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**

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 ($\chi2 = 961$, $df = 265$, 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.

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M.

321

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 ($\chi2 = 961$, 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 (α = .78, .92; ω = .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.

Introduction

To ensure validity of the data collected in intervention studies, it is necessary that the instruments used are evaluated in their ability to identify the studied behaviors. Without proper measurements, even a well-designed clinical trial may not provide valid information about the effectiveness of the program or the results obtained throughout the implementation. Therefore, researchers and professionals who wish to use instruments that have already been validated in other countries must initially assume the task of translating them into their own language and validate them for their use in different cultural contexts (Coster & Mancini, 2015). Then, in order to implement health programs that have proven effective in other countries, it is necessary to adapt not only the programs themselves, but also the instruments that measure their efficiency and effectiveness. It is very important to reflect on the need for cultural adaptation of school mental health programs and their assessment systems for countries with different socio-educational systems, due to the intervention of cultural and psychosocial dimension in behavior problems (Murphy et al., 2017).

Between 2013 and 2016, the General Coordination of Mental Health, Alcohol and Other Drugs of the Ministry of Health of Brazil carried out a project to implement evidence-based international preventive programs, aiming to adapt them to the Brazilian reality and constitute public policies in the field of drug abuse prevention (Pedroso, Abreu, & Kinoshita, 2015, Schneider et al., 2016). A partnership between the Federal University of São Paulo (UNIFESP) and the Federal University of Santa Catarina (UFSC) executed the evaluation of the process and the efficiency of these programs.

The Good Behavior Game (GBG), a North American preventive program for children's mental health, was one of these programs tested and adapted to the reality of Brazilian schools. The United Nations Office on Drugs and Crime (UNODC) recommended it, due to its evidence- based produced in several countries (Bayer et al., 2009, Kellam et al., 2011; Tingstrom et al., 2006). The American Institutes for Research (AIR) advised the Brazilian experience, and his team trained the first program's coaches and suggested the tools for monitoring and evaluating GBG (Schneider et al., 2016). In his Brazilian adaptation, the program was renamed as "Elos - Building Collective", aimed at children from 6 to 10 years old, and implemented by educators in classrooms from the first to fifth year of elementary school education (Schneider et al., 2016).

The GBG / Elos Program does not contain specific curricular content; instead, it is translated into a pedagogical method of classroom management. Children are designated to work together to create a positive learning environment by observing their own behaviors and social interactions. Therefore, the mediation of peer groups, which are based

on daily academic activities, focuses on attitudes such as self-control, task engagement, emotional management and behavioral change (Ford, Keegan, Poduska, Kellam, & Littman, 2013). The main goal is the reduction of aggressive, disruptive and socially isolating behaviors in children (Ford et al., 2013; Kellam et al., 2011).

By focusing on a child's personal and social vulnerabilities, the program addresses the risk factors for future antisocial behaviors, such as problems related to drug use and involvement in violent situations (Kellam et al., 2011; Kellam et al., 2014), acting as a kind of "behavioral vaccine" (Embry, 2002). Thus, by enabling the construction of healthy and inclusive relationships, the GBG / Elos Program improves the children's social and academic abilities, placing itself as a project that acts in the promotion of children's mental health (Schneider et al., 2016). The program has had various applications and cultural adaptations in different countries of the world (Nolan, Houlihan, Wanzek, & Jenson, 2014), including experiences in Latin American countries such as Chile (Pérez, Fernández, Rodríguez, & De la Barra, 2004; Pérez, Rodríguez, De la Barra, & Fernández, 2005).

GBG has as a theoretical base and guide, the concept of Life course/Social field, which conceives that early developmental risk factors are associated with adult problem outcomes and therefore prevention strategies need to be established as early as possible. This theory provides a dual view of mental health, including adaptation, which, on the one hand, is considered as a social dimension that implies how a person, at every stage of their life, is viewed by society and required to perform some social tasks; and, on the other hand, regarded as an individual dimension, geared towards psychological well-being. (Kellam et al., 2011). "According to life course/social field theory, improving the way teachers socialize children in the classrooms will result in the children's improved social adaptation to the classroom social field. The theory also predicts that this early improved social adaptation will lead to a better adaptation in other social fields over the life course". (Kellam et al., 2011, p. 76).

To ensure the quality of a program evaluation effectiveness, such as GBG / Elos Program, it is essential to certify the psychometric properties of the instrument used in the evaluation. Therefore, to guarantee valid and reliable evidence it is fundamental not only to determine the internal and external consistency of the measurement (Pasquali, 2010), but also to assess the internal and external validity of the program, such as the methodological consistency of the results in relation to the implementation and the extrapolation of these results beyond the specific contingencies of the study. (Durlak & DuPre, 2008; Flay et al., 2005).

The main tool proposed for evaluating the effectiveness of GBG is the Teacher Observation of Classroom Adaptation (TOCA) Scale, in which teachers answer questions about their perception of each of their students' behaviors (Werthamer-Larsson, Kellam, & Wheeler, 1991; Koth, Bradshaw, & Leaf, 2009). The Woodlawn Research Center (Chicago, IL) developed this assessment tool more than 40 years ago, around 1975, designed to describe the behavior of each child in the classroom and the tasks requested in that space. The categories created are not intended to be classifications of symptomatic clinical behaviors, but rather assessments of primary school students social adaptation to classroom behaviors, observed and defined by teachers (Koth et al., 2009). The original scale, with 110 items, has undergone several adaptations over various research years (Werthamer-Larsson et al. 1991; Koth et al., 2009, Kourkounasiou & Skordilis, 2014; Wang et al, 2014; Murphy et al., 2014; Guzmán et al., 2015). TOCA-R was the first revision, made in 1991. "The initial set was examined and reduced to 58 items, which were then refined to correspond with the behavioral aspects of DSM- III child disorders criteria" (Werthamer-Larsson et al., 1991, p. 590). In the final version, TOCA-R got 31 items and in the exploratory and confirmatory factor analysis three behavioral factors were defined: (a) social contact versus shy behavior (α = .85); (b) acceptance of authority versus aggressive behavior (α = .92); and (c) concentration problems (α = .96). The test–retest correlation was .75 or higher for each subscale. The original scale is applied in the form of an interview with teachers by technicians trained for the task. (Werthamer-Larsson et al., 1991, Koth et al., 2009).

One of these revisions will be the subject of analysis in this article: the 2010 version of TOCA-R of the American Institute of Research (AIR), which will be detailed later, as it was the basis for the Brazilian adaptation discussed here (AIR, 2010). There are even other previous adaptations for Latin American countries, such as the version of *Teacher Observation of Classroom Adaptation – Re-Revised* (TOCA RR), developed in Chile in the 90's for the evaluation of child mental health programs. This version has good psychometric indicators, with Cronbach alpha values ranging from .92 to .96 on the TOCA-RR's subscales and .74–.95 on the TOCA-R's subscales. (George et al., 2004; Murphy et al., 2014; Guzmán et al., 2015).

Although there are indications of TOCA's high reliability in test-retest evaluations, in addition to their convergent validity with other measures (Werthamer-Larsson, Kellam, & Wheeler, 1991), some studies have found its application expensive and time-consuming in large-scale studies. (Koth, Bradshaw, & Leaf, 2009, Kourkounasiou & Skordilis, 2014). As a result, alternatives were sought

that would reduce costs and application time. One example was the creation of a shortened 24-item version, a checklist entitled TOCA-C, which included the possibility of teacher self-administered application. Some studies have shown that the self-administered questionnaire is an acceptable alternative to the original structured interview format because it requires fewer teacher and researcher combined hours and makes research faster and more financially feasible (Koth, Bradshaw, & Leaf, 2009; Kourkounasiou & Skordilis, 2014).

Therefore, researchers altered and adapted TOCA to use it in the program evaluations. The adaptations are mainly justified by the more appropriate applicability of the instrument to the reality of each educational context, since it constitutes a better measurement of the target outcomes and facilitates their application.

The aim of this article is to describe the psychometric analysis and some aspects of the cultural adaptation of the Teacher Observation of Classroom Adaptation-Revised (TOCA-R) Scale to be used in the future evaluation of the efficacy of the *Elos* Program in Brazilian schools, hoping it will be useful for other child mental health assessments.

Method

Type of study

This psychometric study of the Brazilian adaptation of the TOCA-R scale is based on data obtained from the 2014 *Elos* Program pilot study, whose purpose was to adapt and validate the measurement instruments and the content and method of the preventive program for a future evaluation of efficacy. The evaluative study was longitudinal and had single group quasi-experimental characteristics, in pre and post-test evaluation. Participants were selected in a not randomized way and a control group was not included. Teachers answered the scale before they started the program, in August 2014, and did it again four months later, at the end of the school year.

Cross-cultural adaptation of the TOCA-R scale was based on the steps proposed by Borsa, Damásio & Bandeira (2012): (1) forward translations from the original language to the target language; (2) synthesis of different translations; (3) expert committee review; (4) evaluation by the target population; (5) back-translation; (6) pilot study.

Participants

The Ministry of Health was responsible for implementing the *Elos* Program in Brazilian public schools in the year 2014, and the partner universities were in charge of its

evaluation. The sample included all of the implementation participating students, including all the schools and teachers indicated by the Municipal Departments of Education in the cities chosen for the pilot study.

Sampled students came from 68 elementary school classrooms (first to fifth grade), of 19 public schools in four Brazilian cities (Curitiba, São Bernardo do Campo, Florianópolis, and Tubarão). 68 teachers answered the TOCA adapted for the behavior assessment of 1448 students in the pre-test, and 673 in the post-test. 624 students were matched between the pre and the post-test. 775 students were lost for the follow-up. The losses occurred due to changes in the program implementation team made by the Ministry of Health. Some schools dropped out due to the lack of staff available for supervision

In the final sample (n= 624) 51% were male. 7.6% were 1st grade students, 20.5% were 2nd grade; 35.2%, 3rd grade; 16.1%, 4th grade, and 20.6%, 5th grade. The average age was 8.21 years old.

Instrument

The cross-cultural Brazilian adaptation was based on a version of the TOCA Scale -R. It was devised by the American Institutes for Research in 2010, composed by 55 items measured with a six-point ordinal scale (ranging from "almost never" to "almost always") and derived from three scales (AIR, 2010): a) TOCA-R (Werthamer-Larsson, Kellam, & Wheeler, 1991), which evaluates prosocial and shy behavior, authority acceptance or disruptive behavior, academic readiness and concentration problems; b) Social Competence Subscale (Fast Track Project. In: Conduct Problems Prevention Research Group, 1992), which measures emotion regulation and social competence; and c) Student-Teacher Relationship Scale (Pianta & Steinberg, 1992), which evaluates teacher-student proximity, conflict, and dependence.

Procedures

The final Brazilian adaptation of the TOCA-R scale comprises 33 items answered through a three-point ordinal frequency scale, whose adaptation process will be described below. Each teacher evaluated all of his or her students from a single classroom in two moments: in August 2014, before the program began, and in December 2014, four months after the program was implemented. The teachers received detailed instructions from the program and were coached on how to complete the instrument. Fulfillment was considered in the implementation process as part of the teacher's preparation to carry out the program as it made him/her look in a more detailed manner at their students' behavior,

an important aspect for the organization of the GBG/Elos teams and monitoring student progress. The scale was self-administered by the teachers according to the application model used in the TOCA-C (Koth, Bradshaw, & Leaf, 2009). Teachers had one week to complete the TOCA written form for each of their students. As agreed with the pedagogical staff, they could fill it out during the time dedicated to class preparation or training, or they could fill it out at home if they could not complete the task during their school hours. At the end of the term, the coaching staff would collect the completed instruments from the participating schools and hand them in to the research teams..

Statistical analyses

Statistical analyses used to examine psychometric properties of the adapted TOCA-R scale were: (a) exploratory factor analyses performed with the pre-test participants' subset; (b) confirmatory factor analyses with a held-out sample; (c) longitudinal measurement invariance; and (d) reliability analysis. All analyses were performed in the R programming environment (version 3.3.2), with the help of Psych (version 1.7.3, Revelle, 2017) and Lavaan (version 0.5, Roseel, 2012) packages.

Exploratory Factor Analysis. Due to the modifications made in the TOCA-R scale, an Exploratory Factor Analysis (EFA) was performed to identify the best factorial solution. The analysis was evaluated by cross-validation. First, 70% ($n = 1014$) of the pre-test sample was randomly selected to compose the training data analyzed by EFA. The number of factors was decided based on several criteria: (a) Kaiser's criterion (number of eigenvalues greater than one); (b) Cattel's scree-plot; (c) Horn's parallel analysis; (d) Revelle's Very Simples Structure (VSS) analysis; (e) Velicer's Mean Average Partial (MAP); (f) minimizing Bayesian Information Criterion (BIC); (g) minimizing Root Mean Squared Error of Approximation (RMSEA); and (h) minimizing goodness-of-fit $\chi 2$ statistic. Three criteria evaluated the scale items: item communality should be larger than .5; item complexity should be less than 1.5; and each factor should remain with at least five items. Factor analytic models were fit to the Pearson correlation matrix using maximum likelihood estimator. Although violations of the multivariate normality assumption do not affect point estimates, they might cause underestimation of standard errors and overestimation of the $\chi 2$ statistics (Kaplan, 2000). To address possible violations of multivariate normality, the $\chi 2$ statistics were corrected by the Satorra-Bentler (2001) mean based scaling factor and the standard errors were based on the Satorra-Bentler robust sandwich estimator.

The factor-loading matrix was rotated using the *Oblimin* criterion to obtain a sparse solution with correlated factors.

Confirmatory Factor Analysis. The final model obtained by EFA was cross-validated with a pre-test hold-out sample (n = 434) and a Confirmatory Factor Analysis (CFA) procedure. The CFA model was fitted to the Pearson correlation matrix using the maximum likelihood estimator. To address possible violations of the multivariate normality assumption, as explained in the EFA section above, mean-adjusted $\chi 2$ goodness-of-fit statistic (Satorra-Bentler) and robust standard errors were used. Comparative Fit Index (CFI) evaluated the model fit, the Root Means Squared Error of Approximation (RMSEA) and the Standardized Root Mean Square Residuals (SRMR).

Longitudinal Measurement Invariance. Pre-post comparisons require that the outcome instrument remains stable between applications, reflecting solely the changes caused by the intervention. The final factorial model was fit to the paired pre-post sample ($n = 624$), testing a sequence of increasingly restrictive Longitudinal Confirmatory Factor Analysis models: configurable invariance (same factorial structure in both moments); metric invariance (same factor loadings in both moments) and scalar invariance (same intercept in both moments). Restricted model acceptability, and thus maintenance of the longitudinal measurement invariance hypothesis, was mainly based on the difference in CFI and RMSEA (Cheung & Rensvold, 2002), which were computed from the mean-adjusted χ^2 goodness-of-fit statistic (Satorra, & Bentler, 2001).

Reliability analysis. Cronbach's α and McDonald's ω evaluated each subscale reliability, separately for both moments pre and post-test. Both reliability measures are reported because: (a) Cronbach's α is ubiquitously reported and easily understood in a psychological scale development, even when its assumptions are not theoretically feasible; (b) McDonald's ω gives a better estimate of a scale reliability when the scale is congeneric, which is the case for most psychological scales.

Ethics

The Ethics in Research Committees at the University of São Paulo (#473.498) and the Federal University of Santa Catarina (#711.377) in Brazil approved this study in all its methodological steps, which are in accordance with the requirements of the Declaration of Helsinki. All teachers signed the informed consent to participate in the study. As for the students, due to their age, the term *passive consent*

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M.

325

was adopted, whereby parents were informed of the research by the school management and would manifest whether or not they allowed their child to participate in it.

Results

Brazilian adaptation of the TOCA scale

According to the method proposed by Borsa, Damásio & Bandeira (2012) for cross-cultural instrument adaptation, the tree first steps (1- forward translation from the original language to the target language; 2- synthesis of different translations; 3- expert committee review) were performed as follows: in 2013, the *Elos* pre-pilot study applied the original TOCA-R translation, made by a specialist in the field, with his 55 items. It was applied to the 27 teachers participants from six public schools in São Paulo, Florianópolis, and Tubarão, who evaluated 603 students. Before the program was implemented they applied the scale, and its results were used to evaluate if the translated scale could be used to assess the program in a future study (Schneider et al., 2016).

In 2014, another specialist with a Masters' degree in translation performed the new version from English to Brazilian Portuguese. One expert of the Ministry of Health who coordinated the implementation of *Elos* Program and the four program coaches compared the 2013 translation with the new translation performed in 2014. Each person involved made their own analysis, took notes item by item of their observations and then, in collective meetings, discussed individual evaluations and determined which were the best translations, arriving at a synthesis agreed by all the coaches. Finally, the two researchers responsible for the program evaluation and one Experimental Behavior Analyst (as this approach is one of the main theoretical foundations of GBG) reviewed the final 55 original items in the new Brazilian version.

The teachers interviewed after applying the scale in 2013 complained about the number of TOCA-R items, considering that there were so many items with similar meaning and an excessive number of levels in the ordinal frequency scale, that it was very difficult to complete the evaluation. Filling out the scale for each student in the classroom was considered a grueling activity for most teachers. They said they would not want to participate next year if they had to fill out such a large scale again.

Based on this feedback from the teachers, the two researchers responsible for evaluating the program modified the adapted scale to reduce its size by excluding items that were repeated with similar content. The researchers were careful to maintain the balance of items within each original

dimension: pro-social and shy behavior, acceptance of disruptive authority or behavior, academic readiness, and concentration problems. The final number set was 33 items. The ordinal frequency scale was also reduced from six to three levels (1 - "rarely", 2 - "sometimes", 3 - "frequently"). The program's coaches and their national coordinator reviewed the proposals. A specialist in Experimental Behavior Analysis made its final version and the assessment of the construct validity of the adapted scale.

The other two steps by Borsa et al. (2012) were: 4 - evaluation by the target population and 6 - pilot study, will be described in detail in the sequence of the article. The back translation was the only step that was not performed.

Exploratory Factor Analysis

Criteria for the optimal number of latent factors were not in agreement. Kaiser criterion suggested three factors; *scree-plot* suggested between three and five; parallel analysis suggested five factors; VSS analysis suggested between one and two factors, accordingly to the maximum acceptable complexity; MAP suggested four factors; both minimization of BIC and RMSEA suggested five factors; and the $\chi 2$ goodness-of-fit statistic only reached non-significance at $\alpha = .05$ with more than 16 factors. Item fit based on communality, complexity and a similar number of items per factor was better in the five-factor solution. Since the five-factor solution is suggested by three out of six criteria and has the best item fit, it was chosen as the final solution and is presented below.

Comparisons between the pattern and structure matrices indicated that item-factor correlations are similar to item loadings. Thus, only the factor loading pattern matrix for the five-factor solution is presented in Table 1.

Five factors explained 60% of the total item variance (16%, 13%, 13%, 9%, and 9% for each factor). Goodness-of-fit χ^2 test is statistically significant ($\chi^2 (373) = 1323$, $p < .001$), which is expected, given the sample size, but global fit indices suggested a good fit (RMSEA = .05; TLI = .94, SRMR = .02). The five-factor solution also has the advantage of discriminating between aggressive and disruptive behavior items, as shown next.

Item communality is relatively high. 79% of the items have more than 50% of their variance explained by the five latent factors. 66% of the items have complexity lower than 1.5, which means that, for most items, a single factor is responsible for more than half its variance.

Items 6, 11, 14, 17, 20, 27, 28 (21% of the total number of items) were removed because they did not fit the item criteria described in the Method. Table 2 presents the remaining items and their respective factors. The first and

Table 1.
Training data five-factor solution factor loading matrix (pattern matrix) of TOCA-R adapted*

Item	F3	F1	F2	F4	F5	h^2	U^2	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

* Loadings < 0.3 whose h2 stands for communality, and U2 stands for uniqueness. Comp. stands for mean item complexity.

second factors' items describe aggressive and disruptive behavior, respectively. The third factor's items suggest instruction acceptance and task engagement. The fourth factor's items suggest social skills and prosocial behavior. The fifth factor's items describe self-control and autonomy.

The factor correlation matrix suggests higher-order factors. Aggressive and Disruptive Behavior have a high positive correlation ($r = .77$), and negative correlations with other factors. Task Engagement, Socialization, and Self-control have all positive correlations above .5.

Confirmatory Factor Analysis

The scale simple structure discovered in EFA was used to fit a CFA model to the held-out pre-test data ($n = 434$). Factors' covariance's freely estimated, in line with the high factor correlation presented by the rotated loadings solution. Standardized latent factor and observed indicators based the estimated coefficients, which makes comparisons between EFA and CFA results easier. Figure 1 presents these results in graphic form.

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M.

327

Table 2

Final item list grouped by latent factor obtained from Exploratory Factor Analysis of TOCA-R adapted*

Factor 1: Aggressive behavior
8. *Produz danos ou destrói a propriedade de propósito*
9. *Inicia brigas*
15. *Provoca os colegas*
21. *Grita com os outros*
22. *Machuca os outros fisicamente*

Factor 2: Disruptive behavior
2. *Quebra regras*
12. *Interrompe ou intromete-se na fala dos outros*
23. *É teimoso*
24. *Está fora do lugar com frequência*
26. *Espera a sua vez (Inverted)*

Factor 3: Task engagement
1. *Trabalha bem sozinho (a)*
16. *Completa as tarefas*
18. *Concentra-se nas tarefas*
19. *Presta atenção*
33. *Esforça-se pouco nas atividades (Inverted)*

Factor 4: Socialization
3. *Outras crianças o/a procuram para brincar e realizar atividades*
7. *É socialmente isolado(a) (Inverted)*
10. *Comporta-se timidamente (Inverted)*
13. *Ajuda aos outros, coopera com pares, compartilha com outros*
32. *Passa bastante tempo sozinho (Inverted)*

Factor 5: Self-control
4. *Mostra empatia e compaixão para sentimentos dos outros*
25. *Lida bem com desapontamento ou frustração*
29. *Consegue controlar as emoções*
30. *Resolve autonomamente problemas com os colegas*
31. *Está ciente do efeito de suas ações nos outros*

*Negative correlated items are marked as "(Inverted)".

Global goodness-of-fit indices are acceptable but not optimal (Satorra-Bentler $\chi2(265) = 961, p < 0.001$; CFI = .91; RMSEA = .07 90% CI [.07, .08]; SRMR = 0.09). Modification indices suggest that the source of misfit is mainly a residual correlation between some items and items cross loading in other factors. Since the model fit is already acceptable, and the specification search does not produce substantially different results, the final model was not changed based on modification indices.

Although the model fit is not optimal, factor loadings are stable between EFA and CFA results, with a root mean square deviation of .11. This implies that the loading sign and magnitude is similar between training and test data, suggesting that the construct remains identical between samples.

Longitudinal measurement invariance

Table 3 presents global fit indices and indicates a small difference given the increasingly restrictive invariance models fitted to the pre-post paired sample ($n = 624$).

Likelihood-ratio tests are statistically significant – which is expected, given the sample size – but CFI difference between successive models is lower than .01, suggesting that the longitudinal measurement invariance hypothesis holds (Cheung & Rensvold, 2002). The remaining fit indices, although not as sensitive to violation of measurement invariance, also present negligible differences between models. Results suggest that the Brazilian adaptation of the TOCA-R scale can be used to evaluate longitudinal changes.

Figure 1. Test data (n = 434) CFA results presented as a graph for TOCA-R adapted. Circle nodes indicate latent variables; square nodes indicate observed indicator variables. Single directed arcs indicate factor loadings, and bidirectional arcs indicate covariance (between different nodes) or error variance (in a single node).

Table 3.

Measurement invariance of TOCA-R adapted goodness-of-fit indexes

Indexes	χ^2 (df)	CFI	RMSEA	SRMR
Configural invariance	2835 (1125)	0.893	0.055	0.084
Metric invariance	2915 (1145)	0.890	0.056	0.087
Scale invariance	3033 (1165)	0.885	0.057	0.088

Table 4.

Reliability coefficients for each TOCA-R adapted subscale.

	Aggressive behavior	Disruptive behavior	Task Engagement	Socialization	Self-control
α - Pré	0.91	0.89	0.92	0.78	0.82
α - Pós	0.88	0.86	0.88	0.81	0.82
ω - Pré	0.91	0.89	0.92	0.76	0.83
ω - Pós	0.88	0.86	0.88	0.82	0.82

Reliability analysis

Cronbach's α and McDonald's ω reliability measures are acceptable and remained stable in the pre and post-test. Table 4 presents their values.

The Aggressive Behavior and Task Engagement subscales presented the highest reliability coefficients in the pre-test. The Socialization subscale, both before and after the intervention, presented the lowest, but still acceptable, reliability.

The name given to this TOCA-R scale adapted to Brazil was *Mapeamento das Interações dos Estudantes* (MINE), whose translation is Mapping of Student Interactions.

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M.

329

Discussion

MINE, the Brazilian TOCA-R adaptation, proved to be reliable, according to the results of the exploratory and confirmatory factorial analysis, as well as its reliability study. To evaluate the dimensionality of the adapted instrument, an exploratory factorial analysis was conducted in a subset of the data. The best solution found suggested that five latent factors were sufficient to explain more than 60% of the item variance. The five dimensions were significant and consistent with the general perception of teachers.

The exploratory analysis suggested, on the other hand, the removal of items that were not correlated with the factors or with cross loads in more than one factor. Thus, eight problematic items were removed from the TOCA translated version.

The 25 items selected in the exploratory factorial analysis and their five dimensional solution were evaluated through confirmatory factorial analysis and the overall fit of the model was acceptable, indicating the suitability of the solution found both for being generalizable for new samples and for their temporal invariance. Five dimensions evaluated the reliability using two internal consistency coefficients. This analysis indicated that the dimensions found are reliable and stable between the two moments of application, once again stressing the psychometric qualities of the instrument.

Concerning the three scales that gave rise to the TOCA-R adapted by AIR, the "Social Competence Subscale" (Conduct Problems Prevention Research Group, 1992) had six of its eleven original items maintained, two items of the dimension "emotion regulation", denominated "self-control" in the Brazilian version, and four items of the "social competence" scale, related to the factors of "self-control" and "socialization".

Finally, of the 25 items, 19 came from the TOCA-R (Werthamer-Larsson, Kellam, & Wheeler, 1991), related to its various dimensions: seven items related to what they defined in the original as "authority acceptance", which were distributed among the factors called "aggressiveness" and "disruptiveness" in the Brazilian version. Three items from the original scale attributed to the "hyperactivity/impulsivity" dimension were included in the "disruption" factor in the Brazilian scale. Likewise, two items of the dimension "attention/concentration" were classified as "task engagement" in the Brazilian version. In the same way, three items of academic readiness of the TOCA-R classified in the Brazilian instrument as the factor "engagement in the task". Finally, three items related to "social isolation", which belong to the "socialization" factor in their inverted version, classified as a single item of this factor in the Brazilian version related to "prosocial behavior" in the original scale.

On the other hand, none of the items in the *Student-Teacher Relationship Scale* (Pianta & Steinberg, 1992) was maintained in the Brazilian version, since they had a low correlation with the factors suggested in the analysis. Therefore, it was verified that all the dimensions of the TOCA-R were maintained in the Brazilian instrument, although in a smaller number, in the sense that the reduction of the items was intended, but maintaining the central elements of the scale, that is, their content validity and reliability.

The aim of reducing the number of items was to validate an instrument with applicability in large-scale research, with prospects of cost reduction and mainly of application time, so it would be feasible for teachers to answer it in the work context of Brazilian public schools. The study by Koth, Bradshaw and Leaf (2009), building the TOCA-Checklist, confirmed the feasibility of the instrument to be self-administered by the teacher, aimed at similar objectives and served as a basis for this Brazilian validation. Therefore, it is interesting to note the proximity of the selected items through the exploratory factorial analysis of the MINE (Mapping of Student Interactions) and the 21 items suggested for the reduced TOCA-C scale (Koth, Bradshaw, & Leaf, 2009).

Although the TOCA-C consists of only three dimensions (Pro-Social Behavior, Disruptive Behavior and Concentration Problems), while in the Brazilian version there are five, 16 of its 21 items are the same or equivalent to the selected items. The items that do not appear in the Brazilian version of the TOCA-C mainly concern disruptive behaviors. On the other hand, the Brazilian version seems to distance itself a little more from other adapted versions of the TOCA-R for other cultures, as it is the case of the description by Guzmán et al. (2015) (Cronbach alpha values ranging from .74 to .05). They describe the validation of the so-called TOCA-RR (Re-Revised), used to evaluate the Skills for Life Program in schools in Chile, adapted and tested positively in its validity (Cronbach α values ranging from .74, .95 on all subscales).

This Chilean adaptation was similar to the Brazilian adaptation process: it consisted of reducing the number of items of the instrument, seeking to maintain the psychometric characteristics of content validity and reliability present in the original instrument (George et al., 2004), facilitating

its applicability in large-scale research. It resulted in 31 items clustered into six subscales that measure *acceptance of authority*, *social contact*, *cognition, emotional maturity*, *attention*, and *activity*. In this sense, it was verified that the emphasis of the subscales is somewhat different from the Brazilian version coming from the factorial analysis implemented here. In any case, the behavioral profiles are similar: *aggressive, dispersive/disruptive, withdrawn*, as they approach the main outcomes sought by the child mental health preventive programs analyzed. Other countries, such as Greece, also tested this TOCA version and obtained good internally consistent results (Kourkounasiou & Skordilia, 2014).

Thus, it was observed that there are different versions of TOCA within the United States, in the same way as other versions that aim at their cross-cultural adaptation to other realities. The scale validated by the adaptation to the Brazilian school context was renamed MINE (Mapping of Student Interactions) defining the horizon of the Brazilian instrument's variant. The present study intended to show that it might serve as a basis for future analyses of the effectiveness and efficiency of the *Elos* Program. However, as a scale assessing varying dimensions of children's mental health status, other studies can use it, focusing on the outcome of children's behavioral changes.

This scale, now validated for the Brazilian reality, can be used in other research and projects that require an assessment of children's behavior. It can help teachers to have a more careful observation about the behavior of their students, thus providing a pedagogical use, as well as fulfilling their main function of program evaluation in child mental health. Future studies with similar scales already validated in Latin America may be carried out to adopt measures that allow comparing data on child behavior between these countries.

The analysis allows to conclude that the psychometric properties of TOCA adaptation to Brazil were satisfactory. It is recommended, however, that the scale's items are evaluated on a six- point scale, returning to the original measurement model, as a strategy to avoid the floor effect and the ceiling effect.

Likewise, for the evaluation of Elos Program it is advisable to consider the use of other instruments to evaluate students' behavior with TOCA-R, preferably based on observation techniques of external observer teams, as already used in the evaluation of some international GBG implementations. This is not to be restricted to the teacher's perception, which, depending on the context of the relationship in the school environment, can bring bias to the evaluation.

In the cultural adaptation process, the limitation was the back-translation, which was not done, due to the time it took to start implementing the Elos Program in schools, as had been negotiated with the municipal departments of education.

Although it can be considered that this was a pilot study to adjust both the instruments and content of the program, nevertheless, it is highlighted that one of the limitations is the absence of the randomization and the control group, which limits the causal inference about the effect of the program.

The number of subjects lost for the follow-up (53.5%) is also a study limitation. Measurement invariance analysis simply assumes that data are missing completely at random and reduces the sample to complete cases. Although this is a strong assumption, it is unlikely, given that dropping out is related to a lack of available staff and not to the sample characteristics.

Another limitation, related to data analysis, was the lack of grouping of data in multiple hierarchical levels: students in classes, classes in schools, schools in cities. Although some of these levels may show little variation, as the teacher fills out the instrument, a considerable variance between different classes can be expected. One solution to these data feature is to use multilevel modeling techniques. This research opted for the simplicity of a single-level model to facilitate the presentation of results and make them comparable to other TOCA studies, which also do not take into account the dependence of observations within classes.

References

American Institutes for Research - AIR. (2010). *TOCA-R and other scales as used in the Good Behavior Game Professional Development Models Trial*. Washington, DC: AIR. Unpublished document.

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. 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. 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. 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*

Schneider, D.R., Kaszubowski, E., Garcia, D., Scaff, L., Moraes-Cruz, R. & Sanchez, Z.M.

331

Psychopathology, 4, 509-527. Recovered of https://fasttrac-kproject.org/techrept/s/shs/.

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. 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, 327–350. 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. https://doi.org/10.1023/A:1020977107086

Flay, B. R., Biglan, A., Boruch, R., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175. 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, DC: 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*. Vol. XIII, nº 2, 9-20. https://doi.org/10.5354/0719-0581.2012.17651

Guzmán, J., Kessler, R. C., Squicciarini, A. M., George, M., Baer L., Canenguez, K., Abel, M. R., McCarthy, A., Jellinek, M. S., Murphy, M. J. (2015). Evidence for the Effectiveness of a National School-Based Mental Health Program in Chile. *Child & Adolescent Psychiatry*. Volume 54, Issue 10, Pages 799–807. http://dx.doi.org/10.1016/j.jaac.2015.07.005

Kaplan, D. (2000). *Structural equation modeling: foundations and extensions*. Newbury Park, CA: 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. PMCID: 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. 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. 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. https://doi.org/10.2466/08.11.PR0.115c31z5

Murphy, J.M., Guzmán, J., McCarthy, A.E., et al. (2014). Mental health predicts better academic outcomes: a longitudinal study of elementary school students in Chile. *Child Psychiatry & Human Development*, 46:245–56.

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. Harv Rev Psychiatry;25(5):218–28.

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. https://doi.org/10.1177/0143034312471473

Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas* [Psychological instrumentation: fundamentals and practices]. Porto Alegre: Artmed. 559 p.

Pedroso, R. T., Abreu S., & Kinoshita, R. T. (2015). Aprendizagens da intersetorialidade entre saúde e educação na prevenção do uso de álcool e outras drogas. *Textura*, 17(33), 9-24. Recovered in http://www.periodicos.ulbr.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, 91-98.

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. *Psykhe*, 14, 55-62.

Pianta, R. C., & Steinberg, M. S. (1992). Teacher–child relationships and the process of adjusting to school. In 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.

Satorra, A. & Bentler, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika,* 66(4), 507-514.

Schneider D. R., Pereira A. P. D., 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. 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. 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. https://doi.org/10.1007/BF00937993