Psychometric validation of the audio-guided rBVQ instrument for bullying evaluation among students

Valdemir Ferreira-Junior[1]
Fausto Coutinho-Lourenco[1]
Alessandra A. S. Menezes[1]
Julia Dell Gusmoes[1]
Hugo Cogo-Moreira[1, 2]
Zila M. Sanchez[1]

[1]*Universidade Federal de São Paulo, São Paulo, São Paulo, Brasil*
[2]*Østfold University College, Halden, Norway*

Abstract
The revised Olweus Bully/Victim Questionnaire (rBVQ) is an internationally used instrument to identify bullying, but due to Brazilian students' low proficiency in reading and writing, it had to be adapted for audio-guided use through smartphones. We investigated construct validity, using Confirmatory Factor Analysis (CFA), under a two-correlated factor solution. Participated 1,742 fifth graders and 2,316 seventh graders from 30 public schools in the city of São Paulo. We found excellent fit indices models for both grades (fifth grade: CFI=0.985, RMSEA=0.020; seventh grade: CFI=0.990, RMSEA=0.015), as well as factor loading higher than 0.4 for all items, indicating their relevance to the construct. The rBVQ demonstrated excellent psychometric properties and may help future studies that aim to investigate bullying in populations with similar profiles. In addition, this study innovates by evaluating an audio-guided instrument and using mobile technology.
Keywords: School violence, test validity, victimization, conflict, aggressive behavior

Validação Psicométrica do Instrumento Áudio-Guiado rBVQ para Avaliação de *Bullying*

Resumo
O Questionário revisado de *bullying* de Olweus (rBVQ) é um instrumento internacionalmente utilizado para identificar *bullying*, mas devido às dificuldades de estudantes brasileiros na leitura e escrita, foi adaptado para uso áudio-guiado em *smartphones*. Investigou-se a validade do construto por meio de análise confirmatória fatorial em uma solução de dois fatores. Participaram 1742 e 2316 alunos de quintos e sétimos anos de trinta escolas públicas de São Paulo. Encontrou-se excelentes índices de adequação para ambos os anos escolares (quintos: CFI = 0,985; RMSEA = 0,020; sétimos: CFI = 0,990; RMSEA = 0,015) e cargas fatoriais maiores que 0,4 para todos os itens, indicando sua relevância para o construto. O rBVQ demonstrou excelentes propriedades psicométricas e pode ajudar estudos futuros que objetivam investigar *bullying* em populações com perfis semelhantes. Além disso, este estudo inova por avaliar um instrumento áudio-guiado e utilizando tecnologia móvel.

Palavras-chave: violência escolar, validade do teste, vitimização, conflito, comportamento agressivo

Validación psicométrica del instrumento audio guiado rBVQ para la evaluación del bullying

Resumen
El Cuestionario Revisado de Bullying de Olweus (rBVQ) es un instrumento utilizado internacionalmente para identificar bullying, pero debido a las dificultades de estudiantes brasileños en la lectura y la escritura, se ha adaptado su uso por medio de audioguías en smartphones. Se investigó la validez del constructo mediante el Análisis Factorial Confirmatorio, bajo un modelo de dos factores. Participaron 1742 y 2316 estudiantes de quinto y séptimo curso de 30 escuelas públicas de São Paulo. Excelentes tasas de adecuación fueron encontradas para ambos casos (quinto año: CFI=0,985, RMSEA=0,020; séptimo año: CFI=0,990, RMSEA=0,015), además de cargas factoriales superiores a 0,4 para todos los ítems, lo que indica su relevancia para el constructo. El instrumento rBVQ ha presentado excelentes propiedades psicométricas y puede ayudar a futuros estudios pretendan investigar el bullying en poblaciones con perfiles similares. Además, este estudio innova al evaluar un instrumento guiado por audio, empleando la tecnología móvil.
Palabras clave: Violencia escolar, validez de la prueba, victimización, conflicto, comportamiento agresivo

Introduction

The school environment constitutes more than a space for knowledge and the exercise of ethics and reason, it has also been characterized as a place of proliferation of violence and depredation (Giordani et al., 2017). Among the types of school violence, bullying has gained interest in the media and academic research due to its negative consequences. For (Olweus, 1994), bullying differs from other forms of violence because

http://dx.doi.org/10.1590/1413-82712022270214

it occurs repeatedly, among peers and with power asymmetry. Studies have shown the impacts of bullying on the mental health of those involved in it (Arseneault, 2018; Moore et al., 2017) and, although both victims and perpetrators suffer from the results of this violence, victims are at greater risk of developing mental disorders such as depression and anxiety (Strohacker et al., 2019), social phobia (Wu et al., 2016), and eating disorders (Lie et al., 2019), as well as other more severe disorders such as self-harm behavior, psychotic symptoms and suicidal ideation (Strohacker et al. 2019).

Bullying prevalence varies according to country, age, gender and method used in its estimation. Biswas et al. (2020) investigated bullying victimization among school children (aged 12-17 years) across 83 countries, and found a pooled prevalence of 31%. The highest prevalence was observed in the Eastern Mediterranean (45.1%) and African (43.5%) regions, and the lowest in Europe (8.4%). According Cook et al. (2010), the estimated prevalence of bullying victimization among children and adolescents aged 3 to 18 years old was 20.9% in South Africa, 31.5% in Canada, 21.5% in the United States, 16.8% in Finland, 15.5% in France, 23.9% in England, 10.4% in Japan, 42.5% in New Zealand and 6.7% in Switzerland. In Brazil, Malta et al. (2019) found that 7.4% of 9th grade students had been bullied during the month prior to data collection.

Considered a public health issue, in 2015 Brazilian legislation recognized bullying as described by Olweus, and instituted the bullying prevention program "Systematic Bullying Combat Program" (Federal Brazilian Law 13.185), which, among other actions, requires that educational establishments promote bullying awareness and prevention measures. Thus, instruments capable of detecting bullying instances are necessary, not only in screening situations but also in evaluating the efficacy or effectiveness of school interventions. Self-report questionnaires are the most used instruments for measuring bullying and evaluating the success of bullying prevention programs in epidemiological studies of behaviors (Gonçalves et al., 2016).

Although on the rise, internal assessments by the Brazilian Ministry of Education show inequalities in reading proficiency between public and private schools, as well as other socioeconomic factors, including ethnic and racial issues (INEP, 2018, 2019; Menezes et al., 2016). In international evaluations such as the Program for International Student Assessment – PISA, which assesses how prepared young people are to act as citizens in contemporary society, Brazil performs below the average of the 35 countries that make up the Organization for Economic Cooperation and Development (OECD, 2019). Therefore, an instrument validated in Portuguese for use in Brazilian schools should have properties that allow a good performance in these different scenarios, since most of the students are not able to read at the age of 10 (Lima & Ciasca, 2018). Given the difficulties in locating and accessing such instruments, it became another issue for researchers working with Brazilian students.

Widely used due to its interpretability and reproducibility (Solberg & Olweus, 2003) and its well-defined psychometric properties in different countries (Guilheri et al., 2015; Kyriakides et al., 2006), the revised Olweus Bully/Victim questionnaire – rBVQ (Solberg & Olweus, 2003) emerges as an option in this context, as it has already been validated in Brazilian Portuguese by Guilheri (2016) .

The questionnaire presents a detailed description of bullying followed by two blocks of questions, one about victimization (being bullied) and another about perpetration (bullying others). Each block has a global question about frequency of the episodes and seven other specific questions about types of bullying suffered or practiced (from verbal insults to physical aggression and racist motivations). The questionnaire addresses several other variables found to be related to the global ones: social disintegration, negative self-esteem, and depressive tendencies is related to being bullied, while aggression and antisocial behavior is related to bullying others (Solberg & Olweus, 2003). Respondents who indicate a frequency of two or more times per month in any of the situations, depending on the group of questions, are considered victims, perpetrators or perpetrators-victims (Solberg & Olweus, 2003), the latter being those who concomitantly practice and suffer bullying. Respondents who indicate frequencies below two times a month, for all situations, are considered neutral.

Kyriakides et al. (2006) tested validity and reliability of rBVQ using Rasch analysis, and found satisfactory psychometric properties and a high correlation among the factors "being victimized" and "bullying others". In a cross-cultural study between Brazil and France, Guilheri (2016) carried out the validation for the French and Brazilian versions of the rBVQ and found that both have one-dimensional items with factor loads above 0.32; according to Kaiser's criterion, Bullying is constituted by two factors ("Victim" and "Aggressor") with a moderate correlation between them. Regarding the invariancy test, both versions showed no significant

difference between genders, that is, for both the French and the Brazilian versions, boys and girls understand and respond to the instrument similarly.

Although rBVQ has been validated in Brazil by Guilheri (2016), the sample consisted of non-socioeconomically vulnerable students, data were collected by a paper-and-pencil questionnaire, and was not published in a peer reviewed journal. As such, we still need to provide evidence of validity based on the internal structure of an audio-guided rBVQ for bullying victimization and perpetration assessment among students of lower socioeconomic status in public schools. The aim of this study was to investigate construct validity and test invariance, using Confirmatory Factor Analysis, of the two-correlated factor solution (bullying victimization and perpetration) for the audio-guided rBVQ scale across fifth and seventh graders in public schools in Brazil. It is also important to highlight that the instrument validated in this study will be used to evaluate the effects of the Keepin' it REAL (KiR) program in preventing the bullying perpetration and victimization. Although several studies have evaluated the effects of KiR on substance use (Kulis et al. 2019) and the need of a multicultural adaptation (Hecht et al., 2018), none of the existing studies have evaluated the effects of the program on bullying.

Methods

This study used the baseline assessment (i.e., pre-intervention) of a cluster randomized controlled trial (RCT) that evaluated a school-based drug and violence prevention program, named PROERD (Educational Program of Drug and Violence Resistance), among fifth and seventh graders of 30 public schools in São Paulo.

Participants

In total, 1,742 fifth graders [51.21% male, average age = 10.12 years old, standard deviation (SD) = 0.65], and 2,316 seventh graders [51.49% male, average age = 12.27, SD = 0.72], from 30 public schools of São Paulo, Brazil, participated in the study. The schools were randomly selected from those that did not receive PROERD in the last 3 years. A second allocation determined whether each school would be assigned to the control or intervention group according to a random list. All fifth and seventh graders of each of the selected schools contributed to the study, and all students present participated on data collection. All schools were located in regions far from the urban centers and considered to be of low economic and social development. Due to scale misunderstanding, socio-economic status (SES) was assessed only in the seventh grade [39.65% A-B, 53.61% C, and 6.74% D-E]. Students with intellectual disability or cognitive delay were appointed by the school and received a special code to exclusion in this study.

Since the PROERD curricula is designed for different grades and two different age ranges were targeted (fifth and seventh grade), two different sample sizes were calculated to evaluate the effect of PROERD.

For the fifth grade, the required sample size calculated was 1,820 participants (70 per group) for a power of 80%, a significance level of 5%, 0.3 effect size and 0.02 interclass correlation (Ahn et al., 2014). Considering a 25% participant loss, the sample needed to include 2,340 participants. The parameters used were based on the bullying instrument (Guilheri, 2016) and the results of a randomized controlled trial (Sanchez et al., 2017).

Regarding the seventh grade, based on Donner & Klar (2010), the minimum required sample size calculated was 1,608 participants (67 per group) for a power of 80%, a significance level of 5%, difference of proportions of 7% and 0.02 interclass correlation. Also considering a 25% participant loss, the sample needed to include 2,160 participants. The parameters used were based on the results of a study by the KiR USA (Marsiglia et al., 2011).

Instrument

An original translated version of rBVQ (Guilheri, 2016) was applied in our pilot study, which revealed some difficulties and problems during data collection among fifth graders: their lack of proficiency in reading hindered understanding the different categories of answers per item. rBVQ has seven questions to identify different bullying situations for each type of bullying (victimization or perpetration). The response alternatives for all questions are: "I haven't bullied/been bullied…," "only once or twice," "2 or 3 times a month," "about once a week," and "several times a week". A student is considered as a victim or a perpetrator if they answer "2 or 3 times a month" or more. For the 7th grade, we used the original translated version (Guilheri, 2016) adopting the cut point suggested by the authors (Solberg & Olweus, 2003). For the 5th grade a structurally adapted version was necessary, we simplified the frequency, adapting the number of categories of answers per questions by replacing the 5-point scale by binary answers ("yes" or

"no") and included for each specific bullying situation a new question about whether, in addition to happening recently, it also happened in the last year, to ensure that the event was recurrent. Fifth graders were considered bullying victims or perpetrators if they answered any question on each category (victim or perpetrator) for both "recently" and "last year" positively (yes). In order to promote accessibility and enable the participation of students who had reading difficulties, the audio-guided format was adopted for both grades.

Data collection

Data were collected during baseline assessment (data collection prior to the application of the intervention), using an anonymous self-reported audio-guided questionnaire completed by the participants on smartphones, applied by researchers in the classroom, without the teacher present. Using smartphones made participation more enticing and allowed using audios and images to facilitate understanding, making the participation of students with low proficiency in reading and writing possible – a very prevalent problem in Brazilian public schools (OECD, 2019). Through smartphones we can also configure the data collection software to skip questions according to student responses and send alerts of incorrect answers. The collected data was sent directly to a cloud database, eliminating having to manually transcribe responses, and so, avoiding transcription errors.

There were two researchers per classroom, and before answering the questionnaire, the students received ethical information about the research and signed a consent form. Students had 50 minutes to answer the instrument individually and anonymously.

Data analysis

Confirmatory Factor Analysis (CFA) was conducted to determine construct validity of the two-correlated factor solution for the rBVQ scale across both samples. One model was tested for the fifth grade and another for the seventh grade.

Invariance testing for gender was performed using the Multigroup Confirmatory Factor Analysis (MGCFA) approach. Both models were estimated with 8.4 (Muthén & Muthén, 2012) using a weighted least squares with mean and variance adjusted estimator (WLSMV). Due to its multilevel design (children nested in schools), Asparouhov's (2005, 2006) complex strategy to deal with the non-independence was adopted.

After testing the initial model (i.e., two-correlated factor solution), we performed a sequential strategy for testing measurement invariance, following Meredith's (1993) recommendations. This procedure determines if the meaning of the construct and the difficulty of each individual item was equivalent across genders. These criteria for configural and scalar invariance must be met to compare the groups on the latent variable. As MGCFA involves two separate input matrices, constraining factor loadings and threshold parameters in both groups, it is likely to obtain bivariate empty cells.

A common issue when variables have low prevalence rates, empty cells generate statistically perfect correlations between two items, meaning that they are statistically indistinguishable, and, for analysis purpose, one should be removed. Thus, where empty cells appeared, we excluded one of the items to maintain the maximum possible number of original items for both victim and aggressor factors.

First, we tested if the factor structure was similar between groups (i.e., configural invariance). For model fit and adjustment index, we adopted the chi-squared statistic ($\chi 2$), the 90% CI of the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis Index (TLI). To interpret these indices, we followed Hu and Bentler's (2009) recommendations: an adequate model fit is indicated by $p > 0.05$ for the $\chi 2$, RMSEA ≤ 0.06, CFI ≥ 0.95, and TLI ≥ 0.95. Then, we tested if the item thresholds and factor loadings were equivalent between groups (i.e., scalar invariance).

To provide evidence of scalar versus configural invariance, we systematically tested whether imposing restrictions (i.e., if the discrimination and difficulty parameters are equal across genders) did not worsen the model compared with the less constrained model wherein the parameter was freely estimated (Chen, 2007). The main focus of the invariance testing was to achieve scalar level (strong invariance), as it is a prerequisite for mean comparison between groups (Vandenberg & Lance, 2000). We also provide evidences regarding strict invariance, where restrictions are imposed to residual levels, being therefore, the next hierarchical restriction after scalar invariance. As $\chi 2$ statistic is highly sensitive to sample size, we considered that the added restrictions did not worsen the model if ΔCFIs of the free and constrained models differed by less than 0.01 (Cheung & Rensvold, 2002), and if the change in RMSEA between models was less than 0.015.

Lastly, we calculated reliabilities for the two latent factors within each grade (and their confidences intervals estimated via 200-bootstrapped interactions [Kelley

and Pornprasertmanit, 2016]) using Green and Yang's (2009) proposed index.

Ethical procedures

The study protocol was approved by the Universidade Federal de São Paulo's Research Ethics Committee (n:1327/2018), and the RCT was registered at the Brazilian Ministry of Health Register of Clinical Trials (REBEC), under protocol no. 6q23nk. Before data collection, the school principal received an informed consent form, consonant with the research, and participants also consented to participate by signing the assent form.

Results

The models fit indices showed excellent performance for both grades (Table 1); for both grades, the factor loading for all items were higher than 0.4, indicating their relevance to the construct (Table 2).

As the configural model returned good fit index, we investigated scalar invariance by holding the factor loadings and thresholds of items equal between the groups. Both samples had one item removed due to low prevalence causing empty cells (namely, *"I threatened or forced someone to do things that he or she didn't want to do"* in fifth grade, and *"I got money or other things or I purposefully damaged his or her belongings",* in seventh grade) (Table 3).

The scalar versus configural invariance via chi-square difference test showed that only the seventh grade sample remained constant after the restriction: bullying [fifth graders: $\chi^2_{(diff)}(9) = 40.359$, $p = <0.0001$; seventh graders: $\chi^2_{(diff)}(9) = 9.518$, $p = 0.3909$]. Under invariance testing using ΔCFI and ΔRMSEA, however, we achieved scalar invariance for both samples [fifth graders: ΔCFI = 0.007, ΔRMSEA = 0.003; seventh graders: ΔCFI = 0.001, ΔRMSEA = 0.001]. As the minimum invariance level required to directly compare the means between groups was achieved by these latter criteria, and the chi-square testing difference is sensitive to large sampling, the mean in the latent traits could be compared in both models and samples between genders. Strict invariance testing returned the following model fit: for fifth grade, RMSEA = 0.025 and CFI = 0.976; for seventh grade, RMSEA = 0.015 and CFI = 0.991. Comparing the scalar models with the strict invariance, for both grades, the imposed restriction at residual levels did not worsen the models fit.

As for reliabilities, the fifth grade returned victim factor = 0.759 (95% bootstrapped confidence interval [95%BootCI] = 0.739 to 0.785) and perpetrator factor = 0.720 (95%BootCI = 0.670 to 0.784); the seventh grade returned victim factor = 0.763 (95%BootCI = 0.734 to 0.797) and perpetrator factor = 0.775 (95%BootCI = 0.712 to 0.845).

Table 1.
Fit indices of bullying scale for fifth and seventh graders

Model fit indices	5th grade	7th grade
$\chi 2$	129.622	114.057
df	76	76
p-value	0.0001	0.0031
RMSEA	0.020	0.015
90% RMSEA	[0.014; 0.026]	[0.009; 0.020]
CFI	0.985	0.990
TLI	0.981	0.988
SRMR	0.061	0.065

Legend: df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lew Index; SRMR= standardized root mean squared residual

Discussion

This study used CFA to investigate construct validity of the two-correlated factor solution for the audio-guided rBVQ scale across fifth and seventh graders in Brazilian public schools. For seventh graders, we used the original translated version of the instrument without the general questions; for fifth graders, we needed to fully adapt the questionnaire, as specified in the method section, due to the students' reading and comprehension difficulties. We found excellent fit indices for both grades, and all items were relevant for the construct, with factor loading higher than 0.4.

In Brazil, bullying discussion is still recent and lacks tools with well-defined evidences of psychometric properties to assess this construct, which is essential to estimate bullying prevalence and its associated factors. International questionnaires such as Peer Assessment (Rubin et al., 2007), Behavior in School-aged Children survey (HBSC) (Currie et al., 1997), Conduct Disorder Questions (Rutter et al., 1988) and Kidscape (Elliott & Kilpatrick, 1994) have been translated to Brazilian Portuguese, but studies have yet to describe their psychometric features and adequacy for our population, limiting research on bullying in the Brazilian context. In a study reviewing 25 Brazilian articles,

Table 2.
Standardized factor loadings of two-correlated factor solution of bullying scale for fifth and seventh graders

	5th grade			7th grade		
	Estimate	**S.E.**	**Est./S.E.**	**Estimate**	**S.E.**	**Est./S.E.**
Victim						
Item 1	0.772	0.019	40.357	0.773	0.022	34.797
Item 2	0.714	0.027	26.537	0.832	0.015	54.911
Item 3	0.793	0.027	29.899	0.877	0.025	34.459
Item 4	0.734	0.027	26.875	0.807	0.017	48.376
Item 5	0.728	0.026	28.120	0.719	0.031	22.964
Item 6	0.749	0.028	26.752	0.852	0.035	24.513
Item 7	0.624	0.040	15.772	0.809	0.025	32.334
Practice						
Item 8	0.822	0.029	28.539	0.804	0.025	31.806
Item 9	0.694	0.042	16.697	0.867	0.037	23.400
Item 10	0.722	0.037	19.745	0.908	0.033	27.729
Item 11	0.829	0.036	22.930	0.943	0.033	28.559
Item 12	0.758	0.043	17.649	0.949	0.039	24.404
Item 13	0.768	0.071	10.800	0.853	0.035	24.498
Item 14	0.717	0.060	12.016	0.863	0.042	20.668
Victim with Practice	0.636	0.035	18.271	0.594	0.056	10.631

p-value <0.05

Table 3.
Multigroup Confirmatory Factor Analysis (MGCFA) of bullying scale for fifth and seventh graders

		χ2(df)	**CFI**	**TLI**	**RMSEA**	**90% C.I.**
	Model fit for 5th graders (13 items)					
Bullying version	Configural	179.113 (128) *	0.985	0.981	0.022	[0.013; 0.029]
	Scalar	210.987 (137) *	0.978	0.974	0.025	[0.018; 0.032]
	Model fit for 7th graders (13 items)					
	Configural	163.587 (128) *	0.991	0.989	0.016	[0.007; 0.022]
	Scalar	170.315 (137) *	0.992	0.991	0.015	[0.005; 0.021]

Legend: CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; C.I = Confidence interval; * = *p-value <0.05*

Alckmin-carvalho & Izbicki (2014) concluded that the lack of validated instruments for bullying assessment is an important methodological limitation for studies in Brazil, with bullying being identified, in most studies, by questionnaires developed by the researchers or by instruments not yet validated for the Brazilian population. For instance, when performing the National Adolescent School-based Health Survey (PeNSE), Malta and colleagues (2014) assessed bullying by asking only three questions: "In the past 30 days, how

often did your school colleagues treat you well and/or were thoughtful to you?", "In the past 30 days, how often did any of your colleagues bring you down, made fun of you, mock you?", and "In the past 30 days, did you bring down, mock, intimate or tease any of your school colleagues, so that the person was hurt, annoyed, offended or humiliated?". Considering the complexity of bullying episodes, it is reasonable to question if only a few items can assess such violence. Thus, validate an instrument able to assess more layers of this behavior among Brazilian students is of great importance.

Guilheri's (2016) study of the construct validity of rBVQ in Brazil used a sample smaller than ours (N = 1,802), excluded socioeconomic vulnerable students and with low reading proficiency, and did not conduct invariance tests. In this study, using the instrument in its original form was impossible, requiring not only to adapt the questions, but also to include audios and images, as most students did not master reading and had difficulty with text interpretation. As described in *Sampling,* sample size was calculated considering the main hypothesis of the randomized clinical trial designed to evaluate the effects of PROERD in these two grades. For our psychometrics analysis, we obtained and analyzed different sample sizes (1,742 fifth graders and 2,316 seventh graders) with enough power (i.e., 1 – β) to detect very smalls factor loadings (i.e., below 0.4). These samples also have enough power to conduct invariance testing under multigroup confirmatory factor analysis, which *per se* requires a more substantial sample size when compared with other techniques as MIMIC (see Brown, 2015, p.242) or moderated nonlinear factor analysis (Bauer, 2017).

In Brazil, besides the quality of public education being lower than that of private schools, there is also a contrasting difference in the quality of education between central and peripheral public schools (Marques & Torres, 2019). Most students in our study were socially vulnerable, having limited financial resources and living in neighborhoods with precarious paving, housing, and access to opportunities for development as a citizen. Thus, the previous validation of the instrument for the Brazilian population may not contemplate the different realities within the country. In large cities such as São Paulo, where the instrument has been validated, we see regions with complete public infrastructure alongside areas lacking basic services, luxury buildings living alongside slums, and extremely high wages counteracting a huge mass of unskilled unemployed (Marques & Torres, 2019). Our work represents an important contribution to the literature, which lacks studies on instrument validation, especially with socially vulnerable populations, to which many school prevention programs are addressed.

In epidemiological studies involving data collection through questionnaires, the pen and paper method has been the standard procedure for years (S. Zhang et al., 2012). In the last two decades, electronic devices such as personal assistants (PDAs), and more recently mobile phones, have emerged as an alternative to data collection (Zhang et al., 2017). Using smartphones is a recent development, taking advantage of its larger screens and direct wireless data upload, eliminating having to manually insert data in the database, which avoids transcription errors (Zhang et al., 2017; Zhang et al., 2012). Through smartphones, we can configure the software to skip questions according to participant responses and be aware of typing errors or wrong answers, which improves data quality. In addition, it allows including audios and interactive images that can facilitate the participant's understanding and enable the participation of people with disabilities, as performed in our study.

Collecting data electronically not only reduces potential biases during data collection and transcription, but also allows researchers to obtain more reliable responses from participants (Zhang et al., 2012). In our study, besides collecting data anonymously through smartphones, we also used an audio-guided questionnaire with images to reduce possible misinterpretation. With the increasing use of technologies in our daily lives, modernization of data collection becomes indispensable, which besides facilitating and increasing its quality (Zhang et al., 2012), makes participation more attractive.

Validate rBVQ becomes extremely important as the instrument will be used to evaluate the effects of KiR on bullying but can be used for the evaluation of any other program intended to reduce bullying or, even, to identify the distribution of bullying reports in early adolescence. Our data, besides allowing us to evaluate the program in Brazil and the use of the instrument in future studies on bullying, may contribute to the better understanding of bullying among deprived students using a valid instrument.

A limitation of this study could be the sample, as previously stated, we used the data of a RCT, with schools randomly selected from those that did not receive the evaluated prevention program in the last 3 years, and all these schools were located in peripheral

regions, considered to be of low economic and social development, by random selection. But validation studies with this population are important, since many school prevention programs are addressed for this kind of reality. Therefore, it is important that future studies that aim to assess bullying using this instrument consider performing the psychometric validation of the instrument, especially when the profile of the target population is different.

References

Ahn, C., Heo, M., & Zhang, S. (2014). *Sample size calculations for clustered and longitudinal outcomes in clinical research*. CRC Press.

Alckmin-carvalho, F., & Izbicki, S. (2014). Estratégias e instrumentospara a identificação de bullyingem estudos nacionais. *Avaliação Psicológica*, *13*(3), 343–350.

Arseneault, L. (2018). Annual Research Review: The persistent and pervasive impact of being bullied in childhood and adolescence: implications for policy and practice. *Journal of Child Psychology and Psychiatry*, *59*(4), 405–421. https://doi.org/10.1111/jcpp.12841

Asparouhov, T. (2005). Sampling Weights in Latent Variable Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 411–434. https://doi.org/10.1207/s15328007sem1203_4

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

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bentler, P. M. (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika*, *74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Biswas, T., Scott, J. G., Munir, K., Thomas, H. J., Huda, M. M., Hasan, M. M., de Vries, T. D., Baxter, J., & Mamun, A. A. (2020). Global variation in the prevalence of bullying victimisation amongst adolescents: Role of peer and parental supports. *EClinicalMedicine*, 100276.

Brown, T. (2015). Confirmatory Factor Analysis for Applied Research, Second Edition. In *Guilford Publications*.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cook, C. R., Williams, K. R., Guerra, N. G., & Kim, T. E. (2010). Variability in the prevalence of bullying and victimization: A cross-national and methodological analysis. In *Handbook of bullying in schools: An international perspective*.

Currie, C. E., Elton, R. A., Todd, J., & Platt, S. (1997). Indicators of socioeconomic status for adolescents: the WHO Health Behaviour in School-aged Children Survey. *Health Education Research*, *12*(3), 385–397. https://doi.org/10.1093/her/12.3.385

Donner, A., & Klar, N. (2010). *Design and Analysis of Cluster Randomization Trials in Health Research* (1st ed.). Wiley.

Elliott, M., & Kilpatrick, J. (1994). *How to Stop Bullying: A Kidscape Training Guide*. Kidscape.

Giordani, J. P., Seffner, F., & Dell'Aglio, D. D. (2017). Violência escolar: percepções de alunos e professores de uma escola pública. *Psicologia Escolar e Educacional*, *21*(1), 103–111.

Gonçalves, F. G., Heldt, E., Peixoto, B. N., Rodrigues, G. A., Filipetto, M., & Guimarães, L. S. P. (2016). Construct validity and reliability of Olweus bully/victim questionnaire–Brazilian version. *Psicologia: Reflexão e Crítica*, *29*(1), 27.

Green, S. B., & Yang, Y. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Guilheri, J. (2016). *Jogos de asfixia, jogos de agressão, bullying em contexto escolarEstudo transcultural França-Brasil com escolares de 9-12 anos*. Universidade Federal de São Paulo.

Guilheri, J., Cogo-Moreira, H., Kubiszewski, V., Yazigi, L., & Andronikof, A. (2015). Validité de construit du questionnaire rBVQ d'Olweus pour l'évaluation du harcèlement scolaire (bullying) auprès d'élèves français de cycle 3. *Neuropsychiatrie de l'Enfance et de l'Adolescence*, *63*(4), 211–217. https://doi.org/10.1016/j.neurenf.2015.01.010

Hecht, M. L., Shin, Y., Pettigrew, J., Miller-Day, M., & Krieger, J. L. (2018). Designed Cultural Adaptation and Delivery Quality in Rural Substance Use Prevention: an Effectiveness Trial for the Keepin' it REAL Curriculum. *Prevention Science*, *19*(8), 1008–1018. https://doi.org/10.1007/s11121-018-0937-y

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2018). *Relatório SAEB (ANEB e ANRESC) 2005-2015: panorama da década*.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2019). *Relatório Saeb 2017*. http://portal.inep.gov.br/documents/186968/484421/RELATÓRIO+SAEB+2017/fef63936-8002-43b6-b741-4ac9ff39338f?version=1.0

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, *21*(1), 69–92. https://doi.org/10.1037/a0040086

Kulis, S. S., Marsiglia, F. F., Porta, M., Avalos, M. R. A., & Ayers, S. L. (2019). Testing the Keepin'it REAL substance use prevention curriculum among early adolescents in Guatemala city. *Prevention Science*, *20*(4), 532–543.

Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *British Journal of Educational Psychology*, *76*(4), 781–801. https://doi.org/10.1348/000709905X53499

Lei Nº 13.185, de 6 de novembro de 2015. Institui o Programa de Combate à Intimidação Sistemática (Bullying), Pub. L. No. Diário Oficial da União, 152(213) (2015). http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13185.htm

Lie, S. Ø., Rø, Ø., & Bang, L. (2019). Is bullying and teasing associated with eating disorders? A systematic review and meta-analysis. *International Journal of Eating Disorders*, *52*(5), 497–514. https://doi.org/10.1002/eat.23035

Lima, N., & Ciasca, M. I. (2018). ANA AND PROOF BRAZIL (5TH YEAR): WHAT DOES THE RESULTS REVEAL IN PORTUGUESE LANGUAGE. *Educação & Linguagem*, 43–60.

Malta, D. C., Mello, F. C. M. de, Prado, R. R. do, Sá, A. C. M. G. N. de, Marinho, F., Pinto, I. V., Silva, M. M. A. da, & Silva, M. A. I. (2019). Prevalência de bullying e fatores associados em escolares brasileiros, 2015. *Ciência & Saúde Coletiva*, *24*(4), 1359–1368. https://doi.org/10.1590/1413-81232018244.15492017

Malta, D. C., Porto, D. L., Crespo, C. D., Silva, M. M. A., Andrade, S. S. C. de, Mello, F. C. M. de, Monteiro, R., & Silva, M. A. I. (2014). Bullying em escolares brasileiros: análise da Pesquisa Nacional de Saúde do Escolar (PeNSE 2012). *Revista Brasileira de Epidemiologia*, *17*, 92–105.

Marques, E., & Torres, H. (2019). *São Paulo: segregação, pobreza e desigualdades sociais. Editora Senac São Paulo.* Editora Senac São Paulo.

Marsiglia, F. F., Kulis, S., Yabiku, S. T., Nieri, T. A., & Coleman, E. (2011). When to Intervene: Elementary School, Middle School or Both? Effects of keepin' It REAL on Substance Use Trajectories of Mexican Heritage Youth. *Prevention Science*, *12*(1), 48–62. https://doi.org/10.1007/s11121-010-0189-y

Menezes, I. G., Duran, V. R., Mendonça Filho, E. J., Veloso, T. J., Sarmento, S. M. S., Paget, C. L., & Ruggeri, K. (2016). Policy Implications of Achievement Testing Using Multilevel Models: The Case of Brazilian Elementary Schools. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01727

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Moore, S. E., Norman, R. E., Suetani, S., Thomas, H. J., Sly, P. D., & Scott, J. G. (2017). Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis. *World Journal of Psychiatry*, *7*(1), 60. https://doi.org/10.5498/wjp.v7.i1.60

Muthén, L., & Muthén, B. (2012). Mplus. Seventh edition. In *Los Angeles, CA: Muthén & Muthén*.

OECD. (2019). *Programme for International Student Assessment (PISA) - Results from PISA 2018*.

Olweus, D. (1994). *Bullying at School* (pp. 97–130). https://doi.org/10.1007/978-1-4757-9116-7_5

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2007). Peer Interactions, Relationships, and Groups. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0310

Rutter, M., Olweus, D., Block, J., & Radke-Yarrow, M. (1988). Development of Antisocial and Prosocial Behavior: Research, Theories, and Issues. *Contemporary Sociology*, *17*(2), 252. https://doi.org/10.2307/2070637

Sanchez, Z. M., Valente, J. Y., Sanudo, A., Pereira, A. P. D., Cruz, J. I., Schneider, D., & Andreoni, S. (2017). The #Tamojunto Drug Prevention Program in Brazilian Schools: a Randomized Controlled Trial. *Prevention Science*, *18*(7), 772–782. https://doi.org/10.1007/s11121-017-0770-8

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior*, *29*(3), 239–268. https://doi.org/10.1002/ab.10047

Strohacker, E., Wright, L. E., & Watts, S. J. (2019). Gender, Bullying Victimization, Depressive Symptoms, and Suicidality. *International Journal of Offender Therapy and Comparative Criminology*, 0306624X1989596. https://doi.org/10.1177/0306624X19895964

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Wu, W.-C., Luu, S., & Luh, D.-L. (2016). Defending behaviors, bullying roles, and their associations with mental health in junior high school students: a population-based study. *BMC Public Health*, *16*(1), 1066. https://doi.org/10.1186/s12889-016-3721-6

Zhang, J., Sun, L., Liu, Y., Wang, H., Sun, N., & Zhang, P. (2017). Mobile device–based electronic data capture system used in a clinical randomized controlled trial: Advantages and challenges. *Journal of Medical Internet Research*, *19*(3), e66.

Zhang, S., Wu, Q., van Velthoven, M. H., Chen, L., Car, J., Rudan, I., Zhang, Y., Li, Y., & Scherpbier, R. W. (2012). Smartphone Versus Pen-and-Paper Data Collection of Infant Feeding Practices in Rural China. *Journal of Medical Internet Research*, *14*(5), e119. https://doi.org/10.2196/jmir.2183

Recebido em: 25/08/2020
Reformulado em: 08/01/2021
Aprovado em: 01/03/2021

Annex

Revised Olweus Bully/Victim Questionnaire (rBVQ)

Item
Victim
i1 = I was called mean names, was made fun of, or teased in a hurtful way.
i2 = Other students left me out of things on purpose, excluded me from their group of friends, or completely ignored me.
i3 = I was hit, kicked, pushed, shoved around, or locked indoors.
i4 = Other students told lies or spread rumors about me and tried to make others dislike me.
i5 = I had money or things taken away from me or damaged.
i6 = I was threatened or forced to do things I did not want to do.
i7 = I was bullied with mean names or comments about my race or color.
Practice
i8= I called another student(s) mean names, made fun of, or teased him/her in a hurtful way.
i9 = I kept him/her out of things on purpose, excluded him or her from my group of friends, or completely ignored him or her.
i10 = I hit, kicked, pushed, and shoved him/her around or locked him or her indoors.
i11 = I spread false rumors about him/her and tried to make others dislike him/her.
i12 = I took money or things from him/her or damaged his/her belongings.
i13 = I threatened or forced him/her to do things he/she did not want to do.
i14 = I bullied him/her with mean names or comments about his/her race or color.

Acknowledgments:

The authors Hugo Cogo-Moreira and Zila M. Sanchez are sharing senior authorship.

About the authors:

Valdemir Ferreira-Junior, BSc, is a MSc student in the Department of Psychiatry at Universidade Federal de São Paulo. His research focuses on school violence and substance use prevention. He also has research experience in the field of education, especially in science teaching.
ORCID: https://orcid.org/0000-0003-0148-4281
E-mail: valdemir.ferreira@unifesp.br

Fausto Coutinho-Lourenco is a Psychologist and a PhD student in the Department of Psychobiology at Universidade Federal de São Paulo.
ORCID: https://orcid.org/0000-0002-5656-5137
E-mail: fausto.lourenco@usp.br

Alessandra A. S. Menezes, MSc., is a Mathematician, an Educational Technician, and a Ph.D. student in the Department of Preventive Medicine at Universidade Federal de São Paulo.
ORCID: 0000-0002-3102-4748
E-mail: aasmenezes@unifesp.br

Julia D Gusmoes received her Master in Science in 2018 (Universidade Federal de São Paulo/Escola Paulista de Medicina) and is, currently, a phD student in the same university.
ORCID: 0000-0002-1228-9791
E-mail: dell.julia@unifesp.br

Dr. Hugo Cogo-Moreira received his PhD in Science in 2012 (Universidade Federal de São Paulo/Escola Paulista de Medicina) and currently he is a professor at Østfold University College, Department of Education, ICT and Learning.
ORCID: https://orcid.org/0000-0001-9411-9237
E-mail: hugo.c.moreira@hiof.no

Dr. Zila M. Sanchez, PhD, is an Associate Professor in the Department of Preventive Medicine, Universidade Federal de São Paulo. She has authored more than 80 peer- reviewed articles, served as PI or MPI of multiple funded grants. She is an associate editor of PLOS One and BMC Public Health. Her research focuses mainly on substance use prevention.
ORCID: https://orcid.org/0000-0002-7427-7956
E-mail: zila.sanchez@unifesp.br

Contact:

Rua Sena Madureira, 1.500, 1º Andar
São Paulo-SP, Brasil
CEP: 04021-001