



# Effectiveness of the Elos 2.0 Program, a Classroom Good Behavior Game Version in Brazil

Sheila C. Caetano<sup>1</sup> · Marília Mariano<sup>1</sup> · Anderson Ribeiro da Silva<sup>1</sup> ·  
Marília Moreira Mendes-Sousa<sup>1</sup> · Natália Alexandre Ferreira<sup>2</sup> ·  
Hugo Cogo-Moreira<sup>3</sup> · Jair J. Mari<sup>1</sup> · Zila M. Sanchez<sup>1,4</sup>

Accepted: 9 February 2024  
© The Author(s) 2024

## Abstract

The Good Behavior Game (GBG) is a universal classroom behavior management method to reduce disruptive problems. This study aimed to evaluate the effectiveness of the Elos 2.0 Program, a GBG version from Brazil, on the TOCA-C subscales: *concentration problems*, *disruptive behavior*, and *prosocial behavior*. We evaluated 2030 children aged 6–10 years. The Elos 2.0 program had a positive effect on concentration improvement by 0.254 standard deviations (Cohen's  $d=0.254$ , 95%CI=0.038–0.469), and on prosocial behavior by 0.436 standard deviations (Cohen's  $d=0.436$ , 95%CI=0.139–0.734). The odds of no disruptive behavior in Elos 2.0 were 4.88 times (95%CI<sub>OR</sub>=1.89–12.60), compared to the control group. Our findings indicate that the Elos 2.0 program is associated with modest/robust improvements in children's prosocial and concentration skills and a decrease in disruptive behavior. This study is a preliminary support for the implementation of the Elos 2.0 Program in Brazilian public schools. REBEC/Brazil: U1111-1228–2342.

**Keywords** Good Behavior Game · Children · Behavior problems · Prevention · RCT · Brazil

The Good Behavior Game (GBG), a universal classroom behavior management method, is among the most effective programs for promoting mental health in children and adolescents, preventing and/or reducing students' disruptive behavior and increase on-task

---

Sheila C. Caetano and Marília Mariano share first authorship.

✉ Sheila C. Caetano  
sheila.caetano@unifesp.br

<sup>1</sup> Department of Psychiatry and Medical Psychology, Universidade Federal de São Paulo, Rua Major Maragliano, 241, São Paulo, Brazil

<sup>2</sup> School of Nursing, Department of Public Health, Universidade Federal Do Ceará, Rodolfo Teófilo, Fortaleza, Brazil

<sup>3</sup> Department of Education, ICT and Learning, Østfold University College, Halden, Norway

<sup>4</sup> Department of Preventive Medicine, Epidemiology and Biostatistics Section, Universidade Federal de São Paulo, Rua Botucatu, 740, São Paulo, Brazil

behaviors (Joslyn et al., 2019; Smith et al., 2021; Tingstrom et al., 2006). GBG provides teachers with classroom strategies for reducing aggressive or disruptive behavior, shyness, and social isolation. These behaviors are risk factors for future negative mental health-related outcomes, such as depression, violence, and suicide (Johansson et al., 2020; Kellam et al., 2014). Early interventions benefit the short- and long-term mental health of children and adolescents (Baker-Henningham, 2014), which is particularly relevant because up to one-fifth of children worldwide are affected by emotional and behavioral problems (Caetano et al., 2021). Evidence shows that the implementation of preventive interventions in early life that have pre-existing risk factors within educational or familial contexts can have a positive effect on reduction of the incidence and/or progression of mental disorders throughout life (Humphrey et al., 2018; Joslyn et al., 2019).

GBG has been used worldwide, including in the United States (Barrish et al., 1969), Belgium (Saigh & Umar, 1983), the Netherlands (Witvliet et al., 2009), the United Kingdom (Leflot et al., 2010, 2013), Sudan (Coombes et al., 2016;), and Estonia (Streimann et al., 2020). It is an evidence-based practice with positive effects in the following areas: (1) reduction of disruptive and aggressive behavior in the school context (Tingstrom et al., 2006), (2) prevention of the use and abuse of psychoactive substances (Kellam et al., 2014; Poduska et al., 2008; Wilcox et al., 2008), (3) prevention of suicidal ideation and suicide attempts (Kellam et al., 2008; Newcomer et al., 2016; Wilcox et al., 2008), and (4) prevention of the development of violent behavior and antisocial personality disorder in adulthood (Petras et al., 2008). Nonetheless, recent randomized controlled trial (RCT) meta-analytic reviews reported only a modest effect from classroom-based behavioral programs, primarily in reducing the occurrence of conduct problems (Ashworth et al., 2020; Smith et al., 2021; Veenman et al., 2018). However, there were no significant results regarding inattention or teacher-rated shyness, which led the authors to suggest that GBG may not have the effect magnitude originally reported (Smith et al., 2021).

If GBG can present a modest or robust effect, proposing its integration into school curricula would be innovative, considering its potential utility for improving public health in short and long term (Braun et al., 2023; Durlak & DuPre, 2008). GBG may facilitate teaching and learning processes and improve interactions between teachers and students (Kellam et al., 2011). Positive teacher-student and social interactions have been proven to be correlated to better academic and behavioral skills in students. This is primarily because the teacher diminishes negative remarks and fosters a friendly environment using social reinforcers, such as acknowledgments, praise, and descriptions of appropriate behavior (Baker et al., 2008; Gomes & Pereira, 2014). Moreover, GBG seems to diminish negative teacher remarks and mitigate student disruptive behavior, thereby reinforcing positive teacher-student relationships. This phenomenon arises from a reinforcing loop, wherein students' pro-social behavior and heightened attentiveness continually advance, ultimately contributing to improved group dynamics and teacher-student relationships (Holmdahl et al., 2023; Leflot et al., 2010).

In 2013, the GBG program was culturally adapted to and implemented in Brazil for children between 6 and 10 years of age. This program was titled "*Programa Elos: construindo coletivos*" (Elos program: constructing collectives) to promote cooperative and democratic interactions between teachers and students (Lorenzo et al., 2018). The Elos program also aimed to promote the early integration of protective factors to minimize the risks of drug use/abuse in the long term, which was an important outcome identified in other GBG RCTs (Kellam et al., 2014; Petras et al., 2008). The program was implemented in Brazilian schools across 17 municipalities on a quasi-experimental pilot basis in 2014 and 2015, with no control groups (Schneider et al., 2018). An RCT evaluating the effectiveness of

this program in Brazil has not yet been conducted. The program materials were revised in 2018 to include adaptations identified during qualitative studies of the content (Lorenzo et al., 2018). The updated version of the program was named Elos 2.0. Before disseminating this tool as part of a national public policy, it was essential to first test its effectiveness. This was recommended by the PROMISE project, which provides mental health promotion training guidelines and resources for healthcare professionals (Greacen et al., 2012).

This RCT aimed to evaluate the effectiveness of the Elos 2.0 program. Elos 2.0 is a version of the GBG that was adapted in and for Brazil, an upper-middle-income country, to assist in the reduction of problem behavior and promotion of prosocial behavior in children attending public schools. If the Elos 2.0 program proves to have a positive effect, it can then be implemented by the Ministry of Health in Brazilian public schools as a national policy.

## Method

This study was a cluster-randomized controlled trial (RCT) with two parallel groups (intervention and control) of children aged 6 to 10 years (students in the first to fourth years of elementary education) from Brazilian public schools. The schools that were enrolled in this RCT did not offer programs or activities regarding drug use prevention nor the promotion of mental health as of 2019. The National Coordination of Mental Health, Alcohol, and Other Drugs of the Brazilian Ministry of Health (BMH) was responsible for implementing the program in collaboration with the Municipal and State Departments of Health and Education. The inclusion criteria for schools in this study were as follows: (1) being a public school, (2) having at least one class in each grade (first, second, third, and fourth), and (3) belonging to one of the 3 municipalities selected by the BMH. Further details of this specific RCT design can be found in Mariano et al. (2021).

According to the sample size calculation, a sample of at least  $n=3800$  (1900 children in the control group and 1900 in the intervention group; i.e., 19 clusters with an average of 100 individuals per group in each cluster) was needed to achieve a power of 84% to detect a difference of at least 0.2 in the mean scores for the Teacher Observation of Child Adaptation-revised (TOCA) between the two groups. The units of randomization were the schools. Thus, there were no intervention classes offered at control schools and vice versa, as a way to avoid sample contamination. The standard deviation was set at 0.80, based on Storr et al.'s (2002) work. The cluster correlation coefficient was set at 0.050, which is a conservative value for intraclass correlation compared to the previously reported value of 0.039 (Koth et al., 2009). The coefficient of variation for the cluster sizes was set to 0.700. A *t*-test with a significance level of 0.050 was used, and the degrees of freedom were based on the number of clusters. The PASS 15.0 software program's cluster sample calculation module for randomized controlled trials was used to test two proportions; this was based on Donner and Klar's (1996) equation.

We randomly selected 30 schools (15 experimental and 15 control) from 3 cities: Fortaleza and Eusébio (in the northeastern Brazilian state of Ceará) and São Paulo city (in the southeastern Brazilian state of São Paulo). We began by evaluating the schools in the state of Ceara in 2019. In 2020, we intended to introduce the program in São Paulo city; however, its implementation was halted due to the COVID-19 pandemic. Fortunately, the study in Ceará had enough sample size power to allow for an analysis of the program's effectiveness in the 5th poorest of Brazil 27 states (federative units).

## Participants

In 2019, a total of 2030 elementary school children studying in 22 schools in two cities, Fortaleza and Eusébio, in the state of Ceará, were enrolled in the study. There were 917 students (mean age = 7.7 years; 50.3% girls) who received the intervention and 1113 students (mean age = 7.8 years; 51.5% girls) in the control group (see Table 1).

## Randomization Process

Schools were allocated to the control and experimental groups at a ratio of 1:1 per municipality, as they were the units of randomization. This process was performed using a simple random draw from the list of the Anísio Teixeira National Institute of Educational Studies and Research (the Portuguese acronym is *INEP*), which included all Brazilian schools as separated by municipality. All eligible schools ( $n = 30$ ) were randomized via Efron's biased coin algorithm (PASS, 2023), which ensured that both allocation groups would be balanced. Efron's bias was implemented by a biostatistician who did not work in-field with data collection and intervention implementation. Those assessing outcomes were blind to group designation. Figure 1 shows the Elos 2.0 progress flow through the phases of a parallel randomized trial of two groups. Randomization was performed before the baseline assessments.

## Procedure

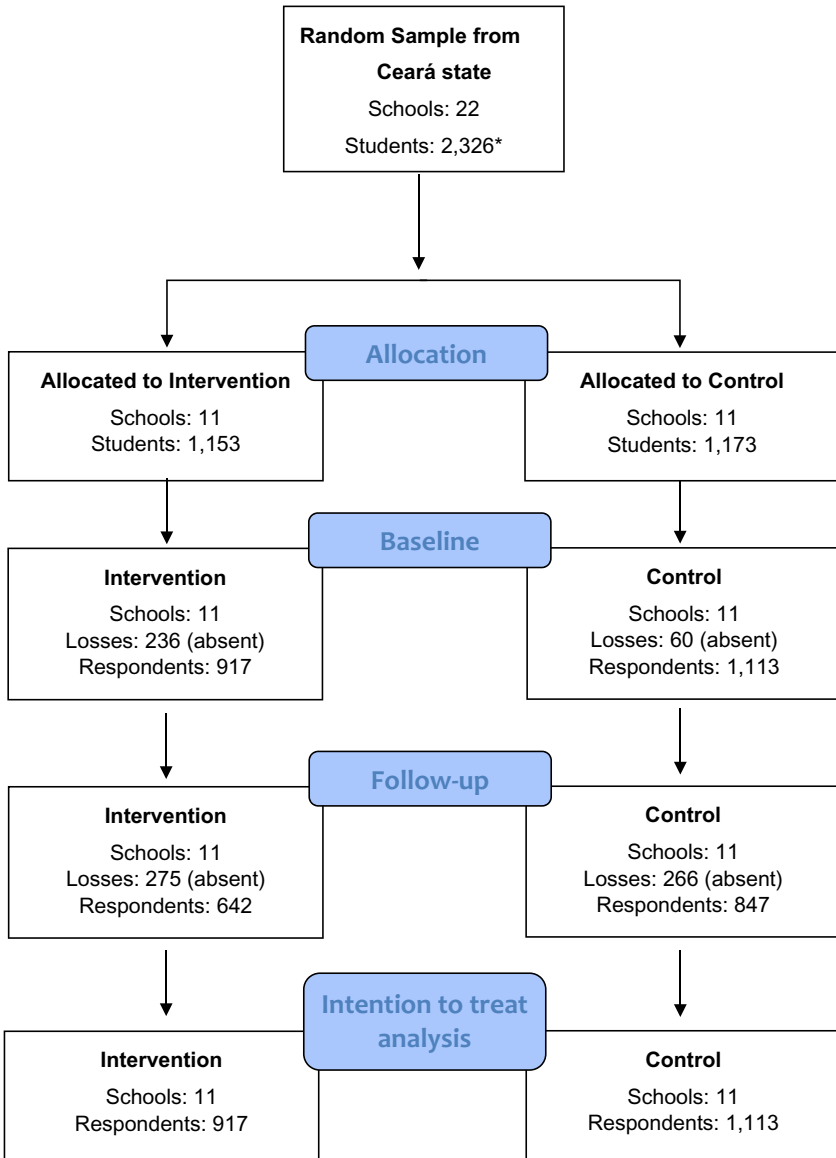
Every school that was invited by the BMH agreed to participate. At the intervention schools, all students in the first through fourth grades participated in the Elos program as a part of their regular school activities. Each school designated one teacher per class who would receive training on the program. In Brazil, there is usually only one teacher in each elementary grade classroom. The BMH trained teachers in Ceará between March and April of 2019. Baseline data were collected from the control and intervention groups simultaneously 2 weeks before the beginning of program implementation in May 2019. The BMH implemented the Elos 2.0 program from June to October, and we collected the follow-up data in November 2019.

This RCT was registered with the Brazilian Ministry of Health Register of Clinical Trials (Registro Brasileiro de Ensaios Clínicos; REBEC) under protocol number U1111-1228-2342 (<https://ensaiosclinicos.gov.br/rg/RBR-86c6jp>). This methodology has

**Table 1** Means and robust standard errors for the three subscales of TOCA

| Measurements           | Control ( $n = 1113$ ) |         |        |       | ELOS ( $n = 917$ ) |         |        |       |
|------------------------|------------------------|---------|--------|-------|--------------------|---------|--------|-------|
|                        | Minimum                | Maximum | Mean   | RSE   | Minimum            | Maximum | Mean   | RSE   |
| Concentration problems | 0                      | 7       | 4.425  | 0.219 | 0                  | 7       | 4.690  | 0.238 |
| Disruptive behavior    | 0                      | 9       | 2.093  | 0.282 | 0                  | 9       | 1.670  | 0.153 |
| Prosocial behavior     | 0                      | 5       | 3.952  | 0.144 | 0                  | 5       | 3.848  | 0.108 |
| Age                    | 5                      | 15      | 7.695  | 0.184 | 5                  | 14      | 7.827  | 0.219 |
| Male                   |                        |         | 51.50% |       |                    |         | 50.30% |       |

RSE, robust standard error (teachers' level)



Legend: \*Students that were enrolled in the classroom before the beginning of classes, not necessarily attending schools. We did not include the sample from São Paulo that did not receive the Program.

Fig. 1 Flow diagram of the RCT to evaluate the effectiveness of the Elos 2.0 Program in the State of Ceará, Brazil

previously been published (Mariano et al., 2021). The Research Ethics Committee at the UNIFESP approved this study (protocol number 3.099.878). All school principals provided informed consent before randomization, and the teachers and parents signed after randomization.

## Measures

The Teacher Observation of Classroom Adaptation Checklist (TOCA-C) (Koth et al., 2009) was used to assess elementary school students' primary outcome behaviors. It was adapted to binary answers ("Never" and "Very often"). The teacher answers 21 items that classify the frequency of each student's behaviors in the classroom. The questions were divided into three dimensions: concentration problems (seven items, e.g., "In the last three weeks, would you say the following statements were never or very often true of this child... pays attention", "stays on task"), disruptive behavior (nine items, e.g., "Breaks rules", "Harms others"), and prosocial behavior (five items, e.g., "Is friendly", "Shows empathy & compassion for others' feelings"). Each dimension outcome is the sum of the score of their answers divided by the total number of items for that subscale. In the original version, the alphas for the TOCA-C subscales ranged from 0.86 to 0.96 (Koth et al., 2009). The instrument was adapted to Brazil with a satisfactory fit of the model ( $\chi^2=961$ ,  $gI=265$ ,  $RMSEA=0.078$  [0.07–0.08 IC95%], and  $CFI=0.9$ ; Schneider et al., 2020). The TOCA-C was completed pre- and post-test for all children by teachers with the mediation of a trained interviewer, who was blind to the group's designation. Teachers notably observed the students for a minimum of 2 months before this pre-test administration. The interviewers were graduate students from the health sector (nurses and psychologists) who were trained by our team.

## Intervention

The teachers underwent a 16-h training program conducted by the BMH during which they received printed material explaining the program, oral presentations, and role-playing. Moreover, at the first implementation of Elos 2.0, the teachers were accompanied by one of the trainers from the BMH.

To participate in the Elos 2.0 program, the teacher followed three steps: (a) they divided the students in the same classroom into heterogeneous teams according to sex, academic performance, and behavior patterns (concentration problems, disruptive behavior, and prosocial); (b) they then reminded each team of the four rules that must be followed (the instructions for activities, voice levels [silence, whisper, group voice, presentation, or street voice], the arrangement of seats [remain seated, stand and walk as agreed, or stand and walk freely], and to be kind [behaviors were combined collaboratively at the beginning of the game by teachers and students]); and (c) teachers announced the start of the game. That is the same procedure adopted in the first GBG version (Barrish et al., 1969). During Elos 2.0, the teachers observed their students, and if the rules were broken, they (1) described the situation to the classroom immediately following the occurrence of inappropriate behavior in a neutral tone of voice; (2) recorded the breaking of the rule on the blackboard, assigning a point to the specific team; and (3) praised the other students who were following the rules.

The rules were illustrated on posters posted in the classrooms while the game took place, similar to the GBG version found in Van Lier et al. (2004). To participate in the Elos

2.0 program, teachers must have selected pedagogical activities that were in accordance with the school curriculum and that could be carried out autonomously by students, such as exercises and artistic activities. Thus, the teacher did not need to offer pedagogical support to the students during GBG, which promoted autonomy and group cooperation.

We also adopted a procedure from Dolan et al. (1993), where one student from each group was the teacher's assistant. Specifically, this student helped to reward their peers at the end of each game. We altered this framework by assigning this task to all children in each group using a weekly rotation. These children had to adhere to the following criteria: the first assistant would be a prosocial child, then followed by the disruptive child, and finally, the child with concentration problems. At the end of 3 weeks, the rotation restarted. This rotation was based on the principles of model learning, where the children learn to behave according to the model of the prosocial peer who spent the first week assisting the teacher.

At the end of the game, the teams that broke no more than five agreements won the game. Noteworthy, all teams can win the game at the same time, which avoids the promotion of competition. The teachers publicly announce the teams that won and do not mention those that lost. The winning teams and students receive prizes (tangible immediate reinforcement, such as stamps in their notebook, colored pencils, and so on). Depending on the progress of the program they may also receive intangible but immediate reinforcement, such as praise and applause. A score is also recorded for teams that win games over a longer period, such as a week, after which delayed reinforcement is given. In summary, awards and recognition for following the rules were provided in three ways: reinforcement given to the student immediately after the game, reinforcement given to the team immediately after the game, and reinforcement given to the team at the end of each week of the game (Barrish et al., 1969; Hansen et al., 2010; Dion et al., 2011; Weis et al., 2015). The teachers had to perform the Elos 2.0 games a minimum of 3 times per week, with each instance lasting between 10 and 30 min, and this lasted for a period of 4 months (see Mariano et al. (2021) for further details).

Notably, the Elos 2.0 program is similar to the PAX GBG version (PAX stands for "peace, productivity, health, and happiness," developed by the Paxis Institute in the United States) (Embry et al., 2016). Compared to other GBGs tested in RCT, the Elos 2.0 and the PAX GBG applied more reinforcement of positive child behaviors in the classroom, thereby decreasing the number of negative punishment processes. Although punitive procedures have been used since the beginning of GBG (Barrish et al., 1969), subsequent adaptations tended to reduce or eliminate punishments (Ialongo et al., 1999; Van Lier et al., 2004; Witvliet et al., 2009). Differently from PAX GBG, the Elos 2.0 program did not use the OK/Not OK desk cards to provide feedback to individuals, and students did not craft a vision of a wonderful classroom, outlining comprehensive lists of elements they wish to observe, hear, engage in, and experience more frequently (Streimann et al., 2017).

Regarding Fidelity, the Elos 2.0 program provided a spreadsheet listing all the steps required for the teacher to execute the game properly, and it was fulfilled by the teacher and sent to the supervisor. The BMH supervisors were the same that provided training to the teachers. When the supervisor was present at the classroom, he/she would observe and fulfill the spreadsheet. The implementation of the Elos 2.0 was divided into the following stages: (1) familiarization, when supervisors would be present at the classroom every 2 weeks; (2) consolidation, when supervisor would be present every month; and (3) expansion, every 2 weeks. In the first phase, visits from supervisors occurred fortnightly. Supervisors were able to observe and assist during the game, correcting if necessary. After the game, the supervisor was able to discuss the fidelity assessment with the teacher. In

the consolidation phase, the autonomy of the teacher became noticeable, as the supervisors only officially observed once a month. Finally, in the expansion phase, the visits were repeated every 2 weeks with the aim of the teacher mastering the game and being prepared to be knowledge spreaders of the activities. The supervisor made the decision regarding when the teacher was ready to move onto the next phase, this decision was made regardless of the time spent in each phase. In general, phase 1 lasted between 6 and 8 weeks, phase 2 for 8–12 weeks, and phase 3 also lasted between 6 and 8 weeks. During the entire implementation process, teachers could review their fidelity assessments and schedule calls with supervisors. Every 4 weeks that a teacher successfully engaged in the program and played Elos, they received a certificate among a closed group of teachers and supervisors on social media (WhatsApp, Facebook) congratulating them. This also occurred when teachers completed a phase. Different certificates were provided on each occasion, and the recipients could customize their next certificate; for example, they could suggest adding their pictures, quotes, and so on to the next certificate. Successful activities were posted in groups. At the end of the program, the teachers received a certificate that listed their total time spent playing and engaging with Elos 2.0.

## Statistical Analysis

Three different regression models were run to evaluate the effects of the Elos program on the three subscales of the TOCA-C. Given the different distributional features of each subscale, different types of models were used. For the concentration problem subscale, linear regression was used due to its continuous nature. For disruptive behavior, zero-inflated Poisson regression was used, given that such an outcome is a count-dependent variable, and approximately 50% of participants scored zero on this outcome. Finally, for prosocial behavior, we used a censored-above regression (continuous outcome, where almost 50% of the sample achieved a ceiling effect for a score of 5).

Regarding disruptive behavior, the zero-inflated Poisson model had two interactive estimated regressions (see Long, 1997). The first evaluates the count-dependent variable for individuals who can assume values of zero and above. For this regression on the count variable, the coefficient was interpreted as the difference between groups on the log count of the number of points in the TOCA's disruptive behavior. The second interactive estimated regression shows the logistic regression of the binary latent inflation variable on the covariates. This predicts the probability of being unable to assume non-zero values and was the focus of this study, with its effect reported in terms of odds ratio (OR). For the *concentration problem* and *prosocial behavior* subscales, the magnitude of the effect was reported in terms of Cohen's *d*.

All the differences between groups' effects are adjusted for sex, age, and baseline measurements of the outcome under assessment. It is crucial to note that, instead of relying on baseline statistical testing on the decision of selecting potential adjustments, we adopted a quantitative policy based on arguments against baseline testing of the total randomized group (Altman, 1985; Senn, 1994) and relevant literature on statistical testing (Molenberghs et al., 2008; Rubin, 1976).

In our attrition analysis, we compared students who provided data at the two different times with those who only completed the initial questionnaire. Supplementary Tables S1 and S2 (Online Resource X) provide a detailed description of the distribution of the baseline continuous outcomes and the socio-demographic features for the groups involved in the attrition analysis.



We assumed missing-at-random, and consistent with intention-to-treat (ITT) analysis, we dealt with missing data using two methods—full information maximum likelihood (FIML) and multiple imputation. We assumed missing-at-random mechanism given that ‘...the empirical data can never provide evidence either for or against the missing at random assumption, and the missing completely at random assumption can be falsified by the data, but it can never be proven true’ (Rhoads, 2012). Additionally, a complete case analysis (which assumes missing completely at random) was performed to evaluate how the FIML and multiple estimates deviated from the former.

In the FIML approach, the model is estimated based on the dependent variables. In our study, these were sex, age, baseline measurements of the outcome under assessment, and group random assignment. In regular regression, the means, variances, and covariances of the dependent variables are not model parameters. However, to avoid missingness from post-assessment measurements (loss of cases), we mentioned the variances of the continuous dependent variables in our model specification (age and baseline assessment). Distributional assumptions were then made regarding these variables, and their means, variances, and covariances are used as model parameters. Assuming missing-at-random mechanism, FIML estimates were unbiased and more efficient than listwise deletion (i.e., complete case analysis, a common routine implemented in statistical software), as well as pairwise deletion and similar-response-pattern imputation (for further details regarding FIML see Enders & Bandalos, 2001).

As a second approach to dealing with missing data and complying with the ITT paradigm and also assuming missing-at-random mechanism, we utilized multiple imputations carried out using the Bayes estimation of an unrestricted variance–covariance model, which was then used to impute the missing values. Regarding the unrestricted model used for imputation, the regression setting available in Mplus (Muthén & Muthén, 2019) version 8.4 was used. This type of setting is recommended when all variables with missing values are of the same nature (e.g., in this study, all variables were considered to be categorical to avoid computation of integer values commencing from continuous assumptions). The following variables were included in the imputation step: random allocation status (control vs. Elos), age, sex, and both baseline as well as follow-up measurements. Fifty imputation datasets were generated and used in the subsequent analysis found in Rubin’s (1987) work.

The estimator used for all analyses was the maximum likelihood with robust standard errors (MLR), which allows for dealing with the non-independence of the observation (e.g., in this study, children nested in classrooms). The standard error was then computed considering the multilevel structure using the TYPE=Complex command in MPlus, as proposed by Asparouhov (2006), using a sandwich estimator. The level of significance was set at 5%.

## Results

Table 1 shows the means and robust standard errors for the three TOCA subscales at the baseline assessment. The robust standard error accounted for 76 classes. The groups that received Elos 2.0 as well as the control groups were all similar in terms of age and sex of the student populations.

Table 2 displays the ICC coefficients for the three subscales at the measurement and establishment of the baseline. This analysis shows that almost 30% of the variance in the assessments comes from teacher-class interactions.

**Table 2** ICC for the three subscales' baseline assessments

| Measurements           | Estimate | SE    | 95% confidence interval |       |
|------------------------|----------|-------|-------------------------|-------|
| Concentration problems | 0.234    | 0.030 | 0.166                   | 0.302 |
| Disruptive behavior    | 0.274    | 0.030 | 0.200                   | 0.348 |
| Prosocial behavior     | 0.297    | 0.039 | 0.220                   | 0.375 |

Table 3 shows the standardized effect sizes, confidence intervals [CI], and  $p$  values for both approaches (FIML and multiple imputation) for the three scale outcomes; this is adjusted for cluster effects, age, sex, and measured baseline assessment. The Elos 2.0 program has a positive effect on all three outcomes. In terms of concentration problems (where the higher the score, the better the concentration), students allocated to Elos schools showed an improvement in concentration by 0.254 standard deviations (95% CI=0.038–0.469).

Regarding disruptive behavior in participants enrolled in Elos schools, the log-odds of membership to the excess zero-generating process (i.e., having no disruptive behavior) increased by 1.586. This corresponded to an odds ratio of 4.88 (95% CI=1.89 to 12.60) after adjusting for baseline disruptive behavior, sex, and age. In other words, the odds of no disruptive behavior being observed in classrooms where Elos 2.0 was deployed are 4.88 times the odds of no disruptive behavior being observed among students nested in the control schools. This demonstrates an increase of 1.586 of having no disruptive behavior in environments where Elos 2.0 is being used. For the count aspect of the zero-inflated regression, under the three scenarios (i.e., complete cases, MI, and FIML), there is a lack of evidence regarding group differences.

Finally, for the prosocial domain (the higher the score, the more prosocial behavior), there was a close-to-moderate effect size in favor of Elos schools (Cohen's  $d=0.436$ , 95%

**Table 3** Standardized effect sizes for the three scales' outcomes after adjusting for cluster effects

| Outcome                | Approach dealing with missingness        | Estimate | 95% confidence interval |        | $p$ value        |
|------------------------|--|----------|-------------------------|--------|------------------|
| Concentration problems | FIML                                     | 0.254†   | 0.038                   | 0.469  | <b>0.024</b>     |
|                        | MI                                       | 0.241†   | 0.068                   | 0.414  | <b>0.006</b>     |
|                        | Complete case                            | 0.252†   | 0.037                   | 0.467  | <b>0.022</b>     |
| Disruptive behavior    | FIML (logistic regression part)          | 4.855*   | 1.892                   | 12.603 | <b>0.001</b>     |
|                        | MI (logistic regression part)            | 4.096*   | 2.075                   | 8.431  | <b>&lt;0.001</b> |
|                        | Complete case (logistic regression part) | 4.884*   | 1.890                   | 12.629 | <b>0.001</b>     |
|                        | FIML                                     | 0.173‡   | -0.105                  | 0.452  | 0.223            |
| Prosocial behavior     | MI                                       | 0.035‡   | -0.190                  | 0.261  | 0.062            |
|                        | Complete case                            | 0.173‡   | -0.106                  | 0.452  | 0.223            |
|                        | FIML                                     | 0.436†   | 0.139                   | 0.734  | <b>0.004</b>     |
|                        | MI                                       | 0.465†   | 0.237                   | 0.693  | <b>&lt;0.001</b> |
| Prosocial behavior     | Complete case                            | 0.436†   | 0.138                   | 0.734  | <b>0.004</b>     |

FIML, full information maximum likelihood; MI, multiple imputation; †Cohen's  $d$  effect size; ‡difference between groups in log count of number of points of TOCA's disruptive behavior; \*odds ratio derived from the latent inflation (dichotomous) variable from the zero-inflated Poisson regression

Values in bold indicate  $p$ -value  $<0.05$

CI=0.139 to 0.734). The participants exposed to Elos showed an average 0.436 standard deviation more prosocial behavior than the control group.

Of the three outcomes, FIML and multiple imputations agreed with each other, showing the robustness of the estimates. Supplementary Tables S1 and S2 (Online Resource X) provide details of attrition across random group assignments.

## Discussion

Elementary school students who participated in Elos 2.0 showed a significant increase in their prosocial and concentration scores at the end of the intervention when compared to children in the control groups. The effects that Elos 2.0 had on reducing disruptive behavior in the classroom were significant and robust; this trial found that children who participated in Elos 2.0 were 4.8 times more likely not to exhibit disruptive behavior than those in the control classrooms. Thus, our study demonstrates the effectiveness of the Elos 2.0 program regarding three behaviors among first- to fourth-grade students.

Our results align with the meta-analysis of RCTs for classroom behavioral programs by Veenman et al. (2018), which reported significant effects on teacher-rated concentration and students' disruptive behavior. Furthermore, we present a modest-to-robust effect, in contrast to the minimal effect found in the meta-analysis. One reason for the difference in our effect size was the large reduction in our initially calculated sample size due to the COVID-19 pandemic and subsequent school closures. Moreover, Smith et al. (2021) conducted a meta-analysis including only the results of GBG RCTs and presented similar outcomes. The exception was the teacher-rated shyness, for which they did not obtain a positive result. Notably, the positive results from this adapted GBG seemed to be more closely related to an increase in the scores of children with or at high risk for behavioral and emotional problems than in children who were not at high risk (Ashworth et al., 2020; Bowman-Perrott et al., 2013; Bradshaw et al., 2009; Streimann et al., 2020).

Although our intervention was universal (rather than limited to at-risk children with higher scores for disruptive behavior), Brazil is among the less positive school disciplinary climates globally, which may have impacted our results. For example, the Teaching and Learning International Survey (TALIS) reports that an average of 12.7% of classroom time is used keeping a classroom in order; Brazilian teachers spent an additional 19.8% of their time doing so. Similarly, students in Brazil were ranked 74th out of 77 countries for their cooperativeness (−0.35 according to the Programme for International Student Assessment—PISA Index, 2018; OECD, 2015). According to the PISA (2018), Brazil's disciplinary climate in language-of-instruction lessons is the penultimate-worst in comparison to other countries studied (−0.37 PISA Index, rank 75/76, 2018; OECD, 2018).

The quality of education remains a challenge in several countries, most of which are LMICs or have low-income communities (OECD, 2018). In these populated areas, evidence-based prevention programs in schools represent an enormous opportunity to reduce the impact of poverty on children's mental health (Belkin et al., 2017). Furthermore, as a universal intervention, GBG may reduce stressful social interactions and increase students' chances of high school graduation and college attendance by providing a better school disciplinary climate and a sense of belonging for the students (Bradshaw et al., 2009; Streimann et al., 2020).

Overall, GBG programs are intended to increase peer reinforcement to achieve prosocial actions and decrease behavioral problems that could undermine positive peer and

student–teacher relationships (Johansson et al., 2020; Newcomer et al., 2016; Witvliet et al., 2009). Johansson et al. (2020) explained that the set of rules used to play a GBG provides a scenario where students can experience the consequences of rule breaking as well as the recognition and reward that accompany appropriate behavior.

The Elos 2.0 program methodology also prioritizes strategies to reinforce positive and appropriate behaviors, such as obedience of rules, cooperative play, and autonomy, which may improve social skills. In Brazil, children aged 4–5 years old showed a 30% delay in socioemotional development in preschools, which is approximately 50% more than the worldwide prevalence (Caetano et al., 2021).

The Elos 2.0 program is similar to previous version of GBG, in particular to the most important GBG version, PAX GBG (Embry et al., 2016), regarding main procedural characteristics. In general, we observed that PAX GBG and other versions of GBGs prioritized the display of rules in a visible place in the classroom, in addition to signaling the exact moment when the game started. The signal could be a specific sound or a pre-established phrase like “The game starts now!” The teachers also selected team leaders and assistants for the duration of the game. This is important because the student team leader assists the teacher in conducting the activities, specifically by distributing the materials to their peers, in addition to serving as a positive role model that other students in the group can imitate (Husband & Chong, 2011). It is worth highlighting that in the subsequent versions of the GBG, teachers and students democratically agreed upon classroom rules. During the game, we observed that the latest GBG versions gradually increased the reinforcing procedures, despite maintaining at least one punitive procedure. Finally, all the evaluated studies emphasized the gradual increase in reinforcing procedures at the end of the game, both immediate and delayed, tangible (e.g., stickers and school supplies) as well as intangible (e.g., praise and positive social recognition). Meanwhile, all studies carried out after the original version withdrew from punitive procedures in the final stage of the game.

Moreover, the PAX GBG was adopted by incorporating extra activities and elements aimed at enhancing compliance and classroom management. The PAX GBG incorporates proven kernels or strategies for influencing behavior (Domitrovich et al., 2010). These kernels are classified into four categories—antecedent, relational, physiological, and reinforcement—each with demonstrated effectiveness in shaping behavior. They aid children in preparing for and attaining goals, reducing anxiety, and providing rewards for desirable behavior (Embry et al., 2016). Regular utilization of these kernels affords students opportunities to practice prosocial skills and self-regulation (Streimann et al., 2017).

Based on a comparison of the diverse applications of GBG in RCT, the Elos 2.0 program had a shorter duration (four months). Veenman et al. (2018) demonstrated that GBG programs with shorter durations had increased effectiveness. This might be related to the teacher burnout and general non-compliance that often occurs with lengthier programs (Berg et al., 2017; Bradshaw et al., 2020). Moreover, we had a fidelity assessment for each game, and the teachers could assess the supervisor, increasing their support. We also intended to increase their engagement using gamification elements, such as positive reinforcement, customized prizes, feedback, and challenges. This new teacher approach should be further investigated as a potential strategy to decrease teacher burnout. Compared to PAX GBG, we did not use the OK/Not OK desk cards to provide feedback to individuals (Streimann et al., 2017), because the Elos 2.0 prioritizes strategies for reinforcing the positive and adequate behaviors that students should exhibit in the classroom.

Considering that the Elos 2.0 program showed effects on increasing prosocial and decreasing disruptive behavior, which are two important risk factors for drug use (NIDA, 2020), we expect that this program can also lead to a reduction in future

drug use among these students. This evidence suggests that the federal government's strategy to include Elos 2.0 as one of its drug-use prevention programs through the School Health Program (PSE) of the Unified Health System (SUS)—Federal decree 6.286/2007—will have positive effects on Brazil's health policy.

This study had several limitations. First, our study ultimately had a smaller sample size than intended. We began scheduling our RCT of the Elos 2.0 program in São Paulo in 2020, but were forced to stop prematurely because of the COVID-19 pandemic. Therefore, we analyzed only the data collected from the sample in Ceará, where the RCT was conducted in 2019. Nevertheless, even this smaller sample size demonstrated a modest/robust improvement in the outcomes. Second, the game was not repeated if a child was absent on a specific day, but this is a common limitation of the prevention intervention designs (Ariza et al., 2013). Third, we were not able to directly assess children because of the high cost of such observations. However, rating scales completed by teachers constitute the most common method used to evaluate children's behavior and functioning at school. Pas et al. (2015) noted that if the teachers who provide the scale ratings are also the teachers who implement the intervention, they would know whether they were in the control or intervention group, which may have influenced their ratings. Nevertheless, the BMH supervisors observed and completed an implementation spreadsheet for the first game (to verify whether students' teams were divided appropriately as planned) and for the subsequent visits. We randomized our sample by school, avoiding the contamination effect that can occur when randomization is performed in the classroom. A student could have characteristics in more than one group; for example, a student could have concentration problems and exhibit disruptive behavior. However, most GBG studies divided their samples in the same manner. Finally, we performed a follow-up assessment only 1 month after the completion of the intervention. Due to the COVID-19 pandemic, we were unable to continue to perform re-evaluations.

In conclusion, our findings indicate that Elos 2.0, as a universal intervention of teacher behavior management, is associated with modest/robust improvements in children's prosocial behavior and concentration skills as well as a decrease in disruptive behavior among all students, regardless of their behaviors in the classroom. This study is a preliminary support for the national implementation of the Elos 2.0 program across Brazilian public schools. More rigorously designed trials with classroom-observed measures that include all Brazilian regions are needed.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11469-024-01256-6>.

**Acknowledgements** We acknowledge past and present workers at the Ministry of Health, who believe and invest in prevention, especially Lidiiane Nogueira Rebouças. We would like to thank Editage ([www.editage.com](http://www.editage.com)) for English language editing.

**Author Contribution** The submitted manuscript was approved by all authors. The study design and grant protocol were elaborated by SCC and ZMS. SCC, MM, and ARS were responsible for the literature review and drafting of the article. MMMS and NAF contributed to the data interpretation. HCM calculated the sample size and school randomization. MM, ARS, MMMS, and NAF organized the data collection and dataset. HCM and JJM were critical reviewers for the study design and manuscript.

**Funding** This study is part of the Research and Innovation Grant for the Prevention of Mental Disorders and Use of Alcohol and other Drugs ("Pesquisas e Inovações em Prevenção de Transtornos Mentais e Uso de Álcool e Outras Drogas"), granted by the Brazilian Ministry of Health (TED #176/2017). The funder's roles consist of selecting the most suitable grant applications and implementing the Elos 2.0 Program.

**Data Availability** The data that support the findings of this study are available on request from the corresponding author (SCC).

## Declarations

**Ethics Approval** The Research Ethics Committee at the Universidade Federal de São Paulo approved this study under protocol number 3.099.878. This study was performed in accordance with the principles of the Declaration of Helsinki.

**Consent to Participate** All schools' principals signed informed consent forms before randomization, and teachers and parents signed after randomization.

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Altman, D. G. (1985). Comparability of randomised groups. *The Statistician*, *34*(1), 125. <https://doi.org/10.2307/2987510>
- Ariza, C., Pérez, A., Sánchez-Martínez, F., Diéguez, M., Espelt, A., Pasarín, M. I., Suelves, J. M., de la Torre, R., & Nebot, M. (2013). Evaluation of the effectiveness of a school-based cannabis prevention program. *Drug and Alcohol Dependence*, *132*(1–2), 257. <https://doi.org/10.1016/j.drugalcdep.2013.02.012>
- Ashworth, E., Humphrey, N., & Hennessey, A. (2020). Game over? No main or subgroup effects of the good behavior game in a randomized trial in English primary schools. *Journal of Research on Educational Effectiveness*, *13*(2), 298–321. <https://doi.org/10.1080/19345747.2019.1689592>
- Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics - Theory and Methods*, *35*(3), 439–460. <https://doi.org/10.1080/03610920500476598>
- Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*(1), 3–15. <https://doi.org/10.1037/1045-3830.23.1.3>
- Baker-Henningham, H. (2014). The role of early childhood education programmes in the promotion of child and adolescent mental health in low- and middle-income countries. *International Journal of Epidemiology*, *43*(2), 407–433. <https://doi.org/10.1093/ije/dyt226>
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*(2), 119–124. <https://doi.org/10.1901/jaba.1969.2-119>
- Belkin, G., Wissow, L., Lund, C., Aber, L., Bhutta, Z., Black, M., Kieling, C., McGregor, S., Rahman, A., Servili, C., Walker, S., & Yoshikawa, H. (2017). Converging on child mental health – toward shared global action for child development. *Global Mental Health*, *4*. <https://doi.org/10.1017/gmh.2017.13>
- Berg, J., Bradshaw, C. P., Jo, B., & Ialongo, N. S. (2017). Using complier average causal effect estimation to determine the impacts of the Good Behavior Game preventive intervention on teacher implementers. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*, 558–571. <https://doi.org/10.1007/s10488-016-0738-1>
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Parker, R., & Greenwood, C. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *In School Psychology Review*, *42*, 1. <https://doi.org/10.1080/02796015.2013.12087490>

- Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology, 101*(4), 926–937. <https://doi.org/10.1037/a0016586>
- Bradshaw, C. P., Shukla, K. D., Pas, E. T., Berg, J. K., & Ialongo, N. S. (2020). Using complier average causal effect estimation to examine student outcomes of the PAX Good Behavior Game when integrated with the PATHS curriculum. *Administration and Policy in Mental Health, 47*, 972–986. <https://doi.org/10.1007/s10488-020-01034-1>
- Braun, S. S., Bradshaw, C. P., Beahm, L. A., Budavari, A. C., Downer, J., Ialongo, N. S., & Tolan, P. H. (2023). Predicting implementation of the PAX Good Behavior Game + MyTeachingPartner interventions. *Frontiers in Psychology, 14*, 1059138. <https://doi.org/10.3389/fpsyg.2023.1059138>
- Caetano, S. C., Ribeiro, M. V. V., Askari, M. S., Sanchez, Z. M., do Rosário, M. C., Perissinoto, J., Resegue, R., Felix, E., Mariano, M., Fidalgo, T. M., Caetano, M., Mari, J. J., Surkan, P. J., & Martins, S. S. (2021). An epidemiological study of childhood development in an urban setting in Brazil (Sao Paulo, Brazil: 1999). *Revista Brasileira de Psiquiatria, 43*(1), 43–54. <https://doi.org/10.1590/1516-4446-2020-0934>
- Coombes, L., Chan, G., Allen, D., & Foxcroft, D. R. (2016). Mixed-methods evaluation of the Good Behaviour Game in English primary schools. *Journal of Community & Applied Social Psychology, 40*(5)(2013), 390–405. <https://doi.org/10.1002/casp.2268>
- Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomized study. *Prevention Science, 12*(1), 70–79. <https://doi.org/10.1007/s1121-010-0182-5>
- Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., ... Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14*(3), 317–345. [https://doi.org/10.1016/0193-3973\(93\)90013-L](https://doi.org/10.1016/0193-3973(93)90013-L)
- Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools, 47*(1), 71–88. <https://doi.org/10.1002/pits.20452>
- Donner, A., & Klar, N. (1996). Statistical considerations in the design and analysis of community intervention trials. *Journal of Clinical Epidemiology, 49*(4), 435–439. [https://doi.org/10.1016/0895-4356\(95\)00511-0](https://doi.org/10.1016/0895-4356(95)00511-0)
- 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. <https://doi.org/10.1007/s10464-008-9165-0>
- Embry, D. D., Fruth, J. D., Roeppcke, E. F., & Richardson, C. (2016). *PAX Good Behavior Game manual*. Tucson: PAXIS Institute.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. [https://doi.org/10.1207/S15328007SEM0803\\_5](https://doi.org/10.1207/S15328007SEM0803_5)
- Gomes, R. M. S., & Pereira, A. S. (2014). Influence of age and gender in acquiring social skills in Portuguese preschool education. *Psychology, 05*(02), 99–103. <https://doi.org/10.4236/psych.2014.52015>
- Greacen, T., Jouet, E., Ryan, P., Cserhati, Z., Grebenc, V., Griffiths, C., Hansen, B., Leahy, E., Silva, K. M., Sabić, A., De Marco, A., & Flores, P. (2012). Developing European guidelines for training care professionals in mental health promotion. *BMC Public Health, 12*, 1114. <https://doi.org/10.1186/1471-2458-12-1114>
- Hansen, W. B., Bishop, D. C., & Jackson-Newsom, J. (2010). Impact of a classroom behavior management intervention on teacher risk ratings for student behavior. *Journal of Drug Education, 40*(1), 81–90. <https://doi.org/10.2190/DE.40.1.f>
- Holmdahl, A., Schad, E., Nilsson, G., & Kaldo, V. (2023). More than just a game: Teachers' experiences of the PAX good behavior game. *European Journal of Psychology and Educational Research, 6*(1), 55–68. <https://doi.org/10.12973/ejper.6.1.55>
- Humphrey, N., Hennessey, A., Ashworth, E., Wo, L., Frearson, K., Petersen, K., Panayiotou, M., Lendrum, A., Wigelsworth, M., Birchall, E., Squires, G., & Pampaka, M. (2018). *Good behaviour game evaluation report and executive summary*. Education Endowment Foundation. [https://educationendowmentfoundation.org.uk/public/files/GBG\\_evaluation\\_report.pdf](https://educationendowmentfoundation.org.uk/public/files/GBG_evaluation_report.pdf). Accessed 3 Nov 2021
- Ialongo, N., Werthamer, L., Kellam, S., Brown, C., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*(5), 599–641. <https://doi.org/10.1023/A:1022137920532>

- Johansson, M., Biglan, A., & Embry, D. (2020). The PAX Good Behavior Game: One model for evolving a more nurturing society. *Clinical Child and Family Psychology Review*, 23(4), 462–482. <https://doi.org/10.1007/s10567-020-00323-3>
- Joslyn, P. R., Donaldson, J. M., Austin, J. L., & Vollmer, T. R. (2019). The Good Behavior Game: A brief review. *In Journal of Applied Behavior Analysis*, 52(3), 811–815. <https://doi.org/10.1002/jaba.572>
- Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence*, 95(SUPPL. 1), S5–S28. <https://doi.org/10.1016/j.drugalcdep.2008.01.004>
- 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.
- Kellam, S. G., Wang, W., Mackenzie, A. C. L., Brown, C. H., Ompad, D. C., Or, F., Ialongo, N. S., Poduska, J. M., & 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(S1), 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>
- Lefflot, G., Van Lier, P. A. C., Onghena, P., & Colpin, H. (2013). The role of children's on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. *Journal of School Psychology*, 51(2), 187–199. <https://doi.org/10.1016/j.jsp.2012.12.006>
- Lefflot, G., van Lier, P. A., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology*, 38(6), 869–882. <https://doi.org/10.1007/s10802-010-9411-4>
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications. ISBN: 9780803973749.
- Lorenzo, F. M., Aveiro, A. G., Vieira, A. G., Ferreira, D. C., Araújo, L. A. S., Tibúrcio, R. R. R., & Gehm, T. P. (2018). Do Good Behavior Game ao Programa Elos: A adaptação transcultural de um programa de prevenção infantil. In Sanchez, Z., M., Schneider, D., R., & Pedroso, R., T. (Eds.) *Prevenção ao uso de drogas: implantação e avaliação de programas no Brasil* (1<sup>st</sup> ed., pp. 151–179). Ministério da Saúde.
- Mariano, M., da Silva, A. R., Lima, J. L. S., de Pinho, N. T., Cogo-Moreira, H., Melo, M. H. S., Mari, J. J., Sanchez, Z. M., & Caetano, S. C. (2021). Effectiveness of the Elos 2.0 prevention programme for the reduction of problem behaviours and promotion of social skills in schoolchildren: Study protocol for a cluster-randomized controlled trial. *Trials*, 22(1). <https://doi.org/10.1186/s13063-021-05408-0>
- Molenberghs, G., Beunckens, C., Sotto, C., & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(2), 371–388. <http://www.jstor.org/stable/20203830>
- Muthén L. K. & Muthén B. O. (2019). *Mplus: Statistical analysis with latent variables: User's guide* (Muthén & Muthén, Ed., 8th ed.).
- Newcomer, A. R., Roth, K. B., Kellam, S. G., Wang, W., Ialongo, N. S., Hart, S. R., Wagner, B. M., & Wilcox, H. C. (2016). Higher childhood peer reports of social preference mediates the impact of the good behavior game on suicide attempt. *Prevention Science*, 17(2), 145–156. <https://doi.org/10.1007/s11121-015-0593-4>
- NIDA. (2020, May 24). *Principles of adolescent substance use disorder treatment: A research-based guide*. <https://www.drugabuse.gov/publications/principles-adolescent-substance-use-disorder-treatment-research-based-guide/acknowledgements> on 2021, November 5
- OECD. (2015). *Education Policy Outlook Brazil*. <https://www.oecd.org/education/Brazil-Country-Profile.pdf>. Accessed 3 Nov 2021
- OECD. (2018). *The Programme for International Student Assessment (PISA) 2018 Brazil*. [https://www.oecd.org/pisa/publications/PISA2018\\_CN\\_BRA.pdf](https://www.oecd.org/pisa/publications/PISA2018_CN_BRA.pdf). Accessed 3 Nov 2021
- Pas, E. T., Bradshaw, C. P., Becker, K. D., Domitrovich, C., Berg, J., Musci, R., & Ialongo, N. S. (2015). Identifying patterns of coaching to support the implementation of the Good Behavior Game: The role of teacher characteristics. *School Mental Health*, 7(1), 61–73. <https://doi.org/10.1007/s12310-015-9145-0>
- PASS 2023 Power Analysis and Sample Size Software. (2023). NCSS, LLC. Kaysville, Utah, USA, [ncss.com/software/pass](https://www.ncss.com/software/pass). Accessed 2 Feb 2023



- Petras, H., Kellam, S. G., Brown, C. H., Muthén, B. O., Ialongo, N. S., & Poduska, J. M. (2008). Developmental epidemiological courses leading to antisocial personality disorder and violent and criminal behavior: Effects by young adulthood of a universal preventive intervention in first- and second-grade classrooms. *Drug and Alcohol Dependence*, *95*(SUPPL. 1), S45–S59. <https://doi.org/10.1016/j.drugalcdep.2007.10.015>
- Poduska, Jeanne M., Kellam, S. G., Wei, W., Brownb, C. H., Ialongo, N. S., Peter, T., & Oyinbo. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*, *95*, S29–S44. <https://doi.org/10.1016/j.drugalcdep.2007.10.009>
- Rhoads, C. H. (2012). *Problems with tests of the missingness mechanism in quantitative policy studies. Statistics, Politics, and Policy*, *3*(1). <https://doi.org/10.1515/2151-7509.1012>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons. ISBN: 9780471087052 Online ISBN: 9780470316696 DOI: 10.1002/9780470316696.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Saigh, P. A., & Umar, A. M. (1983). The effects of a good behavior game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16*(3), 339–344. <https://doi.org/10.1901/jaba.1983.16-339>
- Schneider, D. R., Strelow, M., Langaro, F., & Kaszubowski, E. (2018). A inserção escolar do programa elos: Avaliação do processo de implementação na realidade brasileira. In Z. V. D. M. Sanchez, D. R. Schneider, R. T. Pedroso, A. Sanudo, A. G. Aveiro, A. G. Vieira (Eds.), *Prevenção ao uso de drogas: implantação e avaliação de programas no Brasil* (pp. 181–202). Ministério da Saúde.
- Schneider, D. R., Kaszubowski, E., Garcia, D., Scaf, 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). <https://doi.org/10.14718/acp.2020.23.1.15>
- Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine*, *13*(17), 1715–1726. <https://doi.org/10.1002/sim.4780131703>
- Smith, S., Barajas, K., Ellis, B., Moore, C., McCauley, S., & Reichow, B. (2021). A meta-analytic review of randomized controlled trials of the Good Behavior Game. *Behavior Modification*, *45*(4), 641–666. <https://doi.org/10.1177/0145445519878670>
- Storr, C. L., Ialongo, N. S., Kellam, S. G., & Anthony, J. C. (2002). A randomized controlled trial of two primary school intervention strategies to prevent early onset tobacco smoking. *Drug and Alcohol Dependence*, *66*(1), 51–60. [https://doi.org/10.1016/S0376-8716\(01\)00184-3](https://doi.org/10.1016/S0376-8716(01)00184-3)
- Streimann, K., Selart, A., & Trummal, A. (2020). Effectiveness of a universal, classroom-based preventive intervention (PAX GBG) in Estonia: A cluster-randomized controlled trial. *Prevention Science*, *21*(2), 234–244.
- Streimann, K., Trummal, A., Klandorf, K., Akkermann, K., Sisask, M., Toros, K., & Selart, A. (2017). Effectiveness of a universal classroom-based preventive intervention (PAX GBG): A research protocol for a matched-pair cluster-randomized controlled trial. *Contemporary Clinical Trials Communications*, *8*, 75–84. <https://doi.org/10.1016/j.conctc.2017.08.013>
- 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>
- Van Lier, P. A. C., Muthén, B. O., Van Der Sar, R. M., & Crijnen, A. A. M. (2004). Preventing disruptive behavior in elementary schoolchildren: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology*, *72*(3), 467–478. <https://doi.org/10.1037/0022-006X.72.3.467>
- Veenman, B., Luman, M., & Oosterlaan, J. (2018). Efficacy of behavioral classroom programs in primary school. A meta-analysis focusing on randomized controlled trials. *PLoS ONE*, *13*(10), e0201779. <https://doi.org/10.1371/journal.pone.0201779>
- Weis, R., Osborne, K. J., & Dean, E. L. (2015). Effectiveness of a universal, interdependent group contingency program on children's academic achievement: A countywide evaluation. *Journal of Applied School Psychology*, *31*(3), 199–218. <https://doi.org/10.1080/15377903.2015.1025322>
- Wilcox, H. C., Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., & Anthony, J. C. (2008). The impact of two universal randomized first- and second-grade classroom interventions on young adult suicide ideation and attempts. *Drug and Alcohol Dependence*, *95*(SUPPL. 1), 60–73.
- Witvliet, M., van Lier, P. A. C., Cuijpers, P., & Koot, H. M. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting and Clinical Psychology*, *77*(5), 905–915. <https://doi.org/10.1037/a0014597>