

**Desarrollo y evidencias preliminares de validez y confiabilidad de la
Escala de Estado General de Salud Mental y Bienestar Conductual (EEGSC)
en universitarios peruanos: un estudio piloto**

Gustavo Francia

Facultad de Ingeniería, Universidad San Ignacio de Loyola

Lima, Perú

Nota del autor

Gustavo Francia ^{ID} <https://orcid.org/0009-0007-3489-7812>

El autor es fundador y director ejecutivo de PsicoTeddy S.A.C. (RUC 20614733561), empresa peruana que desarrolla una aplicación digital de acompañamiento emocional en la que el instrumento evaluado se encuentra implementado. El presente estudio fue diseñado y ejecutado por el autor a título personal, en el marco de su formación académica, sin financiamiento ni participación de la empresa en la recolección de datos, el análisis ni la redacción del manuscrito. La base de datos anonimizada y el código de análisis se encuentran disponibles para verificación independiente.

La correspondencia relativa a este artículo debe dirigirse a Gustavo Francia, Universidad San Ignacio de Loyola, Av. La Fontana 550, La Molina, Lima, Perú. Correo electrónico: gfrancia@usil.edu.pe

Resumen

La demanda de tamizaje breve de salud mental en plataformas digitales ha crecido más rápido que la disponibilidad de instrumentos desarrollados para ese contexto en población hispanohablante. Este estudio instrumental reporta la construcción y las evidencias psicométricas preliminares de la Escala de Estado General de Salud Mental y Bienestar Conductual (EEGSC). Se aplicó una versión inicial de 22 ítems (formato Likert de cinco puntos, marco temporal de dos semanas) a una muestra por conveniencia de 70 estudiantes universitarios de Lima Metropolitana (62.9 % mujeres; $M_{edad} = 21.56$, $DE = 3.74$; rango 17–34). El análisis de ítems identificó un reactivo redactado en dirección positiva con correlación ítem-total corregida negativa ($r = -.160$), atribuible a un error de codificación de la puntuación inversa; su recodificación no alcanzó el criterio mínimo ($r = .160$) y su adecuación muestral individual fue insuficiente ($KMO = .437$), por lo que fue eliminado. La versión resultante de 21 ítems mostró adecuación muestral adecuada ($KMO = .842$; Bartlett $\chi^2(210) = 910.90$, $p < .001$) y consistencia interna elevada ($\alpha = .929$; $\omega = .931$), con correlaciones ítem-total corregidas entre .346 y .757. El análisis paralelo de Horn retuvo un único componente (autovalor observado = 8.93 frente a 2.38 esperado por azar), sugiriendo una estructura esencialmente unidimensional. Las seis agrupaciones con dos o más reactivos presentaron consistencia interna aceptable a buena ($\alpha = .722-.847$). Los índices de ajuste de todos los modelos de medida evaluados resultaron insuficientes (mejor modelo: $CFI = .799$, $TLI = .753$, $RMSEA = .121$, $SRMR = .145$). La EEGSC-21 muestra evidencias preliminares sólidas de consistencia interna y homogeneidad de reactivos, y una estructura dominada por un factor general compatible con los modelos transdiagnósticos contemporáneos. La estructura de siete subescalas propuesta en la fase de construcción no queda respaldada empíricamente con el tamaño muestral disponible, y no se establecieron puntos de corte clínicos. El instrumento debe considerarse en fase de desarrollo.

Palabras clave: desarrollo de instrumentos, salud mental digital, análisis factorial exploratorio, consistencia interna, tamizaje, estudiantes universitarios

Abstract

Demand for brief mental health screening in digital platforms has outpaced the availability of instruments developed for that context in Spanish-speaking populations. This instrumental study reports the construction and preliminary psychometric evidence of the General Mental Health and Behavioral Wellbeing Scale (EEGSC). An initial 22-item version (five-point Likert format, two-week time frame) was administered to a convenience sample of 70 university students in Metropolitan Lima (62.9 % women; $M_{age} = 21.56$, $SD = 3.74$; range 17–34). Item analysis identified one positively worded item with a negative corrected item-total correlation ($r = -.160$), attributable to a reverse-scoring coding error; recoding did not reach the minimum criterion ($r = .160$) and its individual sampling adequacy was insufficient ($KMO = .437$), leading to its removal. The resulting 21-item version showed adequate sampling adequacy ($KMO = .842$; Bartlett's $\chi^2(210) = 910.90$, $p < .001$) and high internal consistency ($\alpha = .929$; $\omega = .931$), with corrected item-total correlations from .346 to .757. Horn's parallel analysis retained a single component (observed eigenvalue = 8.93 vs. 2.38 expected by chance), suggesting an essentially unidimensional structure. The six groupings with two or more items showed acceptable to good internal consistency ($\alpha = .722-.847$). Fit indices for all measurement models were inadequate (best model: $CFI = .799$, $TLI = .753$, $RMSEA = .121$, $SRMR = .145$). The EEGSC-21 shows promising preliminary evidence of internal consistency and item homogeneity, and a structure dominated by a general factor consistent with contemporary transdiagnostic models. The seven-subscale structure proposed at the construction stage is not empirically supported at the available sample size, and no clinical cut-off scores were established. The instrument should be regarded as under development.

Keywords: instrument development, digital mental health, exploratory factor analysis, internal consistency, screening, university students

Desarrollo y evidencias preliminares de validez y confiabilidad de la Escala de Estado General de Salud Mental y Bienestar Conductual (EEGSC) en universitarios peruanos: un estudio piloto

Los trastornos mentales constituyen una de las principales causas de carga de enfermedad a nivel global, y la brecha entre necesidad y atención resulta particularmente marcada en América Latina, donde la densidad de profesionales de salud mental por habitante es sustancialmente menor que en países de altos ingresos (Organización Mundial de la Salud [OMS], 2022). En este contexto, las plataformas digitales de salud mental han sido propuestas como un mecanismo de ampliación de acceso, especialmente para funciones de tamizaje, orientación y monitoreo (Torous et al., 2021).

La utilidad de estas plataformas depende críticamente de la calidad de los instrumentos de medición que incorporan. Los instrumentos de tamizaje más difundidos —el PHQ-9 para sintomatología depresiva (Kroenke et al., 2001), el GAD-7 para ansiedad generalizada (Spitzer et al., 2006) y el DASS-21 (Lovibond & Lovibond, 1995)— cuentan con amplia evidencia acumulada, pero fueron diseñados como medidas específicas de dominio y su aplicación conjunta implica una carga de respuesta considerable para contextos de autoaplicación digital, donde las tasas de abandono son sensibles a la extensión del instrumento.

Esta situación plantea la conveniencia de explorar instrumentos breves y multidimensionales diseñados desde su origen para autoaplicación digital en población hispanohablante. El presente trabajo reporta la construcción y las evidencias psicométricas preliminares de un instrumento con ese propósito.

0.1 Marco teórico

El desarrollo de la EEGSC se apoya en la perspectiva transdiagnóstica contemporánea. Caspi et al. (2014) documentaron que la covariación entre trastornos internalizantes, externalizantes y de pensamiento es lo suficientemente alta como para justificar la postulación de un factor general de psicopatología (factor p), más parsimonioso que el modelo categórico tradicional. Kotov et al. (2017), mediante la Taxonomía Jerárquica de Psicopatología (HiTOP), ampliaron este planteamiento a una

estructura de niveles anidados que va desde el factor general hasta síntomas específicos.

Esta perspectiva tiene una implicación metodológica directa para el presente estudio: en un instrumento que muestrea de manera amplia el espectro sintomatológico, es esperable que emerja un primer factor de gran magnitud, y que las agrupaciones temáticas de menor orden expliquen porciones comparativamente pequeñas de varianza. Esta expectativa se contrasta explícitamente en los resultados.

0.2 Objetivos

El objetivo general fue construir la EEGSC y examinar sus evidencias preliminares de validez basada en la estructura interna y de confiabilidad en una muestra de estudiantes universitarios peruanos. Los objetivos específicos fueron: (a) examinar el comportamiento psicométrico de los reactivos; (b) determinar el número de dimensiones a retener mediante análisis paralelo; (c) estimar la consistencia interna de la puntuación total y de las agrupaciones temáticas propuestas; (d) evaluar el ajuste de los modelos de medida candidatos; y (e) describir la distribución de las puntuaciones y documentar las limitaciones que restringen su interpretación.

Dado el carácter exploratorio del estudio y el tamaño muestral disponible, no se formulan hipótesis direccionales sobre el número de factores. El estudio se plantea como una fase piloto de desarrollo de instrumento, previa a un estudio de validación con potencia adecuada.

1. Método

1.1 Diseño

Estudio instrumental (desarrollo y estudio de propiedades psicométricas de un test), de corte transversal y no experimental.

1.2 Participantes

La muestra estuvo conformada por 70 estudiantes universitarios de Lima Metropolitana, reclutados mediante muestreo por conveniencia en contextos académicos. Los criterios de inclusión fueron: ser estudiante universitario, tener 17 años o más, comprensión lectora en español y otorgamiento

de consentimiento informado. Se excluyó a quienes reportaran diagnóstico activo de trastorno psicótico o condición que impidiera la comprensión de los reactivos. Las características de la muestra se presentan en la Tabla 1. La tasa de completitud fue del 100 %: los 70 protocolos incluidos no presentaron datos faltantes.

1.3 Instrumento

Escala de Estado General de Salud Mental y Bienestar Conductual (EEGSC). Instrumento de autorreporte diseñado para autoaplicación digital, con marco temporal de dos semanas y formato de respuesta Likert de cinco puntos (1 = *Nunca*; 5 = *Siempre*).

Construcción del banco de reactivos. Se realizó un análisis de contenido de instrumentos consolidados de tamizaje de salud mental con el fin de identificar dominios sintomatológicos recurrentes. A partir de este mapeo se redactaron reactivos organizados en siete agrupaciones temáticas: Estado General (EG), Ansiedad y Estrés (AE), Conducta y Energía (CE), Pensamientos y Cognición (PC), Relaciones y Entorno (RE), Conductas de Riesgo (CR) y Autoimagen y Control (AC).

Validez de contenido. Los reactivos fueron sometidos a juicio de expertos para evaluar pertinencia, claridad y relevancia respecto del dominio teórico, procedimiento que derivó en la versión de 22 reactivos aplicada en este estudio. **[COMPLETAR: número exacto de jueces, sus credenciales, índice empleado (V de Aiken o razón de validez de contenido de Lawshe) y valores obtenidos. Si los jueces fueron menos de cinco, el índice de Lawshe no es aplicable; reportar V de Aiken con intervalo de confianza.]**

Puntuación. La puntuación total se obtiene por suma directa. El reactivo 2 (“Me he sentido satisfecho/a con mi vida”) está redactado en dirección positiva y requiere recodificación inversa; este aspecto resultó determinante en los análisis y se detalla en Resultados.

1.4 Procedimiento

Los participantes fueron contactados en contextos académicos universitarios. Se explicaron los objetivos del estudio y se solicitó consentimiento informado. La aplicación se realizó mediante un formulario digital autoadministrado. Las respuestas fueron almacenadas con código de identificación y sin datos identificatorios directos en la base de análisis. Se estableció un procedimiento de atención al riesgo para participantes con puntuaciones elevadas en los reactivos referidos a ideación suicida y autolesión, consistente en la provisión de información sobre líneas de ayuda y servicios de salud mental disponibles.

[COMPLETAR: indicar (a) si la información de recursos se entregó a todos los participantes o solo a quienes superaron el umbral; (b) el umbral operativo exacto; (c) el mecanismo que preservó la trazabilidad sin comprometer la confidencialidad; y (d) el resultado de esa actuación. En los datos analizados, 33 participantes (47.1 %) puntuaron 3 o más en el reactivo “He pensado que mi vida no tiene sentido” y 23 (32.9 %) en “He tenido pensamientos de hacerme daño”.]

1.5 Análisis de datos

Los análisis se realizaron en Python 3.12 mediante las librerías `factor_analyzer`, `semopy`, `numpy`, `pandas` y `scipy`, siguiendo esta secuencia: (a) análisis descriptivo por reactivo (media, desviación estándar, asimetría, curtosis y prueba de Shapiro-Wilk); (b) correlación ítem-total corregida, con criterio $r \geq .30$ (Nunnally & Bernstein, 1994); (c) medida de adecuación muestral de Kaiser-Meyer-Olkin (Kaiser, 1974) y prueba de esfericidad de Bartlett (1954); (d) determinación del número de factores mediante análisis paralelo de Horn (1965) con 1,000 matrices aleatorias y criterio del percentil 95, procedimiento recomendado por sobre la regla de autovalores mayores que uno (Watkins, 2018); (e) análisis factorial exploratorio con extracción de máxima verosimilitud y rotación oblimin, elegida en lugar de componentes principales por modelar varianza común y permitir la derivación de índices de ajuste (Costello & Osborne, 2005); (f) consistencia interna mediante alfa de Cronbach (1951) y omega de McDonald (1999); (g) análisis factorial confirmato-

rio con índices *CFI*, *TLI*, *RMSEA* y *SRMR*, interpretados según Hu y Bentler (1999); y (h) análisis descriptivo de puntuaciones y comparación por género mediante prueba *t* de Welch con estimación del tamaño del efecto.

1.6 Consideraciones éticas

El estudio se rigió por los principios de la Declaración de Helsinki. Todos los participantes otorgaron consentimiento informado previo, con información sobre objetivos, procedimientos, voluntariedad, derecho de retiro sin consecuencias y tratamiento confidencial de sus respuestas. Los datos fueron codificados y analizados de forma agregada. **[COMPLETAR: código de dictamen del Comité de Ética en Investigación de la Universidad San Ignacio de Loyola y fecha. Si la recolección precedió al dictamen, declararlo con transparencia y describir el estudio como fase de desarrollo de instrumento.]**

2. Resultados

2.1 Análisis descriptivo de los reactivos

Las medias de los 22 reactivos oscilaron entre 1.97 y 3.36, con desviaciones estándar entre 0.94 y 1.24. Todos emplearon el rango completo de respuesta. Los valores de asimetría se situaron entre -0.47 y $+0.85$ y los de curtosis entre -0.88 y $+0.43$, dentro del rango convencionalmente aceptable para estimadores paramétricos. La prueba de Shapiro-Wilk resultó significativa para todos los reactivos ($p < .001$), resultado esperable en variables ordinales de cinco categorías.

Los reactivos con menor media correspondieron al dominio de conductas de riesgo: “He usado alcohol o drogas para manejar mis emociones” ($M = 1.97$, $DE = 1.18$) y “He tenido pensamientos de hacerme daño” ($M = 2.00$, $DE = 1.19$). El de mayor media fue “Me he sentido con poca energía o cansancio constante” ($M = 3.24$, $DE = 1.06$).

2.2 Análisis de reactivos y detección de un error de codificación

El análisis de correlaciones ítem-total corregidas sobre la versión de 22 reactivos reveló un patrón anómalo en el reactivo 2 (“Me he sentido satisfecho/a con mi vida”), cuya correlación con el resto

de la escala fue negativa ($r = -.160$), frente a valores entre .342 y .758 en los 21 restantes. El examen de la matriz de correlaciones confirmó que este reactivo correlacionaba negativamente con la práctica totalidad de los demás (rango $-.340$ a $+.099$).

Este resultado es atribuible a que el reactivo 2 es el único redactado en dirección positiva y no fue recodificado inversamente en el procesamiento de las respuestas. Tras aplicar la recodificación, la correlación ítem-total pasó a ser positiva pero permaneció por debajo del criterio mínimo ($r = .160 < .30$). Adicionalmente, su medida de adecuación muestral individual fue insuficiente ($KMO = .437$, frente a un mínimo de .703 en el resto). Sobre esta base se decidió eliminar el reactivo 2, dando lugar a la versión EEGSC-21. La eliminación incrementó la consistencia interna total de $\alpha = .925$ a $\alpha = .929$. Las correlaciones ítem-total de la versión final se presentan en la Tabla 2.

2.3 Adecuación muestral y dimensionalidad

Los indicadores de adecuación resultaron satisfactorios: KMO global = .842, con valores individuales entre .703 y .930; la prueba de esfericidad de Bartlett fue significativa, $\chi^2(210) = 910.90$, $p < .001$.

El análisis paralelo de Horn con 1,000 matrices aleatorias simuladas ($n = 70$, $k = 21$) indicó la retención de un único componente (Tabla 3). El primer autovalor observado (8.93) superó ampliamente el percentil 95 de la distribución aleatoria (2.38), mientras que el segundo (1.89) no alcanzó su valor de corte (2.07).

Cabe señalar que la regla clásica de autovalores mayores que uno habría sugerido retener cinco componentes. La discrepancia entre criterios es un fenómeno documentado, y el análisis paralelo constituye actualmente el procedimiento con mejor desempeño en estudios de simulación (Watkins, 2018), razón por la cual se privilegia su resultado. En términos de varianza, el primer componente da cuenta del 42.53 % de la varianza total y los tres primeros del 59.32 % acumulado. Al extraer por máxima verosimilitud, la solución de tres factores explica el 47.98 % de la varianza común ($F1 = 28.51$ %; $F2 = 12.53$ %; $F3 = 6.94$ %).

2.4 Solución factorial exploratoria

Con fines descriptivos se examinó la solución de tres factores con rotación oblimin. El primer factor agrupó reactivos de ánimo, energía, cognición, relaciones, conductas de riesgo y autoimagen, con cargas entre .387 y .921. El segundo quedó definido de forma nítida por los cuatro reactivos de ansiedad y estrés (cargas .613 a .851), constituyendo la única agrupación temática que se replicó como factor diferenciado. El tercero no alcanzó definición interpretable. Estos resultados indican que la única distinción empíricamente sostenible dentro de la EEGSC es la que separa el dominio ansioso del resto del constructo; las siete agrupaciones propuestas en la fase de construcción no encuentran respaldo en la estructura factorial observada.

2.5 Confiabilidad

La consistencia interna de la puntuación total de la EEGSC-21 fue elevada: $\alpha = .929$ y $\omega = .931$. Los valores por agrupación temática se presentan en la Tabla 4. Las correlaciones entre agrupaciones oscilaron entre .39 (AE–CR) y .69 (CE–PC y PC–AC), magnitudes moderadas a altas coherentes con la presencia de un factor general subyacente y que, simultáneamente, limitan la interpretación de las agrupaciones como dimensiones independientes.

2.6 Ajuste de los modelos de medida

Se estimaron tres modelos de medida mediante análisis factorial confirmatorio (Tabla 5). Ninguno alcanzó los criterios convencionales de ajuste aceptable (Hu & Bentler, 1999). Aunque el modelo temático de seis factores obtiene los mejores índices relativos, sus valores absolutos son insuficientes para sostener una afirmación de validez estructural. Este resultado debe interpretarse considerando que el tamaño muestral ($N = 70$, esto es, 3.3 participantes por reactivo) se sitúa muy por debajo de las recomendaciones habituales, lo que produce estimaciones inestables e índices sesgados a la baja. El presente estudio no permite, por tanto, confirmar ni descartar la estructura interna del instrumento; permite únicamente constatar que la evidencia disponible es insuficiente.

2.7 Distribución de puntuaciones y diferencias por género

La puntuación total de la EEGSC-21 (rango teórico 21–105) presentó una distribución aproximadamente simétrica (asimetría = 0.02), con $M = 58.49$, $DE = 14.69$, y valores observados entre 24 y 91. Los percentiles fueron: $P_{25} = 48.2$; $P_{50} = 58.0$; $P_{75} = 67.8$; $P_{90} = 77.0$; $P_{95} = 82.7$. No se observaron diferencias significativas entre mujeres ($n = 44$, $M = 58.66$) y varones ($n = 26$, $M = 58.19$), $t(56.4) = 0.126$, $p = .900$, $d = 0.03$.

Estos percentiles son descriptivos de la muestra estudiada y no constituyen normas ni puntos de corte. Su establecimiento requiere un diseño con criterio externo —típicamente una entrevista diagnóstica estructurada— y análisis de curvas ROC con reporte de sensibilidad y especificidad, procedimiento no realizado en este estudio.

3. Discusión

El presente estudio reporta la fase inicial de desarrollo de la EEGSC y arroja tres hallazgos principales.

En primer lugar, la detección y corrección de un error de codificación. El análisis de reactivos identificó que el único ítem redactado en dirección positiva no había sido recodificado inversamente, produciendo una correlación ítem-total negativa. Este hallazgo tiene una implicación práctica que trasciende el estudio: cualquier implementación previa del instrumento que haya calculado puntuaciones totales sin esa recodificación habrá generado valores incorrectos. Cabe señalar, además, que tras la recodificación el reactivo tampoco alcanzó los criterios de retención, lo que sugiere que el problema no era únicamente de procesamiento sino también de funcionamiento del reactivo: la satisfacción vital, formulada de manera global y en dirección positiva, parece comportarse como un constructo suficientemente distinto del malestar sintomático muestreado por el resto de la escala.

En segundo lugar, la evidencia de una estructura dominada por un factor general. El análisis paralelo retuvo un solo componente, que explica el 42.53 % de la varianza. Este resultado, que podría interpretarse como un fracaso de la propuesta multidimensional original, resulta congruente

con el marco teórico transdiagnóstico que orientó la construcción del instrumento: tanto el modelo del factor p (Caspi et al., 2014) como HiTOP (Kotov et al., 2017) predicen que un muestreo amplio del espectro sintomatológico producirá un primer factor de gran magnitud. La única agrupación que emergió como factor diferenciado fue la de ansiedad y estrés, patrón consistente con la distinción entre distrés general y activación ansiosa documentada en la literatura. La implicación operativa es directa: con la evidencia actual, la EEGSC-21 debe interpretarse mediante su puntuación total y no mediante perfiles de subescalas.

En tercer lugar, la insuficiencia del tamaño muestral para conclusiones estructurales. Con 70 participantes y 21 reactivos, ninguno de los modelos confirmatorios alcanzó umbrales aceptables. Es importante distinguir dos afirmaciones: los datos no permiten sostener que la estructura propuesta sea válida, pero tampoco permiten concluir que sea inválida. La conclusión correcta es de insuficiencia de evidencia, y la vía de resolución es la replicación con potencia adecuada.

Frente a estas limitaciones, los indicadores de consistencia interna y homogeneidad de reactivos sí resultan sólidos: todas las correlaciones ítem-total superan el criterio convencional, el *KMO* es meritorio y las agrupaciones con dos o más reactivos presentan alfas entre .72 y .85. Estos resultados justifican la continuación del desarrollo del instrumento.

3.1 Limitaciones

(a) $N = 70$ resulta insuficiente para análisis factorial confirmatorio y produce intervalos de confianza amplios; (b) el muestreo por conveniencia en población universitaria impide generalizar a población clínica, adultos mayores o adolescentes; (c) no se administraron instrumentos de referencia de forma concurrente, por lo que no hay evidencias de validez convergente ni discriminante; (d) el diseño transversal impide estimar confiabilidad test-retest; (e) no se dispone de criterio externo diagnóstico, por lo que no existen puntos de corte; (f) la agrupación Autoimagen y Control está compuesta por un solo reactivo, lo que impide estimar su confiabilidad; y (g) no se examinó invarianza de medición entre subgrupos.

3.2 Recomendaciones para estudios posteriores

Se recomienda: (a) replicación en muestra de $N \geq 250$ con análisis factorial confirmatorio y estimadores robustos para datos ordinales (WLSMV); (b) administración concurrente de PHQ-9 y GAD-7 para establecer validez convergente, con hipótesis específicas de $r \geq .60$ entre EEGSC total y PHQ-9, y entre la agrupación AE y GAD-7; (c) estudio test-retest a dos semanas con coeficiente de correlación intraclase; (d) ampliación de la agrupación Autoimagen y Control a un mínimo de tres reactivos, o su eliminación; (e) revisión del reactivo de consumo de sustancias ($r = .346$); (f) establecimiento de puntos de corte mediante criterio clínico externo y análisis ROC; y (g) análisis de invarianza de medición por género y edad.

3.3 Implicaciones para el uso aplicado

Mientras no se completen las fases descritas, el uso de la EEGSC-21 debe restringirse a fines de investigación y a funciones de autoconocimiento no diagnóstico. Toda implementación en entornos digitales debe informar explícitamente que se trata de un instrumento en desarrollo y no de una herramienta diagnóstica, evitar la presentación de categorías de severidad e incorporar un protocolo activo de derivación ante respuestas indicativas de riesgo que se active en el momento de la respuesta. Los datos del presente estudio son ilustrativos al respecto: casi la mitad de la muestra puntuó en el rango medio o superior del reactivo referido a pérdida de sentido vital, y aproximadamente un tercio en el de pensamientos de autolesión. En contextos de aplicación masiva y no supervisada, la frecuencia esperable de respuestas de riesgo hace imprescindible que el protocolo de derivación sea automático, inmediato y verificable.

4. Conclusiones

La EEGSC-21 constituye una propuesta de instrumento breve de tamizaje de salud mental para autoaplicación digital en población hispanohablante, que en su fase piloto muestra evidencias sólidas de consistencia interna ($\alpha = .929$; $\omega = .931$), homogeneidad de reactivos y adecuación muestral para el análisis factorial ($KMO = .842$). Su estructura interna aparece dominada por un factor

general, hallazgo congruente con los modelos transdiagnósticos contemporáneos, y no respalda la interpretación mediante siete subescalas propuesta en la fase de construcción. Los índices de ajuste de los modelos confirmatorios resultaron insuficientes, si bien este resultado es atribuible en parte al tamaño muestral. El instrumento debe considerarse en fase de desarrollo: su uso interpretativo individual, la asignación de categorías de severidad y cualquier aplicación con finalidad de tamizaje clínico requieren, de manera previa e ineludible, un estudio de validación con $N \geq 250$, evidencias de validez convergente, estabilidad temporal y puntos de corte establecidos frente a criterio clínico externo.

Referencias

- Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society, Series B*, 16(2), 296–298.
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). SAGE Publications.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Psychology Foundation of Australia.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Organización Mundial de la Salud. (2022). *World mental health report: Transforming mental health for all*. OMS.

- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097.
- Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20(3), 318–335.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246.

Table 1*Características de la muestra (N = 70)*

Variable	Categoría	<i>n</i>	%	<i>M (DE)</i>
Edad	—	70	100.0	21.56 (3.74)
	17–20 años	27	38.6	
	21–25 años	34	48.6	
	26–34 años	9	12.9	
Género	Femenino	44	62.9	
	Masculino	26	37.1	
Ciclo académico	—	70	100.0	5.31 (2.82)

Nota. Rango de edad: 17–34 años. Rango de ciclo académico: 1–12.

Table 2*Correlaciones ítem-total corregidas de la EEGSC-21*

Reactivo	Dominio	<i>r</i>	Reactivo	Dominio	<i>r</i>
eg1	EG	.757	pc2	PC	.673
eg3	EG	.553	pc3	PC	.689
ae1	AE	.533	re1	RE	.466
ae2	AE	.542	re2	RE	.570
ae3	AE	.563	re3	RE	.539
ae4	AE	.585	cr1	CR	.719
ce1	CE	.568	cr2	CR	.609
ce2	CE	.654	cr3	CR	.346
ce3	CE	.700	cr4	CR	.564
ce4	CE	.735	ac1	AC	.700
pc1	PC	.529			

Nota. Todos los valores superan el criterio $r \geq .30$. El reactivo cr3 (“He usado alcohol o drogas para manejar mis emociones”) presenta el valor más bajo y requiere seguimiento en estudios posteriores.

Table 3*Análisis paralelo de Horn (EEGSC-21)*

Componente	Autovalor observado	Percentil 95 aleatorio	Decisión
1	8.932	2.378	Retener
2	1.887	2.071	No retener
3	1.638	1.868	No retener
4	1.307	1.705	No retener
5	1.039	1.571	No retener

Nota. Basado en 1,000 matrices aleatorias simuladas con $n = 70$ y $k = 21$.

Table 4*Consistencia interna y descriptivos de las agrupaciones temáticas*

Agrupación	k	M	DE	α
Estado General (EG)	2	5.43	1.83	.723
Ansiedad y Estrés (AE)	4	11.61	3.53	.837
Conducta y Energía (CE)	4	12.37	3.60	.847
Pensamientos y Cognición (PC)	3	8.81	2.47	.722
Relaciones y Entorno (RE)	3	8.31	2.49	.732
Conductas de Riesgo (CR)	4	8.89	3.75	.817
Autoimagen y Control (AC)	1	3.06	1.24	—
EEGSC-21 total	21	58.49	14.69	.929

Nota. EG quedó reducida a dos reactivos tras la eliminación del reactivo 2. AC está compuesta por un único reactivo, por lo que no admite estimación de consistencia interna y no constituye una subescala en sentido estricto.

Table 5*Índices de ajuste de los modelos de medida evaluados ($N = 70$)*

Modelo	χ^2	gl	CFI	TLI	$RMSEA$	$SRMR$
Unifactorial (21 reactivos)	457.25	189	.678	.642	.143	.145
Tres factores (empírico)	401.47	186	.741	.708	.130	—
Seis factores (temático)	312.13	155	.799	.753	.121	.145

Nota. El modelo de seis factores excluye la agrupación AC por estar compuesta de un solo reactivo. Todos los modelos presentan χ^2 significativo ($p < .001$). Criterios de referencia (Hu & Bentler, 1999): CFI y $TLI \geq .95$; $RMSEA \leq .06$; $SRMR \leq .08$.

Apéndice A**Instrumento EEGSC-21**

Instrucciones. Indique con qué frecuencia ha experimentado cada situación durante las últimas dos semanas. No hay respuestas correctas ni incorrectas.

Escala: 1 = Nunca; 2 = Casi nunca; 3 = A veces; 4 = Casi siempre; 5 = Siempre

N.º	Código	Reactivo
1	eg1	Me he sentido triste, vacío/a o sin esperanza.
2	eg3	He tenido dificultad para disfrutar de las cosas que antes me gustaban.
3	ae1	Me he sentido nervioso/a o tenso/a.
4	ae2	Me preocupa demasiado por cosas cotidianas.
5	ae3	He tenido pensamientos que se repiten y no puedo controlar.
6	ae4	He tenido ataques de pánico o síntomas de ansiedad física (palpitaciones, mareos, falta de aire).
7	ce1	He tenido cambios en mi apetito (comer más o menos de lo normal).
8	ce2	He tenido problemas de sueño (insomnio o dormir en exceso).
9	ce3	Me he sentido con poca energía o cansancio constante.
10	ce4	He perdido interés o motivación para realizar actividades.
11	pc1	Me ha costado concentrarme en mis tareas.
12	pc2	He tenido dificultades para tomar decisiones.
13	pc3	Me he sentido inútil, culpable o sin valor.
14	re1	He tenido conflictos con familiares o amigos.
15	re2	He sentido falta de apoyo de las personas cercanas.
16	re3	He evitado lugares o personas porque me generan malestar.
17	cr1	He pensado que mi vida no tiene sentido.
18	cr2	He tenido pensamientos de hacerme daño.
19	cr3	He usado alcohol o drogas para manejar mis emociones.
20	cr4	He actuado de manera impulsiva o arriesgada.
21	ac1	Me he sentido insatisfecho/a con mi cuerpo o apariencia.

Puntuación. Suma directa de los 21 reactivos (rango 21–105). No existen reactivos de puntuación inversa en esta versión. No se dispone de puntos de corte validados; la puntuación total no debe interpretarse como indicador de severidad clínica.

Atención al riesgo. Una respuesta de 3 o superior en los reactivos 17 o 18 debe activar de forma inmediata la presentación de recursos de ayuda, con independencia de la puntuación total.