

# *Psychometric properties of a Brazilian adaptation of the Teacher Observation of Classroom Adaptation-Revised Scale for Children's Behavior*

Daniela R. Schneider; Erikson Kaszubowski; Douglas Garcia;  
Luísa Scaff; Roberto Moraes-Cruz; Zila M. Sanchez

How to cite this article:

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M. (2020). Psychometric properties of a Brazilian adaptation of the Teacher Observation of Classroom Adaptation-Revised Scale for Children's Behavior. *Acta Colombiana de Psicología*, 23(1), 333-345. doi: <http://www.doi.org/10.14718/ACP.2020.23.1.15>

*Recibido, diciembre 3/2018; Concepto de evaluación, junio 15/2019; Aceptado, septiembre 4/2019*

## **Daniela R. Schneider\***

Universidade Federal de Santa Catarina, Santa Catarina, Brasil  
ORCID: <https://orcid.org/0000-0002-2936-650>

## **Erikson Kaszubowski**

Universidade Federal de Santa Catarina, Santa Catarina, Brasil  
ORCID: <https://orcid.org/0000-0001-5481-1755>

## **Douglas Garcia**

Universidade Federal de Santa Catarina, Santa Catarina, Brasil  
ORCID: <https://orcid.org/0000-0001-5245-6075>

## **Luísa Scaff**

Universidade Federal de Santa Catarina, Santa Catarina, Brasil  
ORCID: <https://orcid.org/0000-0003-0804-4457>

## **Roberto Moraes-Cruz**

Universidade Federal de Santa Catarina, Santa Catarina, Brasil  
ORCID: <https://orcid.org/0000-0003-4671-3498>

## **Zila M. Sanchez**

Universidade Federal de São Paulo, São Paulo, Brasil  
ORCID: <https://orcid.org/0000-0002-7427-7956>

## Abstract

The aim of this study was to describe the process of psychometric analysis of the *Teacher Observation of Classroom Adaptation- Revised Scale* (TOCA-R) for its use in Brazilian schools and to evaluate its validity and reliability. To evaluate the “Elos Program”, which is the Brazilian culturally adapted version of the North American Program “Good Behavior Game”, the TOCA-R was used. The researchers adapted the instrument in 2014, consisting of 33 items in a three-point ordinal response scale. A longitudinal quasi-experimental design with a single group was used. Participants were children aged 6 to 10 years evaluated by their teachers, before ( $n = 1448$ ) and after ( $n = 673$ ) the implementation of the Elos Program in 2014. The study involved initially four schools, 68 classes and their respective teachers. The analytical procedures were exploratory factorial analysis, confirmatory factorial analysis, longitudinal invariance analysis and reliability analysis by precision coefficients. The results of the exploratory factorial analysis showed an acceptable adjustment of five factors with 25 items, with a total explained variance of 60% and mean residual error of 0.02. The confirmatory factorial analysis expressed a satisfactory fit of the model ( $\chi^2 = 961$ ,  $df = 265$ , RMSEA = .078, 95% IC [.07, .08], and CFI = 0.9). A configurational, metric and scalar invariance of latent structure was identified, which, together with the amplitude of variation of the precision coefficients between the instrument dimensions ( $\alpha = .78, .92$ ;  $\omega = .76, .92$ ), demonstrate evidence of validity and reliability for using the TOCA-R in evaluating the *Elos Program* in Brazilian schools.

*Key words:* psychometric analysis, child mental health, preventive programs, Teacher Observation of Classroom Adaptation-Revised, cross-cultural validation.

\* Universidade Federal de Santa Catarina. Departamento de Psicologia, Campus David Ferreira Lima, Trindade, Florianópolis, Santa Catarina, Brasil. C. P.: 88040-900. Tel.: 554837218607 - 5548999717762. [daniela.schneider@ufsc.br](mailto:daniela.schneider@ufsc.br)  
Los autores agradecen al CNPq, al UNODC, al Ministerio de Salud y a los Departamentos Municipales y Estatales de Educación de Brasil que hicieron parte de este estudio, así como a los profesores y estudiantes que participaron en la intervención.

# ***Propiedades psicométricas de la adaptación de la escala *Teacher Observation of Classroom Adaptation-Revised* (TOCA-R) para evaluar el comportamiento infantil en Brasil***

## Resumen

El objetivo de este estudio fue realizar el análisis psicométrico de la escala *Teacher Observation of Classroom Adaptation-Revised* (TOCA-R) para su uso en escuelas brasileñas, y evaluar su validez y confiabilidad. Esta escala ha sido utilizada para evaluar el “Programa Elos”, que es la versión brasileña culturalmente adaptada del programa norteamericano “Good Behavior Game”. El instrumento fue adaptado por los investigadores en 2014 y está compuesto por 33 ítems con una escala de respuesta ordinal de tres puntos. En este estudio, se utilizó un diseño longitudinal cuasiexperimental de grupo único con los niños de 6 a 10 años, evaluados por sus maestros antes ( $n = 1448$ ) y después ( $n = 673$ ) de la implementación del Programa Elos en 2014. Se incluyeron cuatro ciudades brasileñas, 19 escuelas, 68 cursos y sus respectivos maestros, y los procedimientos analíticos fueron el análisis factorial exploratorio, el análisis factorial confirmatorio, el análisis de invariancia longitudinal y el análisis de confiabilidad por coeficientes de precisión. Los resultados del análisis factorial exploratorio mostraron un ajuste aceptable de cinco factores con 25 ítems, con una varianza total explicada del 60 % y un error residual medio de 0.02; el análisis factorial confirmatorio mostró un ajuste satisfactorio del modelo ( $\chi^2 = 961$ ,  $df = 265$ ,  $RMSEA = .078$ ,  $IC\ 95\ \% = .07-.08$ , y  $CFI = 0.9$ ); y se identificó la invariancia configuracional, métrica y escalar de la estructura latente, la cual, junto con la amplitud de variación de los coeficientes de precisión entre las dimensiones del instrumento ( $\alpha = .78, .92$ ;  $\omega = .76, .92$ ), da cuenta de evidencias de validez y confiabilidad que permiten utilizar la escala en la evaluación del Programa Elos en las escuelas brasileñas.

*Palabras clave:* análisis psicométrico, salud mental infantil, programas de prevención, teacher observation of classroom adaptation-revised, adaptación escolar, validación intercultural.

## Introducción

Para garantizar la validez de los datos recolectados en estudios de intervención es necesario que los instrumentos utilizados sean evaluados por su capacidad de identificar los comportamientos estudiados, pues, sin mediciones adecuadas, incluso un ensayo clínico bien diseñado no podría proporcionar información válida sobre la efectividad de un programa ni sobre los resultados obtenidos a través de su implementación. Por consiguiente, los investigadores y profesionales que quieran utilizar instrumentos que ya han sido validados en otros países deben asumir la labor inicial de traducirlos a su propio idioma y validarlos con el fin de usarlos en contextos culturalmente diferentes (Coster & Mancini, 2015). Asimismo, para implementar programas de salud que han demostrado su efectividad en otros países, es necesario no solo adaptar el programa mismo, sino también los instrumentos que miden su eficacia y su efectividad. De este modo, resulta de gran importancia considerar la necesidad de realizar adaptaciones culturales tanto de los programas escolares en salud mental como de sus sistemas de evaluación para países con distintos sistemas socioeducativos, que a su vez se adecuen a la intervención en problemas de comportamiento desde una dimensión cultural y psicosocial (Murphy, Abel, Hoover, Jellinek & Fazel, 2017).

Entre los años 2013 y 2016, la Coordinación General de Salud Mental, Alcohol y Otras Drogas del Ministerio

de Salud de Brasil llevó a cabo un proyecto de implementación de programas preventivos internacionales basados en la evidencia con el propósito de adaptarlos a la realidad brasileña y formular políticas públicas en el campo de la prevención del consumo de drogas (Pedroso, Abreu & Kinoshita, 2015; Schneider et al., 2016). La colaboración entre la Universidad Federal de Sao Paulo (UNIFESP) y la Universidad Federal de Santa Catarina (UFSC) desarrolló la evaluación del proceso y de la eficacia de estos programas.

Específicamente, el “Good Behavior Game” (GBG)—un programa preventivo norteamericano sobre salud mental infantil— fue uno de los programas evaluados y adaptados a la realidad de escuelas brasileñas. La Oficina de Naciones Unidas contra las Drogas y el Delito (UNODC) recomendó el programa por la base en la evidencia que obtuvo en varios países (Bayer et al., 2009; Kellam et al., 2011; Tingstrom, Sterling-Turner & Wilczynski, 2006), y el Instituto Americano para la Investigación (AIR) asesoró la experiencia brasileña, capacitó a su primer equipo de mentores del programa, y sugirió las herramientas de seguimiento y evaluación del GBG para esta población (Schneider et al., 2016).

Ahora bien, ni el GBG ni el Programa Elos presentan contenido curricular específico, sino que se consideran un método pedagógico de manejo del salón de clases. Aquí, los niños son designados para trabajar en equipo y crear un ambiente de aprendizaje positivo a través de la observación de su propia conducta y de sus interacciones sociales. Por

consiguiente, la mediación de grupos de pares —basada en actividades académicas cotidianas— hace énfasis en actitudes como el autocontrol, el compromiso con la tarea, la regulación emocional y el cambio conductual (Ford, Keegan, Poduska, Kellam & Littman, 2013), todo con el objetivo principal de reducir los comportamientos agresivos, disruptivos y de aislamiento social en los niños (Ford et al., 2013; Kellam et al., 2011).

Teniendo en cuenta las vulnerabilidades personales y sociales del niño, el programa aborda los factores de riesgo para futuros comportamientos antisociales, como los problemas relacionados con el consumo de drogas y la participación en situaciones violentas (Kellam et al., 2011; Kellam et al., 2014), actuando como una clase de “vacuna conductual” (Embry, 2002). De este modo, al facilitar la construcción de relaciones sanas e inclusivas, el programa mejora las habilidades sociales y académicas de los niños y es así como logra situarse como un medio que contribuye a la promoción de la salud mental de los niños (Schneider et al., 2016). Como se ha encontrado en la literatura, este programa ha tenido varias aplicaciones y adaptaciones culturales en diferentes países del mundo (Nolan, Houlihan, Wanzek & Jenson, 2014), incluyendo experiencias en países de América Latina como Chile (Pérez, Fernández, Rodríguez & de la Barra, 2004; Pérez, Rodríguez, de la Barra & Fernández, 2005).

Es importante mencionar que el GBG tiene como guía teórica el concepto de “curso de vida/campo social”, con el cual se plantea que los factores de riesgo del desarrollo temprano están asociados a las consecuencias de los problemas de los adultos y, por lo tanto, las estrategias de prevención deben establecerse lo antes posible. Esta teoría proporciona una visión dual de la salud mental en la que, por un lado, incluye la adaptación como una dimensión social que implica la forma en que la sociedad percibe a una persona en cada etapa de su vida, además de cómo esta es requerida en labores sociales; y, por otro, es percibida como una dimensión individual direccionada hacia el bienestar psicológico (Kellam et al., 2011). Incluso, como mencionan Kellam et al. (2011):

De acuerdo con la teoría del curso de vida/campo social, al mejorar la forma en que los profesores se relacionan con los niños en el aula de clase se conseguirá una mejor adaptación social de los niños en el ámbito social del salón de clases. La teoría también predice que esta mejora temprana en la adaptación social conducirá a una mejor adaptación en otros campos sociales a lo largo de la vida (p. 76; traducción propia).

Para asegurar la calidad de la evaluación de la efectividad de un programa como el GBG, o como el Programa Elos, es esencial certificar las propiedades psicométricas de los

instrumentos utilizados en la evaluación. Por tanto, para garantizar una evidencia válida y confiable es fundamental no solo determinar la consistencia interna y externa de la prueba (Pasquali, 2010), sino también evaluar la validez interna y externa del programa, así como la consistencia metodológica de los resultados en relación con la implementación y la extrapolación de los resultados más allá de las contingencias específicas del estudio (Durlak & DuPre, 2008; Flay et al., 2005).

Ahora bien, la herramienta principal propuesta para evaluar la efectividad del GBG es la escala *Teacher Observation of Classroom Adaptation* (TOCA, por sus siglas en inglés), en la cual el profesor responde preguntas sobre su percepción con respecto a los comportamientos de cada uno de sus estudiantes (Koth, Bradshaw & Leaf, 2009; Werthamer-Larsson, Kellam & Wheeler, 1991). El Centro de Investigación de Woodlawn (Chicago, EE. UU.) desarrolló este instrumento de evaluación hace más de 40 años, en 1975 aproximadamente, con el fin de describir los comportamientos de cada niño en el salón de clase y las tareas exigidas en ese espacio. En este, las categorías creadas no se hicieron para ser clasificaciones de comportamientos sintomáticos clínicos, sino, más bien, evaluaciones de adaptación social de las conductas en clase de estudiantes de primaria, observadas y definidas por los profesores (Koth et al., 2009).

Después de años de investigación, la escala original de 110 ítems ha tenido varias adaptaciones (Guzmán et al., 2015; Koth et al., 2009; Kourkounasiou & Skordilis, 2014; Murphy et al., 2014; Wang et al., 2014; Werthamer-Larsson et al., 1991), y TOCA-R fue la primera versión revisada, realizada en 1991, en la que, tal como mencionan Werthamer-Larsson et al. (1991), “la sección inicial fue revisada y reducida a 58 ítems, los cuales fueron luego refinados para ajustarlos con los aspectos comportamentales según los criterios del trastorno infantil del DSM III” (p. 590, traducción propia). En la versión final, el TOCA-R contenía 31 ítems, y en el análisis factorial exploratorio y confirmatorio se definieron tres factores de tipo conductual: (a) contacto social frente a timidez ( $\alpha = .85$ ); (b) aceptación de autoridad frente a agresividad ( $\alpha = .92$ ); y (c) problemas de concentración ( $\alpha = .96$ ); con una correlación test-retest de .75 o más alta en cada subescala. La escala original es aplicada en forma de entrevista con profesores por parte de técnicos entrenados para realizar la tarea (Koth et al., 2009; Werthamer-Larsson et al., 1991).

En el presente artículo, la versión de 2010 del TOCA-R, propuesta por el Instituto Americano de Investigación (AIR, 2010), será objeto de análisis, pues, como se especificará más adelante, fue la base para la adaptación brasileña que aquí se discute. De hecho, existen otras adaptaciones para países latinoamericanos, como la versión *Teacher Observation*

of *Classroom Adaptation-Re-Revised* (TOCA-RR), desarrollada en Chile en la década de 1990, para la evaluación de programas de salud mental para niños. Estas versiones presentan buenos indicadores psicométricos, con un rango de valores alfa de Cronbach que van de .92 a .96 en las subescalas del TOCA-R, y de .74 a .95 en las subescalas de TOCA-RR (George et al., 2004; Guzmán et al., 2015; Murphy et al., 2014).

Sin embargo, aunque hay algunas indicaciones de alta confiabilidad en las evaluaciones del test-retest del TOCA, además de validez convergente con otras medidas (Werthamer-Larsson et al., 1991), en algunos estudios se ha encontrado que su aplicación es muy costosa y que consume mucho tiempo aplicarla en estudios a gran escala (Koth et al., 2009, Kourkounasiou & Skordilis, 2014). Como consecuencia, se han explorado alternativas que redujeran los costos y el tiempo de aplicación. Un ejemplo fue la creación de una versión reducida de 24 ítems, una lista de chequeo titulada TOCA-C, que incluía la posibilidad de una aplicación autoadministrada por el profesor. Algunos estudios han mostrado que el cuestionario autoadministrado es una buena alternativa para el formato original de entrevista estructurada porque requiere menos tiempo invertido por parte de profesores e investigadores, y porque hace la investigación más rápida y más viable financieramente (Koth et al., 2009; Kourkounasiou & Skordilis, 2014).

Teniendo en cuenta estas adaptaciones, los investigadores ajustaron y adaptaron el TOCA para utilizarlo en la evaluación del Programa Elos. Allí, las adaptaciones fueron justificadas principalmente por la aplicabilidad del instrumento a la realidad de cada contexto educativo, ya que mide mejor los resultados de los objetivos y facilita su aplicación. Así, el objetivo del presente artículo fue describir el análisis psicométrico y algunos aspectos de la adaptación cultural de la escala *Teacher Observation of Classroom Adaptation-Revised* (TOCA-R), para ser utilizada en la futura evaluación de la eficacia del Programa Elos en escuelas brasileñas, y para otras evaluaciones de salud mental infantil.

## Método

### *Tipo de estudio*

Este estudio psicométrico de la adaptación brasileña de la escala TOCA-R está basada en datos obtenidos en el estudio piloto del Programa Elos de 2014, cuyo propósito consiste en adaptar y validar los instrumentos de medición, el contenido y el método preventivo del programa para una

futura evaluación de su eficacia. El diseño del estudio fue de tipo longitudinal y tuvo características cuasiexperimentales de un solo grupo, con evaluaciones pre y postest, y con una selección no aleatorizada de los participantes, en la que no se incluyó un grupo control. Para dicho trabajo, los profesores respondieron la escala antes de empezar el programa, en agosto del 2014, y, nuevamente, cuatro meses después, al final del año escolar.

La adaptación transcultural de la escala TOCA-R estuvo basada en los pasos propuestos por Borsa, Damásio y Bandeira (2012): (a) traducciones directas del idioma original al idioma blanco; (b) síntesis de diferentes traducciones; (c) revisión por el comité de expertos; (d) evaluación por la población objetivo; (e) traducción inversa; y (f) estudio piloto.

### *Participantes*

El Ministerio de Salud fue el responsable de la implementación del Programa Elos en escuelas públicas de Brasil en el año 2014, mientras que las universidades asociadas fueron responsables de su evaluación. La muestra incluyó todos los estudiantes que participaron en la implementación, incluyendo todas las escuelas y profesores seleccionados por el Departamento Municipal de Educación, en las ciudades elegidas para el estudio piloto.

La muestra de estudiantes estuvo compuesta por 68 salones de clase de escuela primaria (de primero a quinto grado) de 19 escuelas en cuatro ciudades de Brasil (Curitiba, São Bernardo do Campo, Florianópolis y Tubarão). Asimismo, 68 profesores respondieron la versión adaptada del TOCA en la evaluación conductual de 1448 estudiantes en el pretest y de 673 en el postest. En total, 624 estudiantes fueron emparejados entre el pre y el postest; 775 estudiantes se perdieron en el seguimiento. Estas pérdidas ocurrieron debido a los cambios en la implementación del programa por parte del Ministerio de Salud, y algunas escuelas abandonaron por la falta de disponibilidad de personal para llevar a cabo la supervisión.

En la muestra final (n = 624), 51 % eran hombres; 7.6 % era estudiante de primer grado, 20.5 % de segundo grado, 35.2 % de tercer grado, 16.1 % de cuarto grado y 20.6 % de quinto grado; y el promedio de edad fue de 8.21 años.

### *Instrumento*

La adaptación transcultural brasileña estuvo basada en la versión del TOCA-R revisada por el Instituto Americano para la Investigación en 2010. Específicamente, esta escala está compuesta por 55 ítems con una escala de respuesta ordinal de seis puntos (de “casi nunca” a “casi siempre”), y deriva de tres escalas: (a) el TOCA-R, de Werthamer-Larsson

et al. (1991), que evalúa la timidez, el comportamiento prosocial, la aceptación de autoridad o el comportamiento disruptivo, la preparación académica y los problemas de concentración; (b) la subescala de competencia social (Conduct Problems Prevention Research Group, 1992), que mide la regulación emocional y la competencia social; y (c) la escala de relación entre alumno y maestro (Pianta & Steinberg, 1992), que evalúa la proximidad alumno-maestro, el conflicto y la dependencia.

### *Procedimiento*

La adaptación brasileña definitiva del TOCA-R está compuesta por 33 ítems que se responden en una escala de frecuencia ordinal de tres puntos, cuyo proceso de adaptación será descrito a continuación. Para iniciar, cada profesor evaluó a todos sus estudiantes de cada clase en dos momentos: uno en agosto de 2014, antes de que el programa comenzara, y otro en diciembre del mismo año, cuatro meses después de ser implementado el programa. Los profesores recibieron instrucciones detalladas en el programa de entrenamiento sobre cómo completar el instrumento, y la aplicación fue considerada en el proceso de implementación como parte de la preparación del maestro para hacer que el programa lograra que tuviera una mirada más detallada del comportamiento de los estudiantes, lo que resultó ser un aspecto importante para la organización de los equipos del Programa Elos y para el seguimiento del progreso de los estudiantes.

La escala fue autoadministrada por los profesores según modelo de aplicación utilizado en el TOCA-C (Koth et al., 2009), y los maestros tuvieron una semana para completar el formato escrito TOCA para cada uno de sus estudiantes. Como se acordó con el personal pedagógico, ellos podían diligenciarlo en el tiempo dedicado a la preparación de la clase o a la capacitación, o en la casa si no pueden realizarlo en sus horas en la escuela. Al final del este periodo, el equipo técnico recolectó los instrumentos diligenciados de las escuelas participantes y se los entregó al equipo de investigación.

### *Análisis estadísticos*

Los análisis estadísticos utilizados para examinar las propiedades psicométricas de la adaptación de la escala TOCA-R fueron: el análisis factorial exploratorio, desarrollado con los participantes del pretest; el análisis factorial confirmatorio, con una muestra simple; la medición de la invariancia longitudinal; y los análisis de confiabilidad. Todos los análisis fueron ejecutados en el programa R (versión 3.3.2), con la ayuda de los paquetes Psych (versión 1.7.3, Revelle, 2018) y Lavaan (versión 0.5, Rosseel, 2012).

*Análisis factorial exploratorio.* Como consecuencia de las modificaciones hechas en la escala TOCA-R, el análisis factorial exploratorio (AFE) se ejecutó para identificar la mejor solución factorial, análisis que se realizó con una validación cruzada. Primero, se seleccionó de manera aleatoria el 70 % ( $n = 1014$ ) de la muestra del pretest para constituir los datos de entrenamiento analizados por el AFE; y, segundo, se seleccionó el número de factores según ocho criterios: (a) el criterio de Kaiser —un número de valores propios (*eigenvalues*) mayor que uno—; (b) el gráfico de sedimentación de Cattell; (c) el análisis paralelo de Horn; (d) el análisis de estructura simple de Revelle (VSS); (e) el promedio medio parcial de Velicer (MAP); (f) el criterio de información bayesiano (BIC); (g) el error cuadrático medio de aproximación (RMSEA); y (h) el estadístico de bondad de ajuste  $\chi^2$ . Posteriormente, los ítems de la escala se evaluaron según tres criterios: la comunalidad del ítem, que debía ser mayor que .5; la complejidad del ítem, que debía ser menor que 1.5; y que cada factor debía permanecer con al menos con cinco ítems.

Es importante mencionar que los modelos del análisis de factores fueron ajustados a la matriz de correlación de Pearson usando el máximo estimador de probabilidad; que aun cuando algunas transgresiones del supuesto de normalidad multivariante no afectan las estimaciones puntuales, estas podrían provocar subestimación de errores estándar y de los estadísticos  $\chi^2$  (Kaplan, 2000); que se corrigió el estadístico  $\chi^2$  con el Satorra-Bentler para solucionar las posibles violaciones de normalidad multivariada —el factor de escala basado en la media y los errores estándar estuvieron basados en la estimación robusta de Satorra-Bentler—; y que la matriz de carga factorial se rotó usando el criterio de Oblimin para obtener una solución dispersa con factores correlacionados.

*Análisis factorial confirmatorio.* Al modelo final obtenido por el AFE se le realizó una validación cruzada con una muestra de validación del pretest ( $n = 434$ ) y un procedimiento de análisis factorial confirmatorio (AFC). El modelo del CFA se ajustó a la matriz de correlación de Pearson utilizando el máximo estimador de probabilidad. Para corregir posibles violaciones al supuesto de normalidad multivariante, como se explicó en la sección del AFE más arriba, se utilizó el estadístico de bondad de ajuste para la media ajustada  $\chi^2$  (Satorra-Bentler) y los errores robustos estándar. Finalmente, el índice de ajuste comparativo (CFI) evaluó el ajuste del modelo, el error cuadrático medio de aproximación (RMSEA) y la raíz media del cuadrado residual estandarizado (SRMR).

*Medida de la invariancia de longitudinal.* Las comparaciones pre-post requieren que los resultados del instrumento se conserven estables entre las aplicaciones y que reflejen solo los cambios causados por la intervención, por tanto, se ajustó el modelo factorial final a la muestra pre-post emparejada ( $n = 625$ ), realizando una secuencia de análisis factorial longitudinal confirmatorio cada vez más restrictivo: invariancia ajustada —la misma estructura factorial en ambos momentos—; invariancia métrica —la misma carga factorial en ambos momentos—; e invariancia escalar —la misma intercepción en ambos momentos—. La aceptabilidad restringida del modelo y, por consiguiente, el mantenimiento de la hipótesis de la medida de invariancia longitudinal, fue diferente principalmente en el CFI y el RMSEA (Cheung & Rensvold, 2002), los cuales fueron computados a partir del estadístico de bondad de ajuste para la media ajustada  $\chi^2$  (Satorra & Bentler, 2001).

*Análisis de confiabilidad.* Para evaluar la confiabilidad de cada subescala se utilizó el alfa de Cronbach y el omega de McDonald, de manera separada para los momentos pre y postest. Estas dos medidas de confiabilidad se reportan debido a que: (a) el alfa de Cronbach es ampliamente utilizado y de fácil comprensión en el desarrollo de escalas psicológicas, incluso cuando las asunciones no son teóricamente viables; y (b) el omega de McDonald ofrece una mejor estimación de la confiabilidad de la escala cuando la escala es equivalente, que es el caso de la mayoría de las escalas psicológicas.

#### *Aspectos éticos*

La comunidad de ética de la Universidad de São Paulo (n. °473.498) y la Universidad Federal de Santa Catarina (n. ° 711.377), en Brasil, aprobaron este estudio en todas sus fases metodológicas, las cuales se encuentran en concordancia con las exigencias de la Declaración de Helsinki. Todos los profesores firmaron el consentimiento informado del estudio; y, para los estudiantes, debido a su edad, se adoptó el formato de consentimiento pasivo, en el que los padres eran informados de la investigación por parte de la dirección de la escuela y debían expresar si no querían que sus hijos participaran en él.

## **Resultados**

#### *Adaptación brasileña de la escala TOCA*

Teniendo en cuenta el método propuesto por Borsa et al. (2012) para la adaptación transcultural del instrumento, los tres primeros pasos —primero, la traducción directa del

idioma original al idioma blanco; segundo, la síntesis de diferentes traducciones; y, tercero, la revisión del comité de expertos— se llevaron a cabo como sigue a continuación.

En el 2013, en el estudio prepiloto del Programa Elos se aplicó la traducción original del TOCA-R, elaborada por un especialista en el campo, con sus 55 ítems correspondientes, en un grupo de 27 profesores de seis escuelas públicas de São Paulo, Florianópolis y Tubarão, quienes evaluaron a 603 estudiantes. Después de la implementación del programa, estos mismos aplicaron la escala nuevamente y sus resultados se utilizaron para evaluar si la escala traducida podía utilizarse para evaluar el programa en un estudio futuro (Schneider et al., 2016).

En 2014, otro especialista con maestría en traducción elaboró una nueva versión traducida del inglés al portugués de Brasil. Luego, un experto del Ministerio de Salud —que coordinó la implementación del Programa Elos y los cuatro programas guía— comparó la traducción del 2013 con la nueva traducción del 2014. Cada persona involucrada realizó su propio análisis, tomó notas ítem por ítem con respecto a sus observaciones, y luego, en reuniones grupales, discutieron las evaluaciones individuales y definieron la mejor traducción, llegando a una conclusión en conjunto. Finalmente, los dos responsables de la evaluación del programa y un analista de comportamiento experimental —ya que este enfoque es uno de los fundamentos principales del GBG— revisaron los 55 ítems originales en la nueva versión brasileña.

Los maestros entrevistados luego de aplicar la escala en 2013 expresaron su inconformidad acerca del número de ítems del TOCA-R, al afirmar que fueron demasiados ítems con significado similar y un número excesivo de niveles en la escala de frecuencia, lo cual dificultó su respectiva evaluación. La actividad de completar la escala para cada estudiante del salón de clase fue considerada extenuante para la mayoría de los profesores, y en general expresaron no querer participar el año siguiente si tenían que completar la misma escala tan extensa.

Teniendo en consideración la retroalimentación de los maestros, los dos investigadores responsables del programa modificaron la adaptación de la escala reduciendo su tamaño y excluyendo ítems repetitivos y de contenido similar. Los investigadores se encargaron de balancear los ítems en cada dimensión original: conducta prosocial y timidez, aceptación de la autoridad o comportamiento disruptivo, preparación académica, y problemas de concentración. Con esto, finalmente, se obtuvo un conjunto de 33 ítems. De igual forma, la frecuencia ordinal de la escala también fue reducida de seis a tres niveles (1 = “rara vez”, 2 = “algunas veces” y 3 = “frecuentemente”). Los guías del programa y su coordinador nacional revisaron las propuestas, y el especialista en análisis experimental realizó la versión final,

así como la evaluación de su validez de constructo tras la adaptación de la escala.

Los otros dos pasos de Borsa et al. (2012)—el cuarto, la evaluación de la población objetivo; y el sexto, el estudio piloto— se describen en detalle en la continuación de este artículo. La traducción al idioma original fue el único paso que no se realizó.

#### *Análisis factorial exploratorio*

No se tuvo un acuerdo claro sobre el criterio para el número adecuado de factores latentes, pues el criterio de Kaiser sugirió tres factores; el gráfico de sedimentación sugirió entre tres y cinco; el análisis paralelo sugirió cinco; el análisis VSS sugirió entre uno y dos —según la complejidad máxima aceptable—; el MAP sugirió cuatro

factores; la reducción del BIC y del RMSEA sugirieron cinco factores; y el estadístico de bondad de ajuste  $\chi^2$  alcanzó una no significancia ( $\alpha = .05$ ) con más de 16 ítems. También, el ajuste de los ítems basado en la comunalidad, la complejidad y un número similar de ítems por factor dio como resultado una mejor solución de cinco factores. Finalmente, debido a que la solución de cinco factores fue sugerida por tres de seis criterios y tiene el mejor ajuste de ítems, esta solución fue la que se escogió como una solución final y es la que se presenta a continuación.

La comparación entre el patrón y la estructura de las matrices indicó que las correlaciones ítem-factor son similares a las cargas del ítem. De este modo, solo el patrón de la matriz de la carga factorial para la solución de cinco factores se presenta en la Tabla 1.

Tabla 1.

*Datos del pretest con una solución de cinco factores con matriz de carga factorial (matriz de patrones) del TOCA-R adaptado*

Ítem	F3	F1	F2	F4	F5	h <sup>2</sup>	U <sup>2</sup>	Comp.
18	0.97					0.89	0.11	1.01
19	0.94					0.83	0.17	1.02
16	0.80					0.73	0.27	1.07
01	0.70					0.57	0.43	1.04
33	-0.68					0.57	0.43	1.05
05	-0.57					0.55	0.45	1.31
27	0.42					0.57	0.43	2.16
09		0.88				0.79	0.21	1.02
22		0.82				0.68	0.32	1.03
08		0.72				0.57	0.43	1.06
15		0.69				0.79	0.21	1.21
21		0.62				0.64	0.36	1.29
20		0.40			-0.33	0.55	0.45	2.60
12			0.79			0.66	0.34	1.02
24			0.67			0.59	0.41	1.08
02			0.65			0.71	0.29	1.23
26			-0.65			0.60	0.40	1.29
23			0.56			0.73	0.27	1.50
14		0.38	0.55			0.70	0.30	1.96
28			0.38			0.51	0.49	2.39
17			-0.32			0.54	0.46	3.03
07				-0.72		0.52	0.48	1.01
32				-0.67		0.44	0.56	1.04
10				-0.63		0.50	0.50	1.39
03				0.63		0.56	0.44	1.26
13				0.51		0.62	0.38	1.72
06	0.30			0.42		0.39	0.61	2.08
11				0.40		0.45	0.55	2.09
04				0.30	0.17	0.40	0.60	3.30
29					0.76	0.66	0.34	1.05
25					0.70	0.50	0.50	1.02
30					0.63	0.60	0.40	1.18
31					0.47	0.45	0.55	1.86

Nota. Cargas de  $< .3$ ; h<sup>2</sup> = comunalidad; U<sup>2</sup> = singularidad; Comp. = complejidad media del ítem.

En definitiva, estos cinco factores explicaron el 60 % del total de la varianza de los ítems —16 %, 13 %, 13 %, 9 % y 9 %, respectivamente para cada factor—. El test de bondad de ajuste  $\chi^2$  fue estadísticamente significativo ( $\chi^2(373) = 1323, p < .001$ ), lo cual era esperado, dado el tamaño de la muestra, pero los índices de ajuste global sugirieron un buen ajuste (RMSEA = .05; TLI = .94, SRMR = .02). La solución de cinco factores también tuvo la ventaja de discriminar entre los ítems de comportamiento agresivo y de comportamiento disruptivo, como se muestra más adelante.

Por otra parte, la comunalidad de los ítems fue relativamente alta, pues el 79 % de los ítems tuvo más del 50 % de sus varianzas explicadas por los cinco factores latentes. También, el 66 % de los ítems tuvo una complejidad inferior a 1.5, lo que significa que para la mayoría de los ítems un

solo factor era el responsable para más de la mitad de su varianza.

Es importante mencionar que los ítems 6, 11, 14, 17, 20, 27, 28 (21 % del total de ítems) fueron eliminados porque no se ajustaban a los criterios descritos en el método. En la Tabla 2 se muestran los ítems restantes y sus respectivos factores. Específicamente, el primer y segundo factor describen el comportamiento agresivo y el comportamiento disruptivo, respectivamente; el tercer factor sugiere aceptación de instrucciones y compromiso con la tarea; el cuarto factor sugiere habilidades sociales y comportamiento prosocial; y el quinto factor describe autocontrol y autonomía.

La matriz de correlación de factores sugiere factores de orden superior, pues el comportamiento agresivo y disruptivo tienen una alta correlación positiva ( $r = .77$ ) y correlaciones negativas con otros factores; mientras que el compromiso

Tabla 2.

*Lista final de ítems agrupados según los factores latentes obtenidos en el análisis factorial exploratorio del TOCA-R adaptado*

Factor 1: Comportamiento agresivo
8. <i>Produz danos ou destrói a propriedade de propósito.</i>
9. <i>Inicia brigas.</i>
15. <i>Provoca os colegas.</i>
21. <i>Grita com os outros.</i>
22. <i>Machuca os outros fisicamente.</i>
Factor 2: Comportamiento disruptivo
2. <i>Quebra regras.</i>
12. <i>Interrompe ou intromete-se na fala dos outros.</i>
23. <i>É teimoso.</i>
24. <i>Está fora do lugar com frequência.</i>
26. <i>Espera a sua vez.*</i>
Factor 3: Compromiso con la tarea
1. <i>Trabalha bem sozinho(a).</i>
16. <i>Completa as tarefas.</i>
18. <i>Concentra-se nas tarefas.</i>
19. <i>Presta atenção.</i>
33. <i>Esforça-se pouco nas atividades.*</i>
Factor 4: Socialización
3. <i>Outras crianças o/a procuram para brincar e realizar atividades.</i>
7. <i>É socialmente isolado(a).*</i>
10. <i>Comporta-se timidamente.*</i>
13. <i>Ajuda aos outros, coopera com pares, compartilha com outros.</i>
32. <i>Passa bastante tempo sozinho.*</i>
Factor 5: autocontrol
4. <i>Mostra empatia e compaixão para sentimentos dos outros.</i>
25. <i>Lida bem com desapontamento ou frustração.</i>
29. <i>Consegue controlar as emoções.</i>
30. <i>Resolve autonomamente problemas com os colegas.</i>
31. <i>Está ciente do efeito de suas ações nos outros.</i>

Nota. \* Ítems invertidos: con correlación negativa.





entre los modelos. En conclusión, los resultados indican que la adaptación brasileña de la escala TOCA-R puede ser usada para evaluar cambios longitudinales.

#### *Análisis de confiabilidad*

Las medidas de confiabilidad del alfa de Cronbach y del omega de McDonald son aceptables y permanecen estables en el pre y en el postest. En la Tabla 4 se presentan estos valores.

En particular, las subescalas de comportamiento agresivo y de compromiso con la tarea presentaron una alta confiabilidad en los coeficientes del pretest; mientras que la subescala de socialización tanto antes como después de la intervención presentó la confiabilidad más baja —aunque sigue siendo aceptable—. Por último, el nombre dado a la escala TOCA-R adaptada a Brasil fue *Mapeamento das Interações dos Estudantes* (MINE), cuya traducción sería Mapeo de las Interacciones de los Estudiantes.

### Discusión

MINE, la adaptación brasileña del TOCA-R, probó ser confiable de acuerdo con los resultados del análisis factorial exploratorio y confirmatorio, así como por su estudio de confiabilidad. Para evaluar la dimensionalidad del instrumento adaptado se llevó a cabo un análisis factorial exploratorio en un subconjunto de datos, donde la mejor solución encontrada indicó que cinco factores latentes eran suficientes para explicar más del 60 % de la varianza de los ítems. Las cinco dimensiones fueron consistentes y significativas con la percepción general de los profesores. Asimismo, este análisis factorial exploratorio sugirió la

eliminación de ítems que no estaban correlacionados con ningún factor o que tenían una carga cruzada en más de uno de ellos, por lo cual ocho ítems problema fueron eliminados de la versión traducida del TOCA.

Los 25 ítems seleccionados en el análisis factorial exploratorio y la solución de cinco dimensiones fueron evaluados a través del análisis factorial confirmatorio, con lo que el ajuste general del modelo fue aceptable, indicando así la pertinencia de la solución encontrada tanto para ser generalizada para nuevas muestras como para su invariancia temporal. Las cinco dimensiones demostraron su confiabilidad en los dos coeficientes de consistencia interna, análisis que indicó que las dimensiones encontradas son confiables y estables entre los dos momentos de la aplicación, lo que acentúa una vez más las cualidades psicométricas del instrumento.

Con respecto a las tres escalas que dieron lugar a la adaptación del TOCA-R por parte de AIR, la subescala de competencia social (Conduct Problems Prevention Research Group, 1992) tuvo seis de sus once ítems originales mantenidos, dos ítems de la dimensión “regulación emocional”, denominados “autocontrol” y “socialización” en la versión brasileña, y cuatro ítems sobre la escala de “competencia social”, relacionaron con los factores de “autocontrol” y “socialización”.

De los 25 ítems, 19 provinieron del TOCA-R (Werthamer-Larsson et al., 1991), en relación con sus diversas dimensiones: siete ítems relacionados a lo que ellos definen en el original como “aceptación de la autoridad”, que fueron distribuidos entre los factores llamados “comportamiento agresivo” y “comportamiento disruptivo” en la versión brasileña; tres ítems de la escala original atribuida a la dimensión “hiperactividad/impulsividad” fueron incluidos en

Tabla 3.  
*Medida de invariancia e índices de bondad de ajuste para el TOCA-R adaptado*

Indíces	$\chi^2$ (df)	CFI	RMSEA	SRMR
Variancia configuracional	2835 (1125)	0.893	0.055	0.084
Invariancia métrica	2915 (1145)	0.890	0.056	0.087
Invariancia de escala	3033 (1165)	0.885	0.057	0.088

Tabla 4.  
*Coefficientes de confiabilidad para cada subescala del TOCA-R adaptado*

	Comportamiento agresivo	Comportamiento disruptivo	Compromiso con la tarea	Socialización	Autocontrol
$\alpha$ Pre	0.91	0.89	0.92	0.78	0.82
$\alpha$ Post	0.88	0.86	0.88	0.81	0.82
$\omega$ Pre	0.91	0.89	0.92	0.76	0.83
$\omega$ Post	0.88	0.86	0.88	0.82	0.82

el factor “comportamiento disruptivo” en la escala brasilera; dos ítems de la dimensión “atención/concentración” fueron clasificados como “compromiso con la tarea” en esta versión; tres ítems de “preparación académica” del TOCA-R clasificaron en el instrumento brasileño como “compromiso con la tarea”; y, finalmente, tres ítems relacionados con el “aislamiento social” pertenecen al factor de “socialización” —en su versión invertida— como ítem individual de este factor en la versión brasileña, y se relacionan con el “comportamiento prosocial” en la escala original.

Sin embargo, ninguno de los ítems en la escala de relación entre profesor y estudiante (Pianta & Steinberg, 1992) fue conservado en la escala de la versión brasileña, ya que tenían una baja correlación con los factores sugeridos por el análisis. Por tanto, se verificaron todas las dimensiones del TOCA-R que se mantuvieron en el instrumento brasileño, aunque en un pequeño número, con la intención de reducir los ítems, pero conservando los elementos centrales de la escala; es decir, su validez de contenido y su confiabilidad.

El objetivo de reducir el número de ítems fue validar un instrumento con aplicabilidad a una investigación a gran escala, con perspectivas de reducción de costos y, principalmente, para hacer viable que los maestros respondan la encuesta en el contexto de trabajo en las escuelas públicas brasileñas. El estudio de Koth et al. (2009), en el que se construyó la lista de chequeo del TOCA, confirmó la viabilidad del instrumento para ser autoadministrado por los profesores y apuntó a objetivos similares que sirvieron como base para la validación brasilera. Por consiguiente, es interesante resaltar la proximidad entre los ítems seleccionados en el AFE del MINE y los 21 ítems propuestos para la reducción de la lista de chequeo (TOCA-C; Koth et al., 2009).

A pesar de que el TOCA-C está compuesto por solo tres dimensiones —comportamiento prosocial, comportamiento disruptivo y problemas de concentración—, mientras que en la versión brasileña hay cinco, 16 de los 21 ítems son iguales o equivalentes en las dos opciones en conjunto; los ítems que no aparecen en la versión brasileña del TOCA-C corresponden principalmente a comportamientos disruptivos. Por otra parte, la versión brasileña parece distanciarse un poco más de la versión adaptada del TOCA-R para otras culturas, como en el caso de la descripción realizada por Guzmán et al. (2015) —con un rango de valores alfa de Cronbach de .74 a .05—, pues estos autores describieron la validación de lo que se llamó TOCA-RR (re-revisado) empleado para evaluar el Programa Habilidades para la Vida en escuelas de Chile, adaptado y probado positivamente en su validez —con valores alfa de Cronbach de .74 a .95 en todas las subescalas—.

Esta adaptación chilena fue similar al proceso de adaptación brasileño: se redujo el número de ítems del instrumento, se buscó mantener las características psicométricas en cuando a la validez y confiabilidad presentes en el instrumento original (George et al., 2004), y se buscó facilitar su aplicabilidad en investigación a larga escala. Dicha adaptación terminó en 31 ítems agrupados en seis subescalas que miden la aceptación de la autoridad, el contacto social, la cognición, la madurez emocional, la atención y la actividad. En ese sentido, se pudo verificar que el énfasis de las subescalas son de alguna manera diferentes a lo propuesto en la versión brasileña que proviene del análisis factorial aquí implementado. Sin embargo, los perfiles conductuales son similares —agresivo, dispersivo/disruptivo e introvertido— y se asemejan a los resultados esperados por los programas preventivos de salud mental infantil analizados. En otros países, como en Grecia, también se evaluó esta versión del TOCA y se obtuvieron buenos resultados de consistencia interna (Kourkounasiou & Skordilia, 2014).

Por otra parte, también se encontraron diferentes versiones del TOCA dentro de los Estados Unidos, de la misma manera en que otras versiones apuntaron a su adaptación transcultural a otras culturas. Por esta razón, la escala validada para la adaptación brasileña en contextos escolares se renombró MINE (Mapeo de Interacciones del Estudiante), para definir el horizonte de la variante del instrumento brasileño; con esto, la intención consistía en que pudiera servir de base para futuros análisis sobre la efectividad y la eficiencia del Programa Elos. Sin embargo, como una escala que evalúa varias dimensiones del estado de salud mental en niños, otros estudios pueden utilizarla, haciendo énfasis en los resultados de los cambios comportamentales de los niños.

De este modo, esta escala, ahora validada a la cultura brasileña, puede ser usada en otras investigaciones y proyectos que requieran una evaluación del comportamiento infantil, y puede ayudar a profesores a tener una observación más cuidadosa del comportamiento de sus estudiantes, pues proporciona un uso pedagógico, así como el cumplimiento de su función principal de evaluación de programas en salud mental infantil. Estudios futuros con escalas similares y ya validadas en Latinoamérica podrían llevarse a cabo para adoptar mediciones que permitan comparar datos de comportamientos infantiles entre países.

Finalmente, el análisis permite concluir que las propiedades psicométricas de adaptación del TOCA para Brasil son satisfactorias, aunque se recomienda que los ítems de la escala sean evaluados en una escala de seis puntos, volviendo al modelo de medida del original como estrategia para evitar el efecto de suelo y de techo. Igualmente, para la evaluación del Programa Elos se aconseja considerar el

uso de otros instrumentos para evaluar comportamientos de estudiantes junto con el TOCA-R, preferiblemente basados en técnicas de observación llevadas a cabo por equipos de observadores externos, como ya se usó en las evaluaciones de alguna implementación internacional del GBG. Esto no debe restringir la percepción del profesor, la cual, dependiendo del contexto de la relación en el entorno escolar, puede traer sesgo a la evaluación.

En el proceso de adaptación cultural, una de las limitaciones fue que no se realizó la traducción al idioma original debido al tiempo que tomó empezar la implementación en el Programa Elos en las escuelas, como se acordó con los departamentos municipales de educación. Por otra parte, aunque se considera que este fue un estudio piloto para ajustar los instrumentos y el contenido del programa, aún así cabe resaltar que una de las limitaciones es la ausencia de la aleatorización y de un grupo control, lo cual limita hacer alguna inferencia causal sobre el efecto del programa.

Asimismo, el número de sujetos perdidos para el seguimiento —que correspondió al 53.5 % del grupo inicial— fue también una limitación del estudio. El análisis de la medida de la invariancia asume simplemente que los datos faltantes ocurrieron debido al azar y reduce la muestra a casos completos, y esto, aunque es una suposición sólida, es poco probable dado que la deserción se debió a la falta de personal disponible y no a las características de la muestra.

Por último, otra de las limitaciones relacionada con el análisis de datos fue la falta de agrupación de los datos en múltiples niveles jerárquicos —estudiantes en clases, clases en escuelas, escuelas en ciudades—, y aunque alguno de esos niveles pueda mostrar pequeñas variaciones en momento en que el maestro llenaba el instrumento, era de esperar una varianza considerable entre las diferentes clases. Una solución para este aspecto es usar técnicas de modelado multinivel, pero acá se optó por la simplicidad de un modelo de un solo nivel para facilitar la presentación de resultados y hacerlos comparables para otros estudios del TOCA, en los cuales tampoco se toma en cuenta la dependencia de las observaciones en la clase.

## Referencias

- American Institutes for Research [AIR]. (2010). *TOCA-R and other scales as used in the Good Behavior Game Professional Development Models Trial*. Washington, D. C.: AIR.
- Bayer, J., Hiscock, H., Scalzo, K., Mathers, M., McDonald, M., Morris, A., & Wake, M. (2009). Systematic review of preventive interventions for children's mental health: what would work in Australian contexts? *Australian New Zealand Journal of Psychiatry*, 43(8), 695-710. doi: <https://doi.org/10.1080/00048670903001893>
- Borsa, J., Damásio, B., & Bandeira, D. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: algumas considerações. *Paidéia*, 22(53), 423-432. doi: <http://dx.doi.org/10.1590/1982-43272253201314>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. doi: [http://dx.doi.org/10.1207/S15328007SEM0902\\_5](http://dx.doi.org/10.1207/S15328007SEM0902_5)
- Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorders: The Fast Track Program. *Development and Psychopathology*, 4(4), 509-527. doi: <http://dx.doi.org/10.1017/S0954579400004855>
- Coster, W. J., & Mancini, M. C. (2015). Recomendações para a tradução e adaptação transcultural de instrumentos para a pesquisa e a prática em Terapia Ocupacional. *Revista de Terapia Ocupacional da Universidade de São Paulo*, 26(1), 50-57. doi: <http://dx.doi.org/10.11606/issn.2238-6149.v26i1p50-57>
- Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41 (3-4), 165-407. doi: <http://doi.org/10.1007/s10464-008-9165-0>
- Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, 5(4), 273-297. doi: <https://doi.org/10.1023/A:1020977107086>
- Flay, B. R., Biglan, A., Boruch, R., Castro, F. G., Gottfredson, D., ... Ji, P. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151-175. doi: <http://doi.org/10.1007/s11121-005-5553-y>
- Ford, C. B., Keegan, N., Poduska, J. M., Kellam, S. G., & Litman, J. (2013). *Good Behavior Game Implementation Manual*. Washington, D. C.: American Institute for Research.
- George, M., Squicciarini, A. M., Zapata, R., Guzmán, M. P., Hartley, M., & Silva, C. (2004). Detección Precoz de Factores de Riesgo de Salud Mental en Escolares. *Revista de Psicología de la Universidad de Chile*, 13(2), 9-20. doi: <https://doi.org/10.5354/0719-0581.2012.17651>
- Guzmán, J., Kessler, R. C., Squicciarini, A. M., George, M., Baer L., ... Murphy, M. J. (2015). Evidence for the Effectiveness of a National School-Based Mental Health Program in Chile. *Child & Adolescent Psychiatry*, 54(10), 799-807. doi: <http://dx.doi.org/10.1016/j.jaac.2015.07.005>
- Kaplan, D. (2000). *Structural equation modeling: foundations and extensions*. Newbury Park, C. A.: Sage.
- Kellam, S. G., Mackenzie, A. C., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The Good Behavior Game and the future of prevention and treatment. *Addiction Science & Clinical Practice*, 6(1), 73-84.

- Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3188824/#>
- Kellam, S. G., Wang, W., Mackenzie, A. C., Brown, C. H., Ompad, D. C., Or, F., & Windham, A. (2014). The impact of the good behavior game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science, 15*(1), 6-18. doi: <https://doi.org/10.1007/s11121-012-0296-z>
- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher Observation of Classroom Adaptation-Checklist: Development and Factor Structure. *Measurement and Evaluation in Counseling and Development, 42*(1), 15-30. doi: <https://doi.org/10.1177/0748175609333560>
- Kourkounasiou, M., & Skordilis, E. (2014). Validity and reliability evidence of the TOCA-C in a sample of Greek students. *Psychological Reports: Measures & Statistics, 115*(3), 766-783. doi: <https://doi.org/10.2466/08.11.PR0.115c31z5>
- Murphy, J. M., Abel, M. R., Hoover, S., Jellinek, M., & Fazel, M. (2017). Scope, scale, and dose of the world's largest school-based mental health programs. *Harvard Review of Psychiatry, 25*(5), 218-228. doi: 10.1097/HRP.000000000000149
- Murphy, J. M., Guzmán, J., McCarthy, A. E., Squicciarini, A. M., George, M., ... Jellinek, M. S. (2015). Mental health predicts better academic outcomes: a longitudinal study of elementary school students in Chile. *Child psychiatry and human development, 46*(2), 245-256. doi: <http://doi.org/10.1007/s10578-014-0464-4>
- Nolan, J. D., Houlihan, D., Wanzek, M., & Jenson, W. R. (2014). The Good Behavior Game: A classroom-behavior intervention effective across cultures. *School Psychology International, 35*(2), 191-205. doi: <https://doi.org/10.1177/0143034312471473>
- Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas*. Porto Alegre: Artmed.
- Pedroso, R. T., Abreu S., & Kinoshita, R. T. (2015). Aprendizagens da intersectorialidade entre saúde e educação na prevenção do uso de álcool e outras drogas. *Textura, 17*(33), 9-24. Recuperado de <http://www.periodicos.ulbra.br/index.php/txra/article/view/1339>
- Pérez, V., Fernández, A. M., Rodríguez, J., & de la Barra, F. (2004). Evaluación del efecto de una intervención conductual en la agresividad escolar. *Terapia Psicológica, 23*(1), 91-98. Recuperado de <https://www.redalyc.org/articulo.oa?id=78523110>
- Pérez, V., Rodríguez, J., de la Barra, F., & Fernández, A. M. (2005). Efectividad de una estrategia conductual para el manejo de la agresividad en escolares de enseñanza básica. *Psykhé, 14*(2), 55-62. doi: <https://dx.doi.org/10.4067/S0718-22282005000200005>
- Pianta, R. C., & Steinberg, M. S. (1992). Teacher-child relationships and the process of adjusting to school. En R. Pianta C. (ed.), *Beyond the parent: The role of other adults in children's lives. New directions in child development* (vol. 57, pp. 61-80). San Francisco: Jossey Bass.
- Revelle, W. (2018). Psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston. Recuperado de <https://cran.r-project.org/web/packages=psych>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514. doi: <https://doi.org/10.1007/BF02296192>
- Schneider D. R., Pereira A. P., Cruz J. I., Strelow M., Chan G.; Kurki A., & Sanchez Z. M. (2016). Evaluation of the Implementation of a Preventive Program for Children in Brazilian Schools. *Psicologia: Ciência e Profissão, 36*(3), 508-519. doi: <https://doi.org/10.1590/1982-3703000592016>
- Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The Good Behavior Game: 1969-2002. *Behavior Modification, 30*(2), 225-253. doi: <https://doi.org/10.1177/0145445503261165>
- Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). Effect of First-Grade Classroom Environment on Shy Behavior, Aggressive Behavior, and Concentration Problems. *American Journal of Community Psychology, 19*(4), 585-602. doi: <https://doi.org/10.1007/BF00937993>