

Francisco Rafael de la Peña Olvera <sup>1</sup>   
Andrés Rodríguez Delgado <sup>1</sup>   
Alejandro Irvin Soto Briseño <sup>1</sup>   
Assad Daniel Saad Manzanera <sup>1</sup> \* 

<sup>1</sup> Instituto Nacional de Psiquiatría “Ramón de la Fuente Muñiz”,  
Ciudad de México, México.

\* AUTOR DE CORRESPONDENCIA

Assad Daniel Saad Manzanera  
Correo: [assadsm@imp.edu.mx](mailto:assadsm@imp.edu.mx)

*Estudio de validez y confiabilidad de la escala de irritabilidad: ARI (Índice de Reactividad Afectiva) en población clínica de adolescentes mexicanos*

*Validity and reliability study of the irritability scale: ARI (Affective Reactivity Index) in a clinical population of Mexican adolescents*

---

## RESUMEN

**Introducción:** La irritabilidad crónica (IC) es un constructo transdiagnóstico. Se ha estudiado a través del Índice de Reactividad Afectiva (ARI) el cual es una de las más comúnmente usadas y estudiadas a nivel global. **Objetivos:** Determinar la consistencia interna, la validez del constructo, convergente y divergente, la fiabilidad test-retest y la fiabilidad interinformante, y establecer la prueba diagnóstica para ambas versiones de la escala ARI. **Material y métodos:** Se evaluaron adolescentes y sus padres, se entrevistaron semiestructuradamente (K-SADS-PL-5) y se completaron los instrumentos para los procesos de validez con los adolescentes y sus padres. Se utilizaron contrastes con  $\chi^2$ , correlaciones con r de Pearson y se obtuvieron los valores de la prueba diagnóstica. **Resultados:** Se evaluaron 83 adolescentes y sus tutores. Encontramos una alfa de Cronbach de 0,918 en versión ARI adolescentes (ARI-A) y de 0,904 para ARI padre/tutor (ARI-P). El análisis de componentes principales determinó un solo factor explicando una varianza de 79% en ARI-A y 75% en ARI-P. Se encontró fiabilidad test-retest de 0,42-0,67 en ARI-A y 0,35-0,53 en ARI-P. La fiabilidad interinformante global fue de K= 0,24, p=0,002. Las áreas bajo la curva fueron 0,640 para ARI-A y 0,753 para ARI-P. **Conclusión:** La escala ARI presenta adecuada validez y confiabilidad en la medición de IC en adolescentes mexicanos con psicopatología.

---

## ABSTRACT

**Introduction:** Chronic irritability (CI) is a transdiagnostic construct. It has been studied through the Affective Reactivity Index (ARI) instrument, which is one of the most used and studied globally. **Objectives:** To determine internal consistency, construct validity, convergent validity and divergent validity, test-retest reliability, and inter-informant reliability, and establish the diagnostic test for both versions of the ARI scale. **Materials and methods:** Adolescents and their parents were assessed, they were interviewed with a semi-structured instrument (K-SADS-PL-5) and answered to different instruments for process validity. Contrasts with  $\chi^2$ , correlations with Pearson r were used and diagnostic test values were obtained. **Results:** 83 adolescents and their parents were included. We found a Cronbach's alpha of 0,918 for ARI adolescents (ARI-A) and 0,904 for ARI parent/guardian (ARI-P). The principal components analysis determined a single factor explaining a variance of 79% in ARI-A and 75% in ARI-P. Test-retest reliability of 0,42-0,67 was found in ARI-A and 0,35-0,53 in ARI-P; interinformant reliability was significant K= 0,24, p=0,002. The areas under the curve were 0,640 for ARI-A and 0,753 for ARI-P. **Conclusion:** The ARI scale presents adequate validity and reliability in measuring CI in Mexican adolescents with psychopathology.

**10** **Palabras clave:** psicopatología, irritabilidad, psicometría, psiquiatría adolescente, resultados de reproducibilidad.

**Keywords:** Psychopathology, Irritable Mood, Psychometrics, Adolescent Psychiatry, Reproducibility of Results.

## INTRODUCCIÓN

La irritabilidad es un constructo transdiagnóstico, abarca al menos 15 diagnósticos dentro de la quinta edición del Manual Diagnóstico y Estadístico para Trastornos Mentales de la Asociación Psiquiatría Americana (DSM-5, por sus siglas en inglés) (1). A pesar de la extensión de este fenómeno, el DSM no lo define explícitamente; ha sido difícil de caracterizar y definir en varios ámbitos ya que forma parte de constructos relacionados como el enojo, la agresividad, la hostilidad y los estados depresivos, por lo que recientemente se ha propuesto la siguiente definición: “La irritabilidad es un estado de ánimo de agitación fisiológica parcial caracterizado por un aumento de la sensibilidad a los estímulos sensoriales y una disminución no mediada por la cognición del umbral para responder con ira y/o agresividad a estímulos normalmente menos molestos; está causada por factores que afectan directamente a la fisiología y/o la biología, como el hambre, la falta de sueño, el dolor y la fatiga” (2). Con frecuencia, los niños sin psicopatología presentan diversos periodos de irritabilidad, particularmente durante algunas fases específicas del desarrollo humano; sin embargo, su presencia recurrente, disruptiva y fuera de lo esperado al desarrollo, representa una condición clínica para la salud mental (3,4).

Desde hace más de una década se tiene evidencia que los niños y adolescentes que presentan irritabilidad crónica (IC) en los diferentes trastornos psiquiátricos, tienen diferencias importantes en varios correlatos clínicos (5). La IC presente a edades tempranas está asociada a desenlaces depresivos subsecuentes (6,7), por lo que su identificación temprana es importante. El estudio de un instrumento que pueda evaluar de forma consistente la IC permitiría una identificación temprana de este problema (8). El

Índice de Reactividad Afectiva (ARI, por sus siglas en inglés de *Affective Reactivity Index*), es uno de los instrumentos más usados en todo el mundo (5) y es el instrumento recomendado por DSM-5 para la medición transversal de los síntomas de irritabilidad crónica en niños y adolescentes (9).

El ARI fue diseñado por Stringaris en 2012, momento en que el autor realizó un estudio con población clínica y comunitaria de Estados Unidos de América (EUA) y del Reino Unido. Es un instrumento conciso de 7 ítems en formato de autorreporte en dos versiones para el padre/tutor (ARI-P) y para el niño/adolescente (ARI-A) (8). Este instrumento conceptualiza y operacionaliza la IC de forma breve y sencilla, como una tendencia a ser molestado o presentar labilidad emocional, ira y/o rabietas. La irritabilidad es medida durante los últimos 6 meses, lo cual es considerado como una característica de cronicidad. Examina 3 aspectos de irritabilidad: umbral para reaccionar con enojo, frecuencia de afecto/conducta de enojo y duración de éstos últimos. La escala está diseñada para su uso como instrumento de cribaje en la práctica clínica y en estudios epidemiológicos, se ha traducido a más de 15 lenguajes y ha sido probada en población clínica y comunitaria en varios países: Australia (10), Brasil (11), China (12), Italia (13), España (14) y EUA (15), en los que se ha encontrado propiedades psicométricas similares en cuanto a estructura de un solo factor y validez interna.

El estudio llevado a cabo en Australia incluyó adultos y adolescentes, encontró propiedades psicométricas similares al estudio original de la escala y también describieron correlaciones positivas con síntomas de depresión y ansiedad incluyendo ansiedad social (10). El estudio en Brasil fue constituido por población escolar y encontró buena consistencia interna y con estructura unifactorial,

además describieron mayor correlación de ARI con la dimensión de ansiedad relacionadas a distrés que la correlación con fobias, así como mayores puntajes de disfunción (11). El estudio en China se integró por población clínica ambulatoria, y encontró que la medición de la irritabilidad se asoció con comportamientos agresivos, síntomas ansiosos, depresivos y problemas con pares (12). El estudio en Italia incluyó adultos y adolescentes y encontró adecuada consistencia interna y el análisis factorial apoya la estructura unifactorial (13). El estudio en España estudió el ARI respondido por los profesores y reportó una adecuada consistencia interna y asociaciones con otras escalas, mas no se aplicaron las escalas ARI-P ni ARI-A (14). El estudio de los EUA incluyó a 237 niños y adolescentes, pacientes ambulatorios referidos a terapia, este estudio arrojó un solo factor en el análisis del constructo y la irritabilidad se asoció con problemas internalizados, externalizados y la regulación emocional (15).

Tomando en cuenta la necesidad de contar con un instrumento válido y confiable en población hispanoparlante en las versiones ARI-P y ARI-A, se propuso el estudio de validez y confiabilidad de este instrumento. El objetivo del presente estudio fue determinar la consistencia interna, la validez del constructo, validez de criterio externo, divergente y convergente, la fiabilidad test-retest, la interinformante, así como la determinación de la sensibilidad y especificidad como prueba diagnóstica.

## MATERIAL Y MÉTODOS

El presente fue un estudio de proceso, comparativo y en población de adolescentes y sus padres.

### Población

Se incluyeron adolescentes admitidos para atención y tratamiento externo en la Clínica de la Adolescencia del Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRFM) en la Ciudad de México, con edades de 13 a 17 años. Se excluyeron aquellos que no supieran leer y/o escribir, que presentaran algún grado de discapacidad intelectual, o síntomas psicóticos. Se eliminaron los pacientes que realizaran llenado incompleto de escalas auto-aplicables.

## Consideraciones éticas

El estudio se llevó a cabo con aprobación del Comité de Ética en Investigación del INPRFM (Ref: CEI/C/68/2014) siguiendo todos los principios para la investigación en humanos establecidos en la declaración de Helsinki (16). Se obtuvo el consentimiento/asentimiento informado por escrito de cada uno de los padres/tutores y adolescentes, respectivamente.

## PROCEDIMIENTO

Después de la obtención del consentimiento/asentimiento tanto los padres como los adolescentes completaron las diferentes escalas. En la misma entrevista el investigador principal realizó la evaluación para los apartados del Trastorno Disruptivo de la Regulación Emocional (TDRE) y del Trastorno Oposicionista Desafiante (TOD) de la entrevista del K-SADS-PL-5 (Por sus siglas en inglés *Kiddie-Schedule for Affective Disorders and Schizophrenia Present and Lifetime version*) versión del DSM-5 (17).

Una semana después los psiquiatras certificados adscritos del servicio establecieron el diagnóstico principal mediante una entrevista clínica utilizando la Escala Breve de Apreciación Psiquiátrica para Niños y Adolescentes (BPRS-CA-29, por sus siglas en inglés de *Brief Psychiatric Rating Scale for Children & Adolescents* versión de 29 ítems) (18). En esta misma consulta se realizaron el re-test del ARI-P y del ARI-A.

## INSTRUMENTOS

### ARI

El ARI es una escala breve que consta de 6 ítems que evalúan irritabilidad crónica e incluye un reactivo al respecto del deterioro en el funcionamiento debido a esta en ambas versiones. Los participantes responden sobre 3 puntos en una escala tipo Likert (0 = no es cierto, 1= algo cierto, 2 = totalmente cierto) al respecto de los últimos 6 meses. La versión ARI-P y ARI-A son similares, sólo cambia la redacción de las preguntas, por ejemplo: “la conducta o sentimientos de su hijo” para el caso del ARI-P y “tu conducta

12

o sentimientos” para el caso de la versión ARI-A. Se usó una traducción al español latinoamericano realizada por la Dra. Andrea Abadí en Buenos Aires, Argentina y avalada por el equipo de Stringaris (19), la cual fue obtenida del repositorio en internet y se anexa al final de este manuscrito.

Recientemente se reexaminó psicométricamente las propiedades de la escala ARI en su lenguaje original en una muestra clínica independiente, demostrando buena consistencia interna para la ARI-P (rango alfa/omega 0,88-0,90) y para la ARI-A (rango alfa/omega 0,79-0,90), y una estructura de factor unidimensional tanto el modelo específico de informante y el multi-informante (CFI/TLI >0,95 y >0,98, respectivamente) (15). En este estudio la irritabilidad reportada por padres tenía mayor consistencia que la reportada por niños. Al evaluar la validez convergente de la IC mediante el ARI, se encontró una asociación significativa con variables internalizadas y externalizadas, especialmente con la agresión reactiva, el enojo, la desregulación y los problemas de afrontamiento (15). Mulraney reporta en una pequeña muestra de adolescentes que un punto de corte de 4 en ARI-A y de 3 en ARI-P podría ser un indicador de presencia de psicopatología en general, mas no se ha descrito otro punto de corte con esta escala (20).

Usando la teoría de la respuesta a ítems, se ha encontrado que el ARI es la mejor medición para evaluar la irritabilidad infantil a lo largo de diferentes periodos, pero pierde confiabilidad con niveles más graves de irritabilidad (21). Un estudio psicométrico reciente de análisis de redes reportó que los ítems centrales de la escala eran el reactivo seis “Pierdo la paciencia fácilmente,” cinco “Me enojo fácilmente,” y dos “Pierdo la paciencia frecuentemente”; y que los reactivos tres “Permanezco enojado mucho tiempo” y cuatro “Estoy enojado la mayor parte del tiempo” tienen la capacidad de capturar los niveles más graves de irritabilidad (22). La versión ARI-P también se ha estudiado en población infantil de 3 a 8 años y se ha observado estructura de factor único (23,24).

### Instrumentos para adolescentes

La escala de autorreporte para depresión de Birlson (DSRS, por las siglas en inglés de *Depression*

*Self Rating Scale*) es un instrumento de autorreporte de tipo Likert que consta de 18 reactivos que pueden puntuar de cero al dos, siendo la máxima calificación de 36 (25). La utilización de esta escala como prueba diagnóstica tiene una sensibilidad de 87% y una especificidad de 74% cuando el punto de corte es de 14 puntos. Se ha traducido y validado al español en población mexicana, con una adecuada consistencia interna (alfa de Cronbach=0,85) (26), la traducción se encuentra como anexo en dicha publicación.

El SCARED (SCARED, por las siglas en inglés de *Screen for Child Anxiety Related Emotional Disorders*) (27) es un instrumento de tamizaje para varios trastornos de ansiedad, consta de 41 reactivos y se han reportado 5 factores: somático/pánico, ansiedad generalizada, ansiedad de separación, fobia social y fobia escolar. Se ha observado que el SCARED cuenta con una buena consistencia interna, con alfas de Cronbach de 0,74 a 0,93 y coeficientes de correlación intraclase en un rango de 0,70 a 0,90 (28). La validación de dicha escala en idioma español en población comunitaria mostró los mismos cinco factores y una consistencia interna general de 0,90.

### Instrumento para padres

El padre/tutor respondieron el cuestionario de Swanson, Nolan y Pelham IV (SNAP IV) versión padre/tutor (29). Es una escala que evalúa los síntomas del trastorno por déficit de atención con hiperactividad (TDAH), consta de dos subescalas: inatención e hiperactividad/impulsividad, cada una de las cuales considera nueve preguntas al respecto de cada dimensión de síntomas del cero a tres de acuerdo con la intensidad de dichos síntomas, los cuales son los considerados por el DSM-IV-TR para el diagnóstico de TDAH. Esta escala se ha usado comúnmente en estudios de tratamiento para TDAH (30) a pesar de que sus propiedades psicométricas no se han publicado detalladamente (31). Un estudio reciente de validación en niños concluyó que los puntajes mostraban alta sensibilidad pero baja especificidad para diagnóstico clínico y que puede ser una medida válida para desenlaces en estudios aleatorizados controlados y en entorno clínico (32). Usamos una versión adaptada al español en población argentina, arrojando una sensibilidad de

54% y una especificidad de 78% para inatención, mientras que para hiperactividad-impulsividad estos valores fueron de 86% y 73,5%, respectivamente (29).

### Entrevistas Diagnósticas

El K-SADS-PL-5 es una entrevista semiestructurada diseñada para el diagnóstico de varios trastornos psiquiátricos infantiles en el presente y a lo largo de la vida (33). Esta entrevista es un instrumento diagnóstico que ofrece ventajas como la disminución de la variabilidad en la recolección de la información y permite que los pacientes y sus padres participen en una evaluación integral de las emociones y la conducta del paciente, por lo que es considerada un *gold standard* diagnóstico para fines de investigación. Para este estudio se aplicaron los suplementos de TDRE y TOD del K-SADS-PL-5 versión en español para Latinoamérica; la confiabilidad interevaluador se reportó para el TDRE  $K=0.53$  (IC 0,34 - 0,71) y para el TOD  $K=0,80$  (IC 0,69 - 0,92) (34).

El instrumento BPRS-CA (35) es una entrevista semiestructurada comprehensiva que permite una evaluación de varios grupos de trastornos incluyendo trastornos de conducta, del control de los impulsos, depresivos, psicóticos, de inatención e hiperactividad, del lenguaje, de ansiedad, de estrés y del sueño. Se ha realizado validación al español en Colombia, encontrando alfa de Cronbach de 0,85 (36). Para este estudio se usó la versión mexicana compuesta por 29 reactivos, que se adaptó en la Clínica de la Adolescencia del INPRFM (37). Esta versión tiene cinco reactivos más que la versión original, los cuales evalúan trastornos de la eliminación, hipertimia, uso y abuso de alcohol, tabaco y otras drogas, falta de emociones prosociales, además de evaluar abuso psicológico y sexual. La confiabilidad ínter evaluador y test-retest es de  $r=0,824$  y  $r=0,661$ , respectivamente (37).

### Análisis estadístico

Este estudio se llevó a cabo utilizando pruebas estadísticas con una confianza del 95%. Para la descripción de los datos se utilizaron medidas de tendencia central. La consistencia interna se

estableció mediante la obtención de alfa de Cronbach. La validez de constructo se realizó a través de un análisis de componentes principales con objetivo de reducir dimensiones de una manera interpretable, con rotación ortogonal tipo varimax para simplificar la interpretación sin perder varianza (38), considerando que todos los estudios sobre la escala ARI han reportado sólo un factor. Se construyó una matriz de correlaciones mediante la obtención de un coeficiente de correlación policórico a través del método de extracción de mínimos cuadrados no ponderados. La validez de criterio externo se estableció a través del contraste de los diagnósticos de TDRE y de TOD obtenidos con K-SADS-PL-5 con tres diferentes puntos de corte de la escala ARI, utilizando tres diferentes cuartiles: 25, 50 y 75; por lo que hubieron diferencias en los puntos de corte para ARI-A (cuatro, seis y ocho puntos) y ARI-P (tres, seis, y diez puntos). Para determinar la validez divergente (SNAP-IV) y convergente (DSRS y SCARED) se usaron estos mismos puntos de corte mencionados de ARI-A y ARI-P; esto resulta en una operacionalización en categorías de la ARI frente a operacionalización continua de las otras escalas, por lo que se determinó usar  $\chi^2$  en lugar de las correlaciones de Pearson. Para la fiabilidad test-retest se establecieron  $r$  de Pearson mediante la realización de un test-retest una semana después de la evaluación inicial. La confiabilidad interinformante se determinó contrastando los datos obtenidos de las escalas aplicadas a los hijos con la de los padres y la obtención de coeficientes kappa. Los valores de prueba diagnóstica se determinaron estableciendo la sensibilidad, especificidad y curva ROC (por sus siglas en inglés de *Receptor Operating Characteristic*) del ARI.

### RESULTADOS

Se invitó a participar a 101 adolescentes, 18 individuos no aceptaron participar en el estudio. La muestra final se integró con 83 adolescentes; el 48,2% ( $n=40$ ) hombres y mujeres 51,8%, ( $n=43$ ). La edad media de los participantes fue de 15,4 años ( $DE=1,38$ ). El promedio de los valores del ARI en la versión adolescente fue de 6,65 y de 6,47 para el caso de la versión de padre/tutor.

## Características clínicas

El 50,6% (n=42) presentó como diagnóstico principal algún trastorno internalizado, en el 27,7% (n=23) un trastorno externalizado y el 21,6% (n=18) de los sujetos fueron clasificados en el grupo de psicopatología mixta (internalizado y externalizado). Los diagnósticos más frecuentes fueron: Trastorno depresivo mayor, 56,6% (n= 47); TDAH, 37% (n=31); trastorno de conducta 8,4% (n=7); trastorno de ansiedad generalizada 6% (n=5); trastorno obsesivo compulsivo 3,6%(n=3), entre otros. A pesar de que solo el 2,4% (n=2) de la muestra presentó TOD como diagnóstico principal, el 71,1% (n=60) de los adolescentes lo presentó como segundo o tercer diagnóstico. Así mismo el 10,8% (n=9) de los adolescentes cumplieron con el diagnóstico de TDRE independientemente del diagnóstico principal o del grupo de psicopatología en el cual se encontraban. El 8,4 % (n=7) de los adolescentes presentaban tanto TOD como TDRE.

## Consistencia interna

Se observó que la escala presenta una máxima varianza con un valor de alfa de Cronbach del ARI-A de 0,918 y para la versión ARI-P de 0,904. Además, se realizó un análisis suprimiendo en cada caso uno de los reactivos del ARI, el cual mostró que la supresión de alguno de los reactivos no mejoraba la consistencia interna de la escala como un todo.

## Validez de constructo

Para el análisis de componentes principales se consideraron los primeros 6 reactivos de ambas versiones de la escala ARI, excluyendo de esta manera el reactivo al respecto de deterioro en el funcionamiento. Se observó que la solución más adecuada en ambas versiones del ARI fue a través de un solo factor y que la varianza explicada mediante este es del 79% para ARI-A y del 75% para ARI-P. En la [Tabla 1](#) se muestran los valores de la carga factorial de cada uno de los reactivos del ARI para ambas versiones.

**Tabla 1.** Carga factorial de los primeros 6 reactivos del ARI (n=83).

| Reactivos ARI                           | ARI-A | ARI-P |
|---|-------|-------|
| Los demás me irritan fácilmente         | 0,76  | 0,81  |
| Pierdo la paciencia frecuentemente      | 0,80  | 0,79  |
| Permanezco enojado mucho tiempo         | 0,74  | 0,81  |
| Estoy enojado la mayor parte del tiempo | 0,81  | 0,95  |
| Me enoja frecuentemente                 | 0,88  | 0,96  |
| Pierdo la paciencia fácilmente          | 0,75  | 0,87  |

ARI = Affective Reactivity Index; ARI-A: Affective Reactivity Index versión adolescentes. ARI-P: Affective Reactivity Index versión padres.

## Validez de criterio externo

A través del cuartil 25, se estableció el punto de corte en 4 para la versión de adolescente y de 3 para la versión de padre/tutor; utilizando el cuartil 50 los puntos de corte se establecieron en 6 en ambos casos, y mediante el cuartil 75 los puntos de corte fueron establecidos en 8 y 10, respectivamente. En el caso de ARI-A se observó una correlación estadísticamente significativa con el diagnóstico de TOD cuando se utilizaron los puntos de corte en 6 ( $\chi^2=6,48$ ,  $p=0,01$ ) y 8 ( $\chi^2=4,96$ ,  $p=0,026$ ). Aunque no se observó correlación con el diagnóstico de TDRE, el análisis mostró una tendencia hacia la significancia estadística entre la ARI-A con punto de corte en 6 y el TDRE ( $\chi^2=3,72$ ,  $p=0,054$ ). En el caso del ARI-P, se observó una asociación estadísticamente significativa utilizando los 3 diferentes puntos de corte solamente con el diagnóstico de TOD (punto de corte en 3:  $\chi^2=10,09$ ,  $p=0,001$ ; punto de corte en 6:  $\chi^2=8,17$ ,  $p=0,004$ ; punto de corte en 10:  $\chi^2=12,18$ ,  $p=<0,001$ ).

## Validez divergente

No se observó correlación estadísticamente significativa del ARI en sus dos versiones con escala la de inatención o de hiperactividad de SNAP-IV. La escala ARI-P mostró una tendencia hacia la significancia estadística con punto de corte 10 para la escala de hiperactividad/impulsividad ( $\chi^2=2,733$ ,  $p=0,098$ ).

## Validez convergente

Se observó una correlación estadísticamente significativa entre la versión ARI-A y la escala DSRS utilizando un punto de corte en 4 ( $\chi^2=9,37$ ,  $p=0,002$ ) y con la escala SCARED utilizando los puntos de corte en 6 ( $\chi^2=13,08$ ,  $p<0,001$ ) y 8 ( $\chi^2=6,81$ ,  $p=0,009$ ). En el caso de la versión ARI-P, se observó una correlación estadísticamente significativa utilizando los puntos de corte de 3 ( $\chi^2=5,47$ ,  $p=0,017$ ) y 6 ( $\chi^2=6,25$ ,  $p=0,012$ ).

## Fiabilidad test-retest

De las 83 evaluaciones iniciales se logró obtener una segunda evaluación en 65 de ellas (78,3% de la muestra total). Los datos obtenidos del test-retest permitieron la obtención de un coeficiente de correlación de Pearson a través del cual se observó una correlación positiva entre todos los reactivos de la escala ARI. La correlación global con seis reactivos en la versión ARI-A fue  $r=0,53$  y en la versión ARI-P fue  $r=0,41$ , por lo que estas correlaciones se considerarían moderadas y débil, respectivamente. Los coeficientes de correlación de los diferentes ítems oscilaron entre 0,42 y 0,67 para la escala ARI-A, y entre 0,35 y 0,53 para la escala ARI-P. De todos los reactivos de la escala, el reactivo 3 (“Permanezco enojado mucho tiempo”) y el 4 (“Estoy enojado la mayor parte del tiempo”) en ambas versiones, presentaron las correlaciones más altas. En la [Tabla 2](#) se muestran las correlaciones entre los reactivos del ARI en ambas versiones.

**Tabla 2.** Confiabilidad Temporal de la escala ARI: Correlaciones entre los reactivos del ARI (test-retest) para ambas versiones.

| Reactivos del ARI                              | ARI-A | ARI-P |
|--|-------|-------|
| Los demás me irritan fácilmente                | 0,49  | 0,36  |
| Pierdo la paciencia frecuentemente             | 0,48  | 0,38  |
| Permanezco enojado mucho tiempo                | 0,67  | 0,44  |
| Estoy enojado la mayor parte del tiempo        | 0,66  | 0,53  |
| Me enojo frecuentemente                        | 0,43  | 0,41  |
| Pierdo la paciencia fácilmente                 | 0,42  | 0,35  |
| En general mi irritabilidad me causa problemas | 0,55  | 0,41  |

ARI = Affective Reactivity Index; ARI-A: Affective Reactivity Index versión adolescentes. ARI-P: Affective Reactivity Index versión padres.

## Confiabilidad interinformante

En la confiabilidad interinformante de la escala se encontró una correlación estadísticamente significativa entre ambos informantes, a excepción del reactivo 3 (“Permanezco enojado mucho tiempo”), el cual mostró una tendencia hacia la significancia estadística ( $K= 0,150$ ,  $p=0,059$ ) (Ver [Tabla 3](#)). La confiabilidad global fue de  $K= 0,24$ ,  $p=0,002$ .

**Tabla 3.** Confiabilidad interinformante de la escala ARI: Valores de coeficiente Kappa para cada reactivo del ARI.

| Reactivos del ARI                              | Valor K (valor p)         |
|--|---------------------------|
| Los demás me irritan fácilmente                | <b>0,261 (p=0,002)</b>    |
| Pierdo la paciencia frecuentemente             | <b>0,257 (p=0,002)</b>    |
| Permanezco enojado mucho tiempo                | 0,150 (p=0,059)           |
| Estoy enojado la mayor parte del tiempo        | <b>0,318 (p&lt;0,001)</b> |
| Me enojo frecuentemente                        | <b>0,302 (p&lt;0,001)</b> |
| Pierdo la paciencia fácilmente                 | <b>0,190 (p=0,018)</b>    |
| En general mi irritabilidad me causa problemas | <b>0,234 (p=0,001)</b>    |

ARI = Affective Reactivity Index, K = Valor Kappa

## Prueba diagnóstica

La obtención de la curva ROC (ver [Figura 1](#)) permitió establecer los diferentes puntos de corte para la escala ARI como instrumento de detección del TOD. Dicho análisis mostró que en ambas versiones el mayor equilibrio entre la sensibilidad y la especificidad se encontró a los tres puntos (es decir, al utilizar los puntos de corte entre 2,5 y 3,5) (ver [tabla 4](#)). El área bajo la curva del ARI-P fue 0,753 y para el ARI-A fue de 0,640.

## DISCUSIÓN

Los resultados de esta investigación en su mayoría son concordantes y similares con los obtenidos en otras poblaciones, y apoyan la validez del ARI-P y ARI-A para población adolescente mexicana además aportan nuevas observaciones como prueba diagnóstica. Nuestros hallazgos en población clínica mexicana son congruentes y similares con el resto de la literatura internacional publicada hasta el momento. Se demostraron consistencia interna

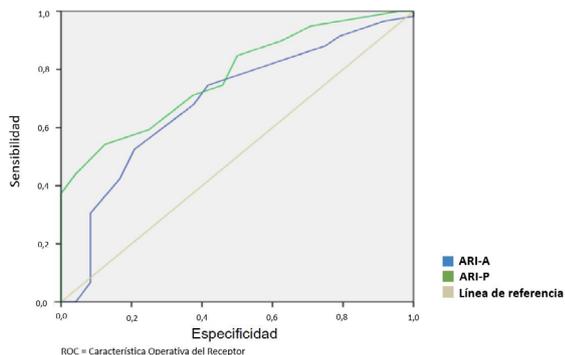
16

adecuada, con validez del constructo reportada en un solo factor. En cuanto a las asociaciones diagnósticas, se reportó validez de criterio externo asociada a TOD (asociándose a criterios diagnósticos de este diagnóstico cuando se puntúa arriba de 4) y convergencia con escalas de depresión y ansiedad, contrastándose con validez divergente de TDAH (ya que no hubo asociación diagnóstica con este diagnóstico). La fiabilidad test-retest fue adecuada y se encontró cierta correlación interinformante entre ARI-P y ARI-A.

**Tabla 4.** Prueba diagnóstica de la escala ARI para la detección de TOD: sensibilidad y especificidad de la ARI utilizando diferentes puntos de corte.

| ARI   | Positivo si $\geq$ | Sensibilidad | Especificidad |
|-------|--------------------|--------------|---------------|
| ARI-A | 2,5                | 0,915        | 0,792         |
|       | 3,5                | 0,881        | 0,75          |
|       | 4,5                | 0,746        | 0,417         |
|       | 5,5                | 0,678        | 0,375         |
| ARI-P | 2,5                | 0,898        | 0,625         |
|       | 3,5                | 0,847        | 0,50          |
|       | 4,5                | 0,746        | 0,458         |
|       | 5,5                | 0,712        | 0,375         |

TOD = Trastorno Oposicionista Desafiante, ARI = Affective Reactivity Index; ARI-A: Affective Reactivity Index versión adolescentes. ARI-P: Affective Reactivity Index versión padres.



**Figura 1.** Curva ROC de la escala ARI

### Características clínicas

La mitad de los adolescentes en nuestra muestra presentó TDM, tres cuartas partes presentó TOD y una tercera parte TDAH. La muestra americana muestra sesgos de selección diagnóstica al tener la evaluación en población comunitaria menor de edad con diagnósticos de TB, SMD, TOD y TDAH orientadas al proceso de validación de SMD (8), mientras que el grupo clínico del Reino Unido de población comunitaria (reclutada en escuelas) incluso mostró participantes con autismo. El grupo estudiado en China incluyó la población clínica ambulatoria con diagnósticos más variados, incluso el 72.2% de la población tenía diagnóstico de TDAH (12). El grupo estudiado en Italia estaba constituido de población comunitaria escolar y excluyó participantes con trastornos psiquiátricos y consumo de sustancias, y encontró que las mujeres reportaban mayor irritabilidad que hombres (13). La población española sobresale por estudiar longitudinalmente a la misma muestra de escolares a los 7 y 11 años y encontró reportes convergentes con conducta agresiva y rasgos de callo emocional (14). Considerando que tanto nuestros resultados como los reportados en todo el mundo identifican la IC asociada a todo tipo de psicopatología, podría considerarse como un especificador transdiagnóstico.

### Consistencia interna

La consistencia interna de la escala en su totalidad en nuestro estudio fue reportada con valores alfa de Cronbach de 0,91 y de 0,90, para el ARI-A y el ARI-P respectivamente. Estos valores son prácticamente iguales y están ligeramente por encima de los reportados en los estudios EUA, Reino Unido (8) y China (12). La versión italiana reporta un valor similar, aunque no especifica si es la versión autorreporte o versión padre/tutor (13). Esto implicaría que la consistencia interna del instrumento parece no verse afectada sea población clínica o comunitaria la que lo responde e independiente del idioma utilizado.

## Validez del constructo

A comparación de la mayoría de los estudios publicados, nosotros realizamos un análisis de componentes principales, el cual es un método distinto a un análisis factorial exploratorio ya que éste se enfoca en las relaciones entre variables (covarianza) y describe probables variables latentes. Por otro lado, estos análisis tienen ciertas similitudes (como uso de valores eigen) y se podría sugerir que para la obtención de componentes y factores pueden resultar análogos uno al otro (39). En nuestro estudio obtuvimos carga unifactorial en ambas versiones del ARI; lo cual corresponde a los hallazgos en los otros estudios publicados, como en la validación italiana (13) que se realizó mediante un análisis factorial exploratorio. En otros estudios que se realizó un análisis factorial confirmatorio también concluyeron en un solo factor en ambas versiones (como fue en EUA y en el Reino Unido (8) y Brasil (11)). Esto es interesante porque, aunque sean diferentes aproximaciones estadísticas se mantiene constante el hallazgo de que la varianza de la escala sea explicada principalmente por un solo factor. Cabe considerar que, por la muestra heterogénea diagnóstica en nuestro estudio, es llamativo concebir como la irritabilidad a pesar de estar constituida de diferentes maneras y en distintos diagnósticos, la variabilidad en esta escala se captura principalmente a través de una variable.

## Validez de criterio externo

En la última década se ha estudiado con más detalle la irritabilidad (4,40) y con base en recientes hallazgos el equipo de trabajo de la Clasificación Internacional de Enfermedades para su undécima edición (CIE-11) decidió agregar un especificador de irritabilidad crónica en el diagnóstico de TOD, en vez de usar una categoría similar al TDRE del DSM-5 (41). La correlación significativa en nuestro estudio con el TOD y no con el TDRE, apoya la idea de la IC como un especificador del TOD y no como un diagnóstico independiente.

## Validez divergente

En nuestro estudio no se encontró asociación estadísticamente significativa con puntajes en la escala SNAP-IV lo que sugiere que el constructo latente de irritabilidad obtenido por la escala ARI es un fenómeno diferenciable de los factores de inatención y de hiperactividad/impulsividad propios del TDAH. Sin embargo, se encontró una tendencia de correlación entre el ARI-P y el factor de hiperactividad/impulsividad, cuando el punto de corte del ARI-P fue el más alto. Este hallazgo coincide con análisis recientes que concluyen que los síntomas de TDAH (particularmente los de hiperactividad/impulsividad) predicen el grado y trayectoria de la irritabilidad a lo largo del desarrollo (42). Nuestros hallazgos se contrastan con los del estudio español donde se encontró que la irritabilidad reportada por los profesores tenía asociación con sintomatología asociada a TDAH y TOD, sin embargo, en esta investigación se tiene la limitante de no haber realizado una entrevista diagnóstica (14).

## Validez convergente

En esta investigación se encontró una adecuada validez convergente del ARI con las escalas de depresión y ansiedad, lo cual es congruente y es un hallazgo esperado con antecedentes comentados previamente (15). Este fenómeno es muy relevante y sigue estudiándose de varias maneras; por ejemplo, se ha indagado en una relación bidireccional entre estos dos componentes (43), y en otro instrumento para estudiar la irritabilidad encontraron una asociación fuerte de una subdimensión de reactividad emocional negativa internalizada con ansiedad y depresión (44). En otros estudios de validación de esta escala también investigaron asociaciones con otros tipos de psicopatología; en el caso de la versión italiana reportó una asociación significativa y positiva con todas las escalas del SDQ (un instrumento de tamizaje sobre varias áreas de dificultades, incluyendo problemas emocionales) excepto con el área prosocial que presentó asociación negativa (13). El estudio en China encontró correlación de comportamiento agresivo

y de irritabilidad en escala CBCL con puntajes en ARI-P y ARI-A ( $r=0,68$  y  $r=0,29$ , respectivamente), así como con puntajes de ansiedad/depresión ( $r=0,43$  y  $r=0,29$ , respectivamente) y problemas sociales ( $r=0,45$  y  $r=0,24$ , respectivamente) (12). En Australia se buscó la validez convergente contrastando el ARI contra la escala de ansiedad del RADS-2 y se encontró la asociación con ansiedad generalizada y la ansiedad social (10). Estos datos de los otros estudios del ARI demuestran que la IC puede asociarse tanto con trastornos internalizados como externalizados a través de diferentes instrumentos, dependiendo de la población estudiada, lo que pareciera implicar que converge con ambas dimensiones. Al igual que en el estudio original de la escala ARI encontró una asociación consistente entre IC y problemas emocionales (8) en nuestro estudio se encontró asociación con depresión y ansiedad. Esto podría implicar que el ARI podría ser un punto inicial para valorar síntomas depresivos y ansiosos, pero no se ha reportado en literatura disponible. De hecho, algunas investigaciones han mostrado que aunque el ánimo bajo o depresivo solo es el síntoma cardinal más común en jóvenes con depresión, la irritabilidad sola o aunada a dicho síntoma se presenta casi en el 42% de los jóvenes con depresión (45). Nuestra investigación añade a la literatura cómo se relaciona el ARI con escalas como DSRS y SCARED, contribuyendo a la relación entre la irritabilidad y síntomas internalizados. En futuros estudios se deberá detallar sobre las similitudes entre la irritabilidad y dichos síntomas, y las diferencias para así poder distinguir más claramente las características propias de la irritabilidad.

### Fiabilidad test-retest

El seguimiento de los participantes de este estudio fue alto, ya que más de 75% de la muestra inicial respondió por segunda ocasión la escala ARI, aunque la fiabilidad test-retest fue moderada para ARI-A y baja para ARI-P. Algo que podría explicar este fenómeno es que por la naturaleza de las entrevistas en un entorno clínico se puede alterar la percepción de algunos síntomas del propio adolescente o del tutor de un momento a otro, o bien que hubiera influencia por parte de intervenciones farmacológicas o

psicoterapéuticas entre las consultas. Nuestros valores obtenidos no se pueden comparar con el estudio original del instrumento ya que desafortunadamente en dicha publicación solo se aplicó por segunda ocasión a una fracción muy pequeña de la muestra y el seguimiento general fue un año después, lo que no permitió establecer fiabilidad test-retest (8). Nuestros resultados difieren del resto, ya que en el estudio de población italiana se repitió la prueba por internet 24 a 48hrs a 224 de 435 adolescentes, reportando una buena fiabilidad (13); y en el estudio en China reportó que ARI-P tenía mayor fiabilidad test-retest que ARI-A cuando repitieron la prueba 2 semanas después, resaltando que los sujetos con trastorno de espectro autista tenían una confiabilidad modesta (12). Es interesante que los dos ítems que presentaron las correlaciones más altas son los que parecen tener la capacidad de capturar los niveles más intensos de irritabilidad acorde al análisis por Lu y cols. (22); esto podría explicarse pensando que los casos más graves de irritabilidad resultarían constantes en cómo se perciben tanto por el sujeto y el tutor.

### Confiabilidad interinformante

La correlación observada entre ARI-P y ARI-A sugiere que existe cierto grado de acuerdo entre ambas versiones de la escala en la evaluación de la irritabilidad crónica que se presenta en el sujeto.

Aunque la mayoría de dichas correlaciones fueron significativas y solo en una de ellas (reactivo 3 “permanezco enojado mucho tiempo”) se observó una tendencia, dichas correlaciones presentaron coeficientes kappa de pobres a débiles, lo cual posiblemente esté relacionado al hecho de que la medición e identificación de la irritabilidad, es pobremente mejor identificado por observadores externos como lo son los padres. Aunque también podría ser explicado por diferencias culturales, ya que por ejemplo el estudio en población China encontró acuerdo moderado interinformante ( $r=0,44$ ) pero sólo encontró diferencia estadística en el primer ítem “los demás irritan fácilmente” (12). Llama la atención que el tercer reactivo que no alcanzó significancia estadística coincide con ser el reactivo que es menos discriminante y provee menor información acorde al análisis por Lu (22), y quizás sólo es relevante en

los casos de irritabilidad más grave. La discrepancia entre estos ítems se ha notado en varios estudios y algunos autores proponen que es consistente y refleja diferentes interpretaciones de los reactivos, por lo que no deben ser promediados, algunos autores han sugerido que la irritabilidad no es un constructo del todo unitario (46).

Recientemente se desarrolló la versión del clínico del ARI (CL-ARI, por sus siglas en inglés) la cual es una entrevista semiestructurada para que personal clínico califique irritabilidad en población pediátrica. En la validez convergente, las puntuaciones de CL-ARI fueron significativamente más altas para los jóvenes con TDRE que para cualquier otro grupo de diagnóstico, y mostraron asociaciones fuertes con otras medidas de irritabilidad informadas por médicos, padres y jóvenes. Esta nueva versión del ARI suma al clínico como observador y en su versión original (41) contrasta con las asociaciones que obtuvimos con el TDRE lo que sin duda genera retos en el establecimiento de este diagnóstico emergente. En futuros estudios sugerimos el uso de esta escala para complementar el uso de la escala ARI.

### Prueba diagnóstica

A excepción del punto de corte en 4 en el ARI-A se observó que la escala permite la identificación del constructo de IC cuando es visto a través del diagnóstico de TOD y no a través del TDRE. Comparando con el estudio de Mulraney (20) para un punto de corte, las diferencias en este puntaje podrían ser por las características de la muestra (en el estudio de Mulraney la población principalmente fue constituida por trastorno depresivo mayor y se enfocó en psicopatología en general) y que nuestro estudio se enfocó para diagnóstico de TOD. En el estudio de población China, aunque no reportan un punto de corte, sí encuentran que los participantes con TDRE tenían los puntajes más altos en escala de ARI-P y los participantes sanos con los puntajes más bajos, este hallazgo también se reportó con ARI-A pero sin significancia estadística (12). Estas asociaciones que contrastan con nuestros hallazgos sin duda representan las dificultades diagnósticas y de ubicación taxonómica de la irritabilidad dentro del DSM y la CIE (4).

### LIMITACIONES

La ausencia de información procedente de un grupo control que permitiera contrastar los datos obtenidos del grupo con psicopatología representa la mayor limitación en este estudio, ya que al no contar con dicha información no fue posible establecer las propiedades de la escala ARI como prueba de tamizaje. Otra limitación importante fue la correlación modesta que se obtuvo en la fiabilidad test-retest. Finalmente, la versión del ARI utilizada tiene la conceptualización sudamericana de su traducción en Argentina, lo cual no necesariamente implicaría los aspectos culturales de la población mexicana.

### CONCLUSIONES

Este trabajo reporta las propiedades psicométricas de las versiones del padre/tutor y de niños/adolescentes de la escala ARI las cuales mostraron validez y confiabilidad adecuadas para la población mexicana, similar a como se ha observado en otros países. Los resultados de este estudio muestran que la ARI cuenta con un gran potencial para ser utilizada en un contexto clínico y/o de investigación en la valoración de IC. La repercusión clínica de esta investigación se centra en la posibilidad de contar con una medida objetiva de autoreporte que facilite el ejercicio clínico y ayude a disminuir la brecha diagnóstica y de atención en esta población vulnerable.

### AGRADECIMIENTOS

Los autores agradecen a los participantes del estudio por aceptar ser parte de éste y a las médicas psiquiatras adscritas en la clínica de la adolescencia INPRFM por su trabajo y apoyo en la atención clínica (Dra. Liz Sosa y Dra. Miriam Feria). También agradecemos al Dr. Sebastián Totxo, Dr. Héctor Taboada, y Dr. José Carlos Medina por sus comentarios al manuscrito, y al Dr. Lino Palacios por sus contribuciones al proyecto inicial.

## FINANCIACIÓN

No requirió de financiamiento. Los resultados no se han presentado previamente en otros medios.

## CONFLICTOS DE INTERÉS DE LOS AUTORES

Los autores niegan conflicto de intereses que reportar.

## CONTRIBUCIONES

Francisco Rafael de la Peña Olvera y Andrés Rodríguez Delgado realizaron el concepto, diseño, análisis, interpretación de datos, escritura y corrección del artículo. Alejandro Irvin Soto Briseño y Assad Daniel Saad Manzanera participaron en la Interpretación de datos y la escritura; este último también participó en la corrección del artículo.

### Tabla de abreviaturas

|            |  |
|------------|--|
| ARI        | Índice de Reactividad Afectiva (Affective Reactivity Index en inglés)                          |
| DSM        | Manual Diagnóstico y Estadística (Diagnostics and Statistics Manual, por sus siglas en inglés) |
| IC         | Irritabilidad Crónica  |
| TB         | Trastorno bipolar  |
| K-SADS-PL  | Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children              |
| EUA        | Estados Unidos de América  |
| RU         | Reino Unido  |
| CFI        | Comparative Fit Index  |
| TLI        | Tucker-Lewis Index   |
| INPRFM     | Instituto Nacional de Psiquiatría "Ramón de la Fuente Muñiz"                                   |
| DSRS       | Depression Self Rating Scale   |
| SCARED     | Screen for Child Anxiety Related Emotional Disorders   |
| SNAP IV    | Swanson, Nolan & Pelham questionnaire  |
| TDRE       | Trastorno Disruptivo de la Regulación Emocional  |
| TOD        | Trastorno Oposicionista Desafiante   |
| BPRS-CA-29 | Brief Psychiatric Rating Scale for Children and Adolescents                                    |
| TDAH       | Trastorno por Déficit de Atención e Hiperactividad   |

|     |   |
|-----|---|
| ROC | Característica Operativa del Receptor (Receiver Operating Characteristic en inglés) |
| DE  | Desviación Estándar   |
| SMD | Disregulación Grave del Ánimo (Severe Mood Dysregulation en inglés)                 |
| CIE | Clasificación Internacional de Enfermedades   |

## REFERENCIAS

1. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, Va.: American Psychiatric Association; 2013.
2. Toohey MJ, DiGiuseppe R. Defining and measuring irritability: Construct clarification and differentiation. *Clinical Psychology Review*. 2017;53: 93-108.  
<https://doi.org/10.1016/j.cpr.2017.01.009>
3. Leibenluft E, Cohen P, Gorrindo T, Brook JS, Pine DS. Chronic Versus Episodic Irritability in Youth: A Community-Based, Longitudinal Study of Clinical and Diagnostic Associations. *Journal of Child and Adolescent Psychopharmacology*. 2006;16(4): 456-66.  
<https://doi.org/10.1089/cap.2006.16.456>
4. Evans SC, Burke JD, Roberts MC, Fite PJ, Lochman JE, autor, et al. Irritability in child and adolescent psychopathology: An integrative review for ICD-11. *Clinical Psychology Review*. 2017;53: 29-45.  
<https://doi.org/10.1016/j.cpr.2017.01.004>
5. Krieger FV, Leibenluft E, Stringaris A, Polanczyk GV. Irritability in children and adolescents: past concepts, current debates, and future opportunities. *Braz J Psychiatry*. 2013;35(s1): S32-S39.  
<https://doi.org/10.1590/1516-4446-2013-S107>
6. Stringaris A, Cohen P, Pine DS, Leibenluft E. Adult Outcomes of Youth Irritability: A 20-Year Prospective Community-Based Study. *AJP*. 2009;166(9): 1048-54.  
<https://doi.org/10.1176/appi.ajp.2009.08121849>
7. Vidal-Ribas P, Stringaris A. How and Why Are Irritability and Depression Linked? *Child and Adolescent Psychiatric Clinics of North America*. 2021;30(2): 401-14.  
<https://doi.org/10.1016/j.jchc.2020.10.009>

8. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, et al. The Affective Reactivity Index: a concise irritability scale for clinical and research settings. *Journal of Child Psychology and Psychiatry*. 2012;53(11): 1109-17. <https://doi.org/10.1111/j.1469-7610.2012.02561.x>
9. American Psychiatric Association. DSM-5-TR Online Assessment Measures. En: American Psychiatric Association [Internet]. Disponible en: <https://www.psychiatry.org:443/psychiatrists/practice/dsm/educational-resources/assessment-measures>
10. Mulraney MA, Melvin GA, Tonge BJ. Psychometric properties of the Affective Reactivity Index in Australian adults and adolescents. *Psychol Assess*. 2014;26(1): 148-55. <https://doi.org/10.1037/a0034891>
11. DeSousa DA, Stringaris A, Leibenluft E, Koller SH, Manfro GG, Salum GA. Cross-cultural adaptation and preliminary psychometric properties of the Affective Reactivity Index in Brazilian Youth: implications for DSM-5 measured irritability. *Trends Psychiatry Psychother*. 2013;35(3): 171-80. <https://doi.org/10.1590/S2237-60892013000300004>
12. Pan P-Y, Yeh C-B. Irritability and Maladaptation Among Children: The Utility of Chinese Versions of the Affective Reactivity Index and Aberrant Behavior Checklist-Irritability Subscale. *Journal of Child and Adolescent Psychopharmacology*. 2019;29(3): 213-9. <https://doi.org/10.1089/cap.2018.0070>
13. Grassi ML, Valente D, Berardi A, Tofani M, Galeoto G. Validation and Analysis of the Psychometric Properties of Two Irritability-Measuring Tools: The Affective Reactivity Index (ARI) and the Born-Steiner Irritability Scale (BSIS) in the Italian Adult and Adolescent Populations. *International Journal of Environmental Research and Public Health*. 2023;20(5): 4607. <https://doi.org/10.3390/ijerph20054607>
14. Ezpeleta L, Penelo E, De La Osa N, Navarro JB, Trepát E. How the Affective Reactivity Index (ARI) works for teachers as informants. *Journal of Affective Disorders*. 2020;261: 40-8. <https://doi.org/10.1016/j.jad.2019.09.080>
15. Evans SC, Abel MR, Doyle RL, Skov H, Harmon SL. Measurement and correlates of irritability in clinically referred youth: Further examination of the Affective Reactivity Index. *Journal of Affective Disorders*. 2021;283: 420-29. <https://doi.org/10.1016/j.jad.2020.11.002>
16. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*. 2013;310(20): 2191-4. <https://doi.org/10.1001/jama.2013.281053>
17. Ulloa RE, Ortiz S, Higuera F, Nogales I, Fresán A, Apiquian R, et al. Estudio de fiabilidad interevaluador de la versión en español de la entrevista Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL). *Actas Españolas de Psiquiatría*. 2006;34(1): 36-40.
18. Shafer AB. Factor structure of the Brief Psychiatric Rating Scale for Children (BPRS-C) among Hospital Patients and Community Clients. *Personality and Individual Differences*. 2013;55(1): 41-6. <https://doi.org/10.1016/j.paid.2013.01.024>
19. Transatlantic-comppsysh/Affective-Reactivity-Index. transatlantic-comppsysh; 2022. Available: <https://github.com/transatlantic-comppsysh/Affective-Reactivity-Index>
20. Mulraney M, Melvin G, Tonge B. Brief report: Can irritability act as a marker of psychopathology? *Journal of Adolescence*. 2014;37(4): 419-23. <https://doi.org/10.1016/j.adolescence.2014.03.005>
21. Dougherty LR, Galano MM, Chad-Friedman E, Olin TM, Bufferd SJ, Klein DN. Using Item Response Theory to Compare Irritability Measures in Early Adolescent and Childhood Samples. *Assessment*. 2021;28(3): 918-927. <https://doi.org/10.1177/1073191120936363>
22. Lu B, Fang Y, Cai J, Chen Z. Psychometric Evaluation of the Affective Reactivity Index Among Children and Adolescents in China: A Multi-Method Assessment Approach. *Assessment*. 2023;31(5): 10731911231199424. <https://doi.org/10.1177/10731911231199424>
23. Wilson MK, Cornacchio D, Brotman MA, Comer JS. Measuring Irritability in Early

- Childhood: A Psychometric Evaluation of the Affective Reactivity Index in a Clinical Sample of 3- to 8-Year-Old Children. *Assessment*. 2022;29(7): 1473–81. <https://doi.org/10.1177/10731911211020078>
24. Sugaya LS, Kircanski K, Stringaris A, Polanczyk GV, Leibenluft E. Validation of an irritability measure in preschoolers in school-based and clinical Brazilian samples. *Eur Child Adolesc Psychiatry*. 2022;31: 577-87. <https://doi.org/10.1007/s00787-020-01701-6>
  25. Birmaher B, Hudson I, Buchanan DG, Wolff S. Clinical Evaluation of a Self-Rating Scale for Depressive Disorder in Childhood (depression Self-Rating Scale). *J Child Psychol & Psychiat*. 1987;28(1): 43-60. <https://doi.org/10.1111/j.1469-7610.1987.tb00651.x>
  26. Lara Muñoz M, de la Peña Olvera F, Cortés J, Nicolini H, Páez F, Almeida L. Traducción al español y validez de la Escala de Birmaher (DSRS) para el Trastorno Depresivo Mayor en la Adolescencia. *Salud Mental*. 1996;19(extra 3): 17-23.
  27. Sánchez Quintana S, Ortiz León S, Autor, Higuera Romero FH. Validez y confiabilidad de la escala de ansiedad (SCARED) para niños y adolescentes en población abierta. *Especialidad en Medicina (Psiquiatría Infantil y de la Adolescencia)*, Universidad Nacional Autónoma de México. 2004. Available: <http://132.248.9.195/pdtestdf/0327421/Index.html>
  28. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale Construction and Psychometric Characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1997;36(4): 545-53. <https://doi.org/10.1097/00004583-199704000-00018>
  29. Grañana N, Richaudeau A, Gorriti CR, O'Flaherty M, Scotti ME, Sixto L, et al. Evaluación de déficit de atención con hiperactividad: la escala SNAP IV adaptada a la Argentina. *Rev Panam Salud Publica*. 2011;29(5): 344-9. <https://doi.org/10.1590/S1020-49892011000500007>
  30. Swanson JM, Kraemer HC, Hinshaw SP, Arnold LE, Conners CK, Abikoff HB, et al. Clinical Relevance of the Primary Findings of the MTA: Success Rates Based on Severity of ADHD and ODD Symptoms at the End of Treatment. *J Am Acad Child Adolesc Psychiatry*. 2001;40(2): 168-79. <https://doi.org/10.1097/00004583-200102000-00011>
  31. Bussing R, Fernandez M, Harwood M, Hou W, Garvan CW, Eyberg SM, et al. Parent and Teacher SNAP-IV Ratings of Attention Deficit Hyperactivity Disorder Symptoms: Psychometric Properties and Normative Ratings From a School District Sample. *Assessment*. 2008;15(3): 317-28. <https://doi.org/10.1177/1073191107313888>
  32. Hall CL, Guo B, Valentine AZ, Groom MJ, Daley D, Sayal K, et al. The Validity of the SNAP-IV in Children Displaying ADHD Symptoms. *Assessment*. 2020;27(6): 1258-71. <https://doi.org/10.1177/1073191119842255>
  33. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36(7): 980-8. <https://doi.org/10.1097/00004583-199707000-00021>
  34. De la Peña Olvera F, Villavicencio LR, Palacio JD, Félix FJ, Larraguibel M, Viola L, et al. Validity and reliability of the kiddie schedule for affective disorders and schizophrenia present and lifetime version DSM-5 (K-SADS-PL-5) Spanish version. *BMC Psychiatry*. 2018;18: 193. <https://doi.org/10.1186/s12888-018-1773-0>
  35. Lachar D, Randle SL, Harper RA, Scott-Gurnell KC, Lewis KR, Santos CW, et al. The Brief Psychiatric Rating Scale for Children (BPRS-C): Validity and Reliability of an Anchored Version. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001;40(3): 333-40. <https://doi.org/10.1097/00004583-200103000-00013>
  36. Cujíño Cano MF, Chaskel Heilbronner RE, Toledo Arenas JD. Validación de la versión en español de la Brief Psychiatric Rating Scale for Children para trastornos psiquiátricos en niños, en dos hospitales de Bogotá, Colombia. *Rev peru psiquiatr*. 2010;1(2): 55-62.

37. De la Peña Olvera F, Palacios L, Ulloa R. Validity of Mexican Modified Children's Brief Psychiatric Rating Scale (BPRS-C-25). 52 Annual Meeting of the American Academy of Child and Adolescent Psychiatry; 2005 Oct 18; Toronto, Canada.
38. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016;374(2065): 20150202.  
<https://doi.org/10.1098/rsta.2015.0202>
39. Gniazdowski Z. Principal Component Analysis versus Factor Analysis. *Zeszyty Naukowe WWSI*. 2021;24(15): 35-88.  
<https://doi.org/10.26348/znwwsi.24.35>
40. Lochman JE, Evans SC, Burke JD, Roberts MC, Fite PJ, Reed GM, et al. An empirically based alternative to DSM-5's disruptive mood dysregulation disorder for ICD-11. *World Psychiatry*. 2015;14(1): 30-3.  
<https://doi.org/10.1002/wps.20176>
41. Evans SC, Roberts MC, Keeley JW, Rebello TJ, Autor, Lochman JE, et al. Diagnostic classification of irritability and oppositionality in youth: a global field study comparing ICD-11 with ICD-10 and DSM-5. *J Child Psychol Psychiatr*. 2020;62(3): 13244.  
<https://doi.org/10.1111/jcpp.13244>
42. Johns-Mead R, Vijayakumar N, Mulraney M, Melvin G, Youssef G, Sciberras E, et al. Categorical and dimensional approaches to the developmental relationship between ADHD and irritability. *Journal of Child Psychology and Psychiatry*. 2023;64(10): 1422-31.  
<https://doi.org/10.1111/jcpp.13818>
43. Losiewicz OM, Metts AV, Zinbarg RE, Hammen C, Craske MG. Examining the indirect contributions of irritability and chronic interpersonal stress on symptoms of anxiety and depression in adolescents. *Journal of Affective Disorders*. 2023;329: 350-8.  
<https://doi.org/10.1016/j.jad.2023.02.125>
44. Dissanayake AS, Dupuis A, Arnold PD, Burton CL, Crosbie J, Schachar RJ, et al. Is irritability multidimensional: Psychometrics of The Irritability and Dysregulation of Emotion Scale (TIDES-13). *Eur Child Adolesc Psychiatry*. 2024;33: 2767-80.  
<https://doi.org/10.1007/s00787-023-02350-1>
45. Stringaris A, Maughan B, Copeland WS, Costello EJ, Angold A. Irritable Mood as a Symptom of Depression in Youth: Prevalence, Developmental, and Clinical Correlates in the Great Smoky Mountains Study. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2013;52(8): 831-40.  
<https://doi.org/10.1016/j.jaac.2013.05.017>
46. Mallidi A, Meza-Cervera T, Kircanski K, Stringaris A, Brotman MA, Pine DS, et al. Robust caregiver-youth discrepancies in irritability ratings on the affective reactivity index: An investigation of its origins. *J Affect Disord*. 2023;332: 185-93.  
<https://doi.org/10.1016/j.jad.2023.03.091>