Pinney et al., 1987) and the SSS-W (Meston & Trapnell, 2005) are scales designed to assess SS for women only. The first evaluates the general SS and the SS with the partner; the second evaluates two relational domains

(*communication and compatibility*) and three interpersonal domains (*feelings of inner happiness, concern for relationship, and personal concern*) of women's SS. The ISS (Hudson et al., 1981) is a one-dimensional scale that assesses the degree, severity, and magnitude of relationship problems in the sexual component. However, the ISS was built based on the clinical and personal experience of the author, so its construct validity is questionable. The GRISS (Rust & Golombok, 1986) is a measure of the existence and severity of sexual dysfunctions (e.g., impotence, anorgasmia, premature ejaculation, and vaginismus), designed for heterosexual couples. The GMSEX (Lawrance & Byers, 1995) evaluates the overall SS; however, the scale is restricted in terms of variance, since the measurement is made from an item and its reliability is limited (Mark, Herbenick, Fortenberry, Sanders, & Reece, 2014). The ESSS (González-Rivera et al., 2017) is a multidimensional scale that evaluates the perception of sexual life, negative feelings after sexual activity, perceived sexual performance, and the individual role during sexual intercourse. However, the ESSS focuses on the evaluation of the SS centered on the person.

The NSSS (Štulhofer et al., 2010) is a useful instrument to assess SS and is based on the theory of Bancroft, Loftus, and Long (2003) called *three windows*. The authors evaluated the scale in diverse samples: adults and young, men and women, heterosexual and different sexual orientation, clinical and community samples. The NSSS can be used for research in non-clinical settings.

Originally the NSSS (Štulhofer et al., 2010) included 20 items. The principal component analysis (PCA) with oblimin rotation derived two dimensions: The Ego-Centered subscale, devoted on personal experiences and sensations (e.g. quality and frequency of one's sensations, orgasms, and sexual excitement) and the Partner -and Sex-

ual Activity-Centered subscale (e.g. trust, commitment and emotional closeness with a partner). The authors evaluated the internal consistency in two student samples, two community samples, and one sample of Croatian non-heterosexual adults, and reported Cronbach's alpha indices of $\geq .94$ for the full-scale and $\geq .90$ for both subscales and the short version.

In a later study, Štulhofer, Buško and Brouillard (2011) derived a short version (NSSS-S), which includes 12 Likert-type items (1 = *Not at all satisfied* to 5 = *Extremely satisfied*). The PCA with oblimin rotation derived a factor that measures personal and partner sexual satisfaction, regardless of gender, sexual orientation, or marital status.

The factorial structure of the NSSS-S has been studied in different countries and languages: Canada and United States (Mark et al., 2014), United States and Croatia (Štulhofer et al., 2011), Portugal (Santos-Pechorro et al., 2016), Spain (Strizzi, Fernández-Agis, Alarcón-Rodríguez, & Parrón-Carreño, 2016) and Germany (Hoy, Strauß, Kröger, & Brenk-Franz, 2019). Previous

studies have included participants from community sample, clinical sample, university students, and people with different sexual orientations, both men and women.

In general, studies have found that the NSSS-S has excellent internal consistency ($\alpha = .90 - .96$) and adequate test-retest temporal stability ($r = .72 - .84$; Mark et al., 2014; Štulhofer et al., 2011). However, the results regarding factor validity are inconsistent. In this sense, some studies have found that the scale is one-dimensional (Hoy et al., 2019; Štulhofer et al., 2011), while others have identified two factors (Mark et al., 2014; Santos-Pechorro et al., 2016; Strizzi et al., 2016).

The NSSS-S was recently translated and adapted for adult women and men in Spain (Strizzi et al., 2016). The principal axis factoring extraction with promax rotation derived two factors. Confirmatory factor analysis (CFA) proved the existence of two factors. It should be noted that factor validity for the NSSS-S has not been evaluated in Spanish-speaking Latino adults. Furthermore, the invariance by sex for this scale has not been studied.

Table 1
Goodness of fit indices for the NSSS-S.

Models	χ^2/SCF	<i>df</i>	<i>p</i>	RMSEA (CI)	SRMR	CFI	TLI
CFA unidimensional (n = 401)	162.137	54	.0001	.084 (.072 - .097)	.040	.934	.919
CFA bifactor (n = 401; [F1 = items 1-6; F2 = items 7-12])	143.921	53	.0001	.078 (.066 - .091)	.038	.944	.930
CFA bifactor re-specified (n = 401; [F1 = items 1-6; F2 = items 7-12])	112.255	52	.0001	.066 (.053 - .079)	.033	.961	.950

Note. χ^2 = Chi square; SCF = scaling correction factor for MLR; *df* = degrees of freedom; *p* = significance; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = root mean square residual; CFI = comparative fit index; TLI = Tucker-Lewis index.

Finally, previous studies have used Cronbach's alpha coefficient to assess the internal consistency of the scale. However, some authors have pointed out that the Cronbach's alpha coefficient has some limitations, for example, it is affected by the number of items, the number of response options, and the proportion of the variance (Dominguez-Lara & Merino-Soto, 2015). Therefore, it has been suggested to include more adequate indices for calculating the internal consistency of scales of an ordinal nature. In this sense, this study aimed to evaluate two previously identified factorial models for the NSSS-S (Mark et al., 2014; Štulhofer et al., 2011), in Spanish-speaking Latino adults using CFA. It also aimed to assess the incremental validity, the internal consistency and the temporal stability of the scale using multi-group confirmatory factor analysis (MGCFA) and finally, to compare the SS between men and women.

Methods

Participants

The sample included 401 university students (232 female and 169 male), sexually active, from 18 to 38 years old ($M = 20.83$; $SD = 2.66$) selected by non-probability convenience sampling, selected from two public universities in the municipality of Nezahualcoyotl, State of México. A subsample of 76 participants answered the scale four weeks after the first application.

Instruments

Sexual satisfaction. The NSSS-S (Štulhofer et al., 2011) was previously described. In the present study, we used the translated version, adapted by Strizzi et al. (2016).

Table 2
Invariance by sex.

Models	χ^2/SCF	<i>df</i>	<i>p</i>	RMSEA (CI)	SRMR	CFI	TLI	Δ RMSEA	Δ CFI
MGCFA									
<i>By sex</i> (n = 401)									
Females re-specified (n = 232)	111.292	52	.0001	.086 (.070 - .104)	.042	.938	.921		
Males re-specified (n = 169)	48.030	52	.259	.026 (.000 - .057)	.032	.993	.991		
Configural invariance	162.792	104	.0001	.069 (.055 - .083)	.038	.958	.947		
Metric invariance	169.167	114	.0001	.064 (.050 - .078)	.043	.960	.954	-.002	.002
Scalar invariance	182.330	124	.0001	.062 (.049 - .076)	.044	.959	.956	.007	-.001
Strict invariance	183.469	136	.0001	.057 (.043 - .070)	.049	.963	.964	.005	.004

Note. χ^2 = Chi square; SCF = scaling correction factor for MLR; *df* = degrees of freedom; *p* = significance; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = root mean square residual; CFI = comparative fit index; TLI = Tucker-Lewis index; Δ = incremental value.

Procedure

The present study followed the ethical principles of the Declaration of Helsinki (World Medical Association, 1964) for medical research involving human subjects. After receiving informed consent, participants were notified of the purpose of the study and proceeded to answer the scale in their classroom.

Data analysis

Multivariate normality analyses (Mardia, 1970) were performed using the R software (R Core Team, 2013). Absolute values of > 5 and a $p < .05$ were indicative of non-normality.

All confirmatory analyses were performed with Mplus 8.0 software (Muthén & Muthén, 2017). Given the discrepancies in the underlying structure of the NSSS-S, we did not use the one-dimensional structure proposed by Štulhofer et al. (2011), instead, we tested two models: one-dimensional (Štulhofer et al., 2011) and bifactorial (Mark et al., 2014). To assess the fit of the model, two absolute fit indices were used: 1) the Satorra-Bentler scaled chi-square (S-B χ^2/df) and 2) the standardized root mean square residual (SRMR). Additionally, two comparative adjustment indices were used: The comparative fit index (CFI) and the Tucker-Lewis index (TLI); and a parsimonious adjustment index, the root mean square error of approximation (RMSEA). To consider that the model has a proper fit, values less than 2 indicate good fit and values between 2 and 3 indicate acceptable fit for the S-B χ^2/df index (Schermelleh-Engel, Moosbrugger, & Müller, 2003); for the SRMR index, values less than .05 are indicative of a good fit (Hooper, Coughlan, & Mullen, 2008); for CFI and TLI, values $> .95$ are

interpreted as a good fit (Hu & Bentler, 1999); and for the RMSEA, an index between .05 and .08 is considered as a reasonable adjustment (Browne & Cudeck, 1993).

The analysis of invariance by sex with MG-CFA included the examination of the configural invariance, the metric invariance, the scalar invariance, and the residual or strict invariance. To evaluate the incremental models, the Cheung and Rensvold (2002) criteria were used: $\Delta RMSEA \leq .015$ and $\Delta CFI \leq .01$.

To calculate the internal consistency of the scale, McDonald's (1999) coefficient omega (ω) was used, which works with factor loads, which reduce the risk of overestimating reliability since it does not depend on the number of items and can be used in continuous variables. The values of ω can be interpreted as: unacceptable ($< .70$), fair (.70 - .79), good (.80 - .89) and excellent ($\geq .90$; Cicchetti, 1994).

Finally, to obtain evidence of temporal stability, Pearson's correlation coefficients (r) and intraclass correlation (ICC) were used. Values of r can indicate: null correlation ($< .30$), low correlation (.30 - .49), moderate correlation (.50 - .69), high correlation (.70 - .90) and very high correlation ($> .90$; Hinkle, Wiersma, & Jurs, 2003). For the ICC, the values are classified as: low reliability ($< .5$), moderate reliability (.5 - .74), good reliability (.75-.90) and excellent reliability ($> .90$; Portney & Watkins, 2000).

Results

Preliminary analysis

Multivariate normality tests indicated that the data was not normally distributed (skewness 15.98, $p = 1.00$; kurtosis 208.17, $p < .001$), for both men (skewness 27.17, $p = 1.00$; kurtosis

Table 3
Internal consistency and temporal stability of the NSSS-S by sex.

Scale and sub-scales	Males			Females			Full sample		
	<i>r</i> (CI, 95%)	ICC (CI, 95%)	ω	<i>r</i> (CI, 95%)	ICC (CI, 95%)	ω	<i>r</i> (CI, 95%)	ICC (CI, 95%)	ω
Ego-Centered	.51 (.04 - .82)	.67 (.38 - .83)	.90	.82 (.60 - .92)	.90 (.81 - .95)	.89	.71 (.48 - .84)	.83 (.73 - .89)	.89
Partner- and Sexual Activity-Centered	.70 (.36 - .87)	.82 (.66 - .91)	.91	.79 (.33 - .91)	.88 (.77 - .94)	.89	.75 (.54 - .87)	.85 (.77 - .91)	.90
Full-scale	.65 (.33-.86)	.79 (.60-.89)	.95	.82 (.49-.92)	.90 (.80-.95)	.94	.74 (.50-.87)	.85 (.77-.91)	.94

Note. Pearson and intraclass coefficients were significant ($p = .0001$).

202.58, $p < .001$), and women (skewness 22.48, $p = 1.00$; kurtosis 206.85, $p < .001$). Therefore, subsequent analyses were performed using the robust maximum likelihood estimation (MLR) method.

Confirmatory factor analysis

The CFA was performed to evaluate the fit of two models: one-dimensional and bifactorial (F1 = items 1-6; F2 = items 7-12). As shown in Table 1, the two models have adequate fit indices, although Model 2 presents better fit indices. After inspecting the modification indices, we released covariance errors between items 11 and 12, thus, improving the fit. For this model, the item-total correlations ranged from .71 to .82. The items loaded significantly in their respective factor ($p < .0001$) and ranged from .69 to .83. Subsequent analyses were performed using the re-specified two-factor model (F1 = items 1-6; F2 = items 7-12).

Invariance analysis by sex

We first evaluated the fit of the re-specified model for each group (men and women). As shown in Table 2, the fit was *acceptable* for women and *good* for men. For the configural invariance, the measure of model fit was good. Furthermore, the changes in CFI and RMSEA indicate that the NSSS-S is invariant for sex, therefore, the scale can be applied to men and women.

Internal consistency

The omega coefficient indicated that the 12 items of the NSSS-S show good internal consistency for the whole sample, for both men and women (Table 3).

Temporal stability analysis

A subsample of 76 participants answered

the NSSS-S four weeks after the first application. Temporal stability analysis showed low Pearson and ICC correlations for men compared to women on the scale and its two subscales (Table 3).

Sex comparisons

Since the NSSS-S was invariant for sex, SS was compared in men and women. Results showed that men have higher global SS ($M = 47.07$, $SD = 8.50$) than women ($M = 46.44$, $SD = 9.51$), $t_{(401)} = .68$, $p = .50$. In addition, in the Ego-Centered subscale, men scored higher ($M = 24.14$, $SD = 4.17$) than women ($M = 23.58$, $SD = 4.77$), $t_{(401)} = 1.26$, $p = .21$. Finally, no differences were observed in the partner -and Sexual Activity-Centered subscale $t_{(401)} = .12$, $p = .90$, between men ($M = 22.93$, $SD = 4.72$) and women ($M = 22.86$, $SD = 5.15$).

Discussion

The purpose of this study was to assess the incremental validity of the NSSS-S. To our knowledge, this is the first study to focus on this aspect of the scale. Therefore, after re-specifying the model, the invariance by sex was tested. The NSSS-S showed metric, scalar, and strict invariance. These findings indicate that the construct represented by the NSSS-S is the same for university men and women, and the scores and means of the NSSS-S can be compared between these groups.

Although there are various instruments to evaluate SS (González-Rivera et al., 2017; Hudson et al., 1981; Lawrance & Byers, 1995; Meston

& Trapnell, 2005; Pinney et al., 1987; Rust & Golombok, 1986), the NSSS-S has a theoretical framework and allows evaluating personal SS as a couple, regardless of gender, sexual orientation or marital status. Furthermore, the scale can be used in non-clinical settings. However, studies have reported inconsistent results regarding the factor structure of the NSSS-S. Some authors have found that the scale is unifactorial (Hoy et al., 2019; Štulhofer et al., 2011), and others point out that it is bifactorial (Mark et al., 2014; Santos-Pechorro et al., 2016; Strizzi et al., 2016). Our results support the two-factor structure of NSSS-S.

Most previous studies have evaluated the internal consistency of the NSSS-S using the Cronbach's alpha coefficient. In this study, internal consistency was assessed using the McDonald's omega coefficient (1999), which allows a more precise estimation of the internal consistency of the items and reduces the effect of the number of the items included in the scale. Consistent with the findings of previous studies (Hoy et al., 2019; Mark et al., 2014; Santos-Pechorro et al., 2016; Strizzi et al., 2016; Štulhofer et al., 2011), our findings support the internal consistency of the NSSS-S and its subscales for both men and women. Furthermore, the scale has adequate temporal stability, which is also consistent with previous studies (Mark et al., 2014; Štulhofer et al., 2011).

In gender comparisons, our findings are consistent with previous studies (for a review, see Petersen & Hyde, 2010), which indicate that SS is higher in men than in women. These results support the conclusions of Hyde (2005) about gender differences in sexuality. Also, each society has different criteria to value sexuality, being that men have greater sexual freedom as compared to women (Grindell, 2009).

Finally, the findings of this study indicate that the NSSS-S is a two-dimensional measure

that allows evaluating SS in Spanish-speaking Latino adults. The scale has incremental validity, adequate internal consistency, and temporal stability.

Limitations

Despite the contribution of this study, we can highlight some limitations, for example, the sample was recruited from only two public universities, which could be unrepresentative of adults from other Spanish-speaking cities, so future studies could include samples representative of other Spanish-speaking regions and may even include older adults.

SS studies have mainly been focused on heterosexual people (Sánchez-Fuentes & Sierra, 2015). Another limitation of this study was that it did not include a sample of participants with different sexual orientation, so this aspect should also be addressed in future studies with Hispanic-Latino populations.

In addition, future studies could assess the invariance of this scale, according to the body mass index and the marital status of the participants, which are variables associated with SS. Furthermore, in this study it was not possible to cross-validate the re-specified model using a new sample, which would represent another contribution to the incremental validity of the scale.

Finally, a further limitation of this study was that convergent and divergent validity of the scale were not evaluated.

References

Ahumada, S. G., Luttges, C. D., Molina, T. G., & Torres, S.

G. (2014). Satisfacción sexual: Revisión de los factores individuales y de pareja relacionados [Sexual satisfaction: Review of related individual and partner factors]. *Revista Hospital Clínico Universidad de Chile*, 25(4), 278-284. Retrieved from <https://www.redclinica.cl/ActividadAcad%C3%A9mica/RevistaHCUCH/DespliegueRevHospClinUnivChile/042014/tabid/1175.aspx>

Bancroft, J., Loftus, J., & Long, J. S. (2003). Distress about sex: A national survey of women in heterosexual relationships. *Archives of Sexual Behavior*, 32(3), 193-208. Retrieved from <https://www.springer.com/journal/10508>

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Thousand Oaks, CA: Sage Publications.

Carrobes, J. A., & Sanz-Yaque, A. (1991). *Terapia sexual [Sexual therapy]*. España: Fundación Universidad Empresa.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. doi: 10.1037/1040-3590.6.4.284

Domínguez-Lara, S. A., & Merino-Soto, C. M. (2015). ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? [Why is it important to report the confidence intervals of the Cronbach's alpha coefficient?]. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 13(2), 1326-1328. Retrieved from <http://revistaumanizales.cinde.org.co/rllcsnj/index.php/Revista-Latinoameri>

[cana/index](#)

- González-Rivera, J. A., & Hernández-Gato, I. (2019). Validación de una versión breve de la Escala de Satisfacción Sexual Subjetiva (ESSS-B) en Puerto Rico [Validation of a Short Version of the Subjective Sexual Satisfaction Scale (ESSS-B) in Puerto Rico]. *Revista Evaluar*, 19(2), 43-57. doi: [10.35670/1667-4545.v19.n2.25082](https://doi.org/10.35670/1667-4545.v19.n2.25082)
- González-Rivera, J. A., Veray-Alicea, J., Santiago-Santos, D., Castro-Castro, S., & Quiñones-Soto, R. (2017). Desarrollo y validación de una escala para medir satisfacción sexual subjetiva en adultos puertorriqueños [Development and validation of a scale to measure subjective sexual satisfaction in Puerto Rican adults]. *Salud y Conducta Humana*, 4(1), 52-63. Retrieved from <http://rsych.com>
- Grindell, V. M. (2009). *Sexual satisfaction in young adults: Are there gender and age differences using the crucible® sexual relationship inventory?* (Publication No.3421101) [Doctoral dissertation, Alliant International University]. ProQuest Dissertations & Theses Global.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60. Retrieved from <http://www.ejbrm.com/main.html>
- Hoy, M., Strauß, B., Kröger, C., & Brenk-Franz, K. (2019). Überprüfung der deutschen Kurzversion der "New Sexual Satisfaction Scale" (NSSS-SD) in einer repräsentativen Stichprobe [Evaluation of the German short version of the "New Sexual Satisfaction Scale" (NSSS-SD) in a representative sample]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 69(3-4), 129-135. doi: [10.1055/a-0620-0002](https://doi.org/10.1055/a-0620-0002)
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- Hudson, W. W., Harrison D. F., & Crosscup, P. C. (1981). A short-form scale to measure sexual discord in dyadic relationships. *Journal of Sex Research*, 17(2), 157-174. doi: [10.1080/00224498109551110](https://doi.org/10.1080/00224498109551110)
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581-592. doi: [10.1037/0003-066X.60.6.581](https://doi.org/10.1037/0003-066X.60.6.581)
- Kalichman, S. C., Johnson, J. R., Adair, V., Rompa, D., Multhauf, K., & Kelly, J. A. (1994). Sexual Sensation-Seeking: Scale development and predicting AIDS-risk behavior among homosexually active men. *Journal of Personality Assessment*, 62(3), 385-397. doi: [10.1207/s15327752jpa6203_1](https://doi.org/10.1207/s15327752jpa6203_1)
- Lawrance, K., & Byers, E. S. (1995). Sexual satisfaction in long-term heterosexual relationship: The interpersonal exchange model of satisfaction. *Personal Relationship*, 2(4), 267-285. Retrieved from <https://onlinelibrary.wiley.com/journal/14756811>
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530. doi: [10.1093/biomet/57.3.519](https://doi.org/10.1093/biomet/57.3.519)
- Mark, K. P., Herbenick, D., Fortenberry, J. D., Sanders, S., & Reece, M. (2014). A psychometric comparison of three scales and a single-item measure to assess sexual satisfaction. *The Journal of Sex Research*, 51(2), 159-169. doi: [10.1080/00224499.2013.816261](https://doi.org/10.1080/00224499.2013.816261)
- McDonald, R. P. (1999). *Test theory: A unified treatment*. New York, NY: Lawrence Erlbaum Associates. doi: [10.4324/9781410601087](https://doi.org/10.4324/9781410601087)
- Meston, C., & Trapnell, P. (2005). Development and validation of a five-factor sexual satisfaction and distress scale for women: The Sexual Satisfaction Scale for

- Women (SSS-W). *The Journal of Sexual Medicine*, 2(1), 66-81. doi: [10.1111/j.1743-6109.2005.20107.x](https://doi.org/10.1111/j.1743-6109.2005.20107.x)
- Moral de la Rubia, J. (2018). Validación de la Escala de Búsqueda de Sensaciones Sexuales en mexicanos casados o cohabitantes [Validation of the Sexual Sensation Seeking Scale among married or cohabiting Mexicans]. *Revista Evaluar*, 18(3), 49-72. doi: [10.35670/1667-4545.v18.n3.22224](https://doi.org/10.35670/1667-4545.v18.n3.22224)
- Muthén, L. K., & Muthén, B. O. (2017). Mplus (8.0). [Computer software]. Retrieved from <https://www.statmodel.com>
- Petersen, J. L., & Hyde, J. S. (2010). A meta-analytic review of research on gender differences in sexuality, 1993-2007. *Psychological Bulletin*, 136(1), 21-38. doi: [10.1037/a0017504](https://doi.org/10.1037/a0017504)
- Pinney, E. M., Gerrard, M., & Denney, N. W. (1987). The Pinney Sexual Satisfaction Inventory. *The Journal of Sex Research*, 23(2), 233-251. doi: [10.1080/00224498709551359](https://doi.org/10.1080/00224498709551359)
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org>
- Rust, J., & Golombok, S. (1986). The GRISS: A psychometric instrument for the assessment of sexual dysfunction. *Archives of Sexual Behavior*, 15(2), 157-165. doi: [10.1007/BF01542223](https://doi.org/10.1007/BF01542223)
- Sánchez-Fuentes, M. M., Santos-Iglesias, P., & Sierra, J. C. (2014). A systematic review of sexual satisfaction. *International Journal of Clinical and Health Psychology*, 14(1), 67-75. doi: [10.1016/S1697-2600\(14\)70038-9](https://doi.org/10.1016/S1697-2600(14)70038-9)
- Sánchez-Fuentes, M. M., & Sierra, J. C. (2015). Sexual satisfaction in a heterosexual and homosexual Spanish sample: The role of socio-demographic characteristics, health indicators, and relational factors. *Sexual and Relationship Therapy*, 30(2), 226-242. doi: [10.1080/14681994.2014.978275](https://doi.org/10.1080/14681994.2014.978275)
- Santos-Pechorro, P., Moneiro-Pascoal, P., Neves-Jesus, S., Almeida, A. I., Soares-Figueiredo, C., & Vieira, R. X. (2016). Propriedades psicométricas da versão portuguesa da Nova Escala de Satisfação Sexual-versão curta [Psychometric properties of the Portuguese version of the New Sexual Satisfaction Scale-Short form]. *Revista Internacional de Andrologia*, 14(3), 94-100. doi: [10.1016/j.androl.2016.04.006](https://doi.org/10.1016/j.androl.2016.04.006)
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74. Retrieved from <https://www.dgps.de/fachgruppen/methoden/mpr-online>
- Strizzi, J., Fernández-Agis, I., Alarcón-Rodríguez, R., & Parrón-Carreño, T. (2016). Adaptation of the New Sexual Satisfaction Scale-Short Form into Spanish. *Journal of Sex & Marital Therapy*, 42(7), 579-588. doi: [10.1080/0092623X.2015.1113580](https://doi.org/10.1080/0092623X.2015.1113580)
- Štulhofer, A., Buško, V., & Brouillard, P. (2010). Development and bicultural validation of the New Sexual Satisfaction Scale. *The Journal of Sex Research*, 47(4), 257-268. doi: [10.1080/00224490903100561](https://doi.org/10.1080/00224490903100561)
- Štulhofer, A., Buško, V., & Brouillard, P. (2011). The New Sexual Satisfaction Scale and its short form. In T. D. Fisher, C. M. Davis, W. L. Yarber & S. L. Davis (Eds.), *Handbook of Sexuality-Related Measures* (3rd ed., pp. 530-533). doi: [10.4324/9781315881089](https://doi.org/10.4324/9781315881089)
- World Medical Association. (1964). *Declaration of Helsinki*. Retrieved from <https://www.wma.net>