Psicologia: Reflexão e Crítica

RESEARCH

CrossMark

Development and refinement of the Peer Aggressive Behavior Scale–PAB-S

Juliane Callegaro Borsa

Abstract

Child aggressive behavior is a growing research topic in several fields of knowledge. Nonetheless, there are few instruments in Brazil designed to assess such a construct. The aim of this study was to present a new instrument to assess aggressive behavior among children: The Peer Aggressive Behavior Scale (PAB-S). Two studies are presented. The first describes the items development procedures, as well as evidence of content validity. It also presents the procedures used to refine the instrument and initial evidence of validity based on the internal structure of the refined version of the PAB-S. The second study evaluates, in an independent sample, the goodness-of-fit of the refined version and compares them with those presented by the initial version. A sample of 974 children (52,3 % girls) aged between 7 and 13 years old, attending public and private schools from Porto Alegre and Rio de Janeiro, participated in the study. The results indicate that the PAB-S fits a one-dimensional instrument in both the initial (39 items) and the refined version (25 items). The results suggest evidence of content validity and validity based on the internal structure of the scale. Further studies are needed to gather evidence of validity based on external variables.

Keywords: Aggressive behavior, Children, Scale, Psychometrics

Background

Aggressive behavior in childhood is a subject that has gained attention in research conducted in Psychology and in other fields, broadening and intensifying related scientific discussions regarding this topic (see Borsa & Bandeira 2014a). Such an emphasis exists because the aggressive behavior is a common problem observed in both children and adolescents and leads to severe social and adaptive problems (Miller & Lynam 2006).

Aggressive behavior can be comprehended as any behavior intended to inflict harm on someone or damage something (Berkowitz 1993; Dodge & Coie 1987). However, this complex and multi-determined phenomenon has not yet achieved a consensual definition. When not modulated or, when intense and persistent, these behaviors may lead to psychosocial losses over the course of life (Borsa & Bandeira 2014a).

This study focuses on disruptive aggressive behavior, that is, non-modulated, intense and persistent aggressive behavior, which poses a risk to one's development and is associated with difficulty interacting with others and rejection of peers, learning difficulties, and dropping out

of school, in addition to depression, anxiety and impulsiveness (Tremblay et al. 2008). One important focus of research addressing aggressive behavior during childhood has been an attempt to classify this behavior in regard to its motivation and manifestations. Hence, there are different classification approaches depending on studies' objectives and underlying theories (Little et al. 2003).

Concerning to how aggressive behaviors manifest, they can be classified as physical (hitting, biting, kicking) or verbal (offending, hurting feelings, gossiping) or as direct and indirect. Direct behaviors include acts such as physically and verbally assaulting someone, destroying objects, arguing, threatening, ridiculing, etc. Indirect behavior includes acts such as disturbing the environment, practicing virtual aggression, making intrigues, gossiping, or damaging the image of people, etc. (Little et al. 2003; Smith et al. 2008).

In terms of etiology, aggressive behavior can be classified as proactive aggression, also called instrumental, offensive, or predatory aggression, and as reactive aggression, also referred to as impulsive, affective or defensive aggression (Crick & Dodge 1996; Dodge & Coie 1987; Vitaro et al. 2006). Studies suggest that these subtypes of aggressive behavior are preceded by different variables, associated with

Correspondence: juliborsa@gmail.com
Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil

Springer Open

different behavioral results, oriented by different emotional and cognitive processes related to different social experiences (Hubbard et al. 2010).

Proactive aggressive behaviors are characterized by deliberate aggression toward an instrumental goal, i.e., to achieve a desired goal (Dodge & Coie 1987). In other words, it is a behavior motivated by the desire to persevere aimed at an objective (for instance, obtaining money or material goods, to hurt or injure a person). In general, proactive aggressive behavior is "cold-blooded" and is related to a higher sense of self-efficacy. Hence, proactive aggressive behavior is associated with the expectation of positive results (Crick & Dodge 1996). Reactive aggressive behavior, in turn, is characterized by defensive impulsive responses in the face of a provocation (Dodge & Coie 1987). It is associated with anger and frustration and, for this reason, this behavior is called "hot-blooded" behavior, due to the physiological response attached to it. Children with reactive behavior present information-processing deficits and therefore tend to perceive hostility in their peers' actions even in ambiguous situations in which there is no clear provocative or aggressive intention (Crick & Dodge 1996; Dodge & Coie 1987).

There is no consensus in the literature regarding the importance of distinguishing between proactive and reactive aggression (Bushman & Anderson 2001) because both types of aggression may occur simultaneously, which leads one to assume that these represent a continuum coexisting at different levels in each child (Hubbard et al. 2010; Poulin & Boivin 2000). Dodge and Coie (1987) developed an instrument composed of three items designed to assess proactive aggressive behavior and three items intended to assess reactive aggressive behavior. Correlation between the two scales was equal to .76, suggesting that the instrument has a one-dimensional structure. Higher correlation (.83) was found in the further studies conducted by Price and Dodge (1989).

The study conducted by Bushman and Anderson (2001), titled "Is it time to pull the plug on the hostile versus instrumental aggression dichotomy?", proposes a discussion of the real need to discriminate between proactive and reactive aggressive behaviors. According to the authors, aggressive behavior is usually hybrid; i.e., it may aim for different goals (concrete and/or subjective) and rely on different motivational factors. Additionally, the aggressor may plan aggression and at the same time experience anger before or during the aggression. The authors consider the dichotomization of aggressive behavior to undermine scientific advancements seeking to understand aggressive behavior and intervene in the face of it.

In Brazil, instruments to assess aggressive behavior are scarce, as demonstrated in the Borsa and Bandeira (2011) systematic review. This lack of instruments hinders studies aiming to investigate aggressive behavior among children using instruments adapted to the Brazilian context. The objective of this study is to present a new measure, called "Peer Aggressive Behavior Scale (PAB-S)", to assess aggressive behavior among children. Study I described the procedures concerning the instrument's development, specifically the development of items and initial evidence of content validity. Furthermore, it also presents the refinement process of the scale while investigated initial evidence of validity based on its internal structure. Finally, study II assessed the refined version with 25 items in an independent sample and compared its goodness-of-fit indexes with those presented by the initial version, which had 39 items. The idea of dividing this paper into two studies was to ease understanding of the different methodological stages used. It is worth noting that this study's main objective was to refine the instrument to obtain a briefer and more economical measure.

Methods
Study I
Study I describes the development of the Peer Aggressive Behavior Scale (PAB-S), specifically reporting the procedures used to develop its items and evidence of content validation. Furthermore, it presents the refining procedures and initial evidence of validity based on the instrument's internal structure.

Participants
A total of 619 children (52 % girls) aged between eight and 12 years old ($M = 9.98$; $SD = 1.9$), attending from the 1st to the 5th grade in both public and private schools in the city of Porto Alegre, RS, Brazil participated in the study.

Instrument
The Peer Aggressive Behavior Scale (PAB-S): The PAB-S is a 39-item self-report questionnaire: 20 items address proactive aggressive behavior and 19 address reactive aggressive behavior. The content of the items involve both direct physical (e.g., "*When a colleague does something that makes me sad, I beat him*") and verbal (e.g. "*I scream to my colleagues for them to do what I want*") manifestations, as well as indirect and relational behavior (e.g., "*I gossip about my colleagues to become more popular*"). Each item is assessed through an analogue-visual five-point Likert scale, which varies according to the frequency of behavior.

Procedures of the PAB-S development
The PAB-S was developed based on an extensive review of literature on child development and on theories explaining aggressive behavior. Specifically, we sought to understand the contributions of Social Learning Theory (Bandura 1973), Frustration-Aggression Theory (Berkowitz

1993), and Information Processing Theory (Dodge & Crick 1994). An unsystematic review of literature was conducted to identify the main instruments currently used to assess aggressive behaviors among children and adolescents. Hence, a search was conducted through the PsycINFO database without restricting papers by date and using the descriptors 'Aggression'; 'Aggressive'; 'Proactive'; 'Reactive' associated with the words 'Child' and/or 'Adolescent' and 'Evaluation'; 'Assessment'; 'Questionnaire'; 'Scale'; 'Instrument'; 'Checklist'. A total of 50 studies were found and the used instruments were carefully read and analyzed. Among the instruments found, the Teacher- Report Scale (Dodge & Coie 1987), the Revised Teacher Rating Scale for Reactive and Proactive Aggression (Brown et al. 1996), the Peer Conflict Scale (PCS – Marsee & Frick 2007), the Reactive–Proactive Aggression Questionnaire (RPQ - Raine et al. 2006), the Parent-rated Scale of Reactive and Proactive Aggression (PRPA – Kempes et al. 2006), and the Aggressive Behavior Scale (Little et al. 2003) were used as examples for the PAB-S construction. Specifically, the instruments were used to understand the types of behavior assessed by the items, the rapport instructions, as well as the structure of items response.

Initially, 39 self-report items were developed in order to match the specificity of aggressive behavior among children that commonly takes place in the school context. The expressions were chosen to ease the understanding of children at different ages and from different areas of Brazil. Items referring to physical, verbal and relational (aggression intended to harm someone socially) forms of aggression were included for both proactive and reactive behaviors.

Content validity analyses included a detailed evaluation of the instrument by expert judges (a group of five graduate students in psychology and experts on psychological assessment and on the construction of psychological instruments). All judges were aware of the definitions of aggressive behaviors and the underlying theories used to construct the scale. Each item was discussed concerning the potential presence of confusing, repetitive, redundant terms or hard-to-understand terms. The quality of the scale's graphic structure and layout was verified. A form was provided to the experts to assess the scale. The form's first part referred to the scale's general aspects and included the questions aiming to evaluate the clearness and objectives of the instructions, the adequacy of the response style, the need of adding further items or reducing redundant ones, etc. The second part contained questions concerning item's adequacy in theoretical, grammatical and idiomatic terms. After an in depth evaluation, minor changes were conduct, specifically in regard to terms which could be generalized to different cultural contexts and to be comprehended for children with different ages.

After the content validity evaluation, a pilot study was conducted with a group of children from three different Brazilian states: three children from Porto Alegre, RS (an 11-year old boy, an 8-year old girl, and a 9-year old girl), three from Aracaju, SE (12-year old boy, an 8-year old girl and an 11-year old girl), and two children from João Pessoa, PB (two 10-year old boys). All items were fully comprehensible. Minor suggestions were given regarding the instructions of the scale, which was modified. After these modifications, the scale was considered by the authors ready to be used.

Procedures of data collection

First, schools were contacted either personally or by phone. Each school received a summary of the research project and clarification in regard to the study's objectives and procedures. The schools that consented to participate in the study signed a consent letter authorizing the study, while the children's parents or legal guardians received an informed consent form.

The PAB-S was collectively applied in classrooms in such a way that the children's routine curricular activities were not interrupted. Ethical issues were ensured according to Resolution 466/2012, Brazilian Ministry of Health. All procedures met the guidelines of the Institutional Review Board at the Federal University of Rio Grande do Sul (Process number: 05283812.2.0000.5334).

Procedures of data analysis

Initially, a Parallel Analysis (Horn 1965), with random permutation of observed data (Timmerman & Lorenzo-Seva 2011), was conducted to verify the PAB-S's dimensionality. After establishing the number of factors to be retained, an Exploratory Structural Equation Modeling (ESEM) was conducted. The ESEM is an exploratory factor analysis technique that presents confirmatory goodness-of-fit indexes. The estimation method Weighted Least-Squares Mean and Variance-Adjusted (WLSMV) was used in a polychoric correlation matrix, acknowledging the data's ordinal nature (Muthén & Muthén 2012).

The following goodness-of-fit indexes were assessed: root mean squared error of approximation (RMSEA); comparative fit index (CFI); Tucker-Lewis index (TLI); standardized root mean square residual (SRMR). RMSEA below .08, with confidence interval (90 %) not exceeding 1.00; CFI and TLI above .90 (preferably above .95); and SRMR below .100 indicate the model's goodness-of-fit (Brown 2006).

In addition to assessing the model's goodness-of-fit, ESEM enabled verifying the instrument's structural problems through modification indexes (MI) (Muthén & Muthén 2012). MIs above 30 were considered as evidence of misfit (Brown 2006). Problematic items were excluded based on the adopted criteria (MI ≥ 30; Brown 2006) and

on methodological decisions detailed below. Parallel analysis was performed using the Factor program, version 9.2 (Lorenzo-Seva & Ferrando 2006). ESEM was performed using Mplus version 7.11 (Muthén & Muthén 2012).

Finally, we sought to assess the instrument's psychometric properties more deeply through Item Response Theory (IRT) by using the Rating Scale Model (Andrich 1978). The Rating Scale Model was implemented with two main objectives: 1) to assess the goodness-of-fit indexes for the items through infit and outfit indicators; and 2) to verify whether the items of the instrument presented differential functioning (DIF) for sex and age groups.

Infit and outfit indexes quantify residue for the items in regard to the tested model (Bond & Fox 2007; Linacre 2011). Infit assesses unexpected response patterns of individuals who present a *theta* level equivalent to the item's level of difficulty. Outfit, in turn, verifies unexpected response patterns of those who present a *theta* level below or above the item's level of difficulty. The ideal value (for infit and outfit) is 1.00 (mean square), considered acceptable values between .50 and 1.50 (Linacre 2011). Items with values beyond expectations were excluded.

For the DIF analyses, age was separated into two groups: Group I (7–10 years old) and Group II (11–13 years old). The separation of two age groups was made considering an important behaviors difference in terms of developmental stages of children aged from 7 to 10 and for those aging from 11 to 13 (Papalia et al. 2013). Mantel-Haenszel procedure and $p \leq .05$ were used to assess items with differential functioning. The magnitude of DIF was interpreted through DIF contrast: values between |.00| and |.43| were considered to be low; values between |.44| and |.64| were considered moderate; and those above |.64| were considered to be high (Linacre 2013). Rasch analyses were performed using WinSteps version 3.72 (Linacre 2013).

Results

The results of the parallel analysis indicate the PAB-S has one dimension (Dataset explained variance = 58.3 % for factor 1 and 4.9 % for factor 2; Parallel Analysis explained variance = 12.6 % for factor 1 and 5.1 % for factor 2). Hence, the scale was submitted to a single-factor ESEM presented in Table 1. The goodness-of-fit indexes of the structural model were: χ^2 df = 1881.55 (702); χ^2/df = 2.68; RMSEA (.052; 90 % CI = .049 − .055); CFI = .94; TLI = .94; SRMR = .066.

Even though the goodness-of-fit indexes were appropriate for the scale's initial structure, a follow-up inspection, using MIs, showed high correlation of errors between a few pairs of items. Item 33 ('When I want to hurt someone, I take or mess up some of his/her things') presented high residual correlation with item 37 ('When someone hurts me, I take or mess up some of his/her things';

Table 1 Factor loadings of the PAB-S −39 items

Items	Factor loads
Item 1	.672
Item 2	.508
Item 3	.773
Item 4	.646
Item 5	.719
Item 6	.698
Item 7	.791
Item 8	.696
Item 9	.827
Item 10	.888
Item 11	.805
Item 12	.806
Item 13	.688
Item 14	.800
Item 15	.843
Item 16	.702
Item 17	.672
Item 18	.817
Item 19	.708
Item 20	.742
Item 21	.591
Item 22	.786
Item 23	.724
Item 24	.724
Item 25	.760
Item 26	.848
Item 27	.666
Item 28	.847
Item 29	.735
Item 30	.872
Item 31	.697
Item 32	.784
Item 33	.819
Item 34	.715
Item 35	.730
Item 36	.844
Item 37	.825
Item 38	.700
Item 39	.838
Explained variance	.45
Alpha reliability	.96

MI = 50.15; r_{33*37} = .57). Item 24 ('When someone speaks ill of me, I call him/her bad names') presented high residual correlation with item 34 ('When someone

speaks ill of me, I also speak ill of him/her'; MI = 38.48; r_{24*34} = .35). Item 13 ('When someone does something I don't like, I leave him/her out of my group') presented high residual correlation with item 19 ('If any of my friends doesn't agree with my rules, I leave him/her out of my play group'; MI = 30.42; r_{13*19} = .38). Considering the content of pairs of items in qualitative terms shows there is considerable conceptual overlap among them. Hence, we opted to exclude the item with the lowest factor load from the pairs, according to Brown (2006). In this way, items 13, 33, and 34 were deleted and the PAB-S became a scale with 36 items.

Rasch's analysis
Rasch's analysis, specifically in regard to infit, showed that all the 36 were between .87 and 1.50. Therefore, they posed no problem in regard to this criterion. In regard to the outfit analysis, however, three items presented values above the cutoff point (1.50), namely: Item 2 ('I argue with my friends to show I'm right'; Outfit mean square = 2.11); Item 21 ('I make fun of my friends to show I'm funny'; Outfit mean square = 1.76) and Item 27 ('I get involved with fights at school'; Outfit mean square = 2.91). These items were excluded for presenting an unexpected pattern of responses. Thus, the scale remained with 33 items.

Differential Item Functioning (DIF) for sex and age groups
DIF analysis was performed for sex and age groups (Group I included 7 to 10 year-old children and Group II included 11 to 13 years old) to investigate which items presented response bias for these two sociodemographic characteristics. Hence, the remaining 33 items were analyzed. According to the Mantel-Haenszel procedure ($p < .05$), seven items presented DIF for sex and one item presented DIF for age (Table 2). Those items presenting bias response for gender (4, 8, 11, 16, 20, 28 and 32) and age (item 17) were excluded. Hence, the PAB-S's refined version retained 25 items.

After refinement procedures for the PAB-S, we compared goodness-of-fit indexes of the scale's initial version (39 items) and the refined version (25 items), performing a new ESEM. The goodness-of-fit indexes for the initial version were, as previously demonstrated: χ^2 df = 1881.55 (702); χ^2/df = 2.68; RMSEA (.052; 90 % CI = .049 − .055); CFI = .94; TLI = .94; SRMR = .066 and for the refined version were (25 items): χ^2 df = 579.183 (275); χ^2/df = 2.11; RMSEA (90 % CI) = .042 (.037−.047); CFI = .98; TLI = .97; SRMR = .052. All the goodness-of-fit indexes performed better for the refined version.

Considering that some improvement was expected in the scale's refined version, since new analysis was performed considering this same refinement sample, a new study was conducted to compare the initial and refined versions with an independent sample.

Study II
This study's aim was to assess the psychometric properties of the refined version (25 items) of the PAB-S's with an independent sample and to compare goodness-of-fit indexes of this version with the initial version (39 items). In this study, the initial version (39 items) was employed.

Participants
A total of 355 children (52.7 % girls) aged between 7 and 13 years old (M = 9.92; SD = 1.26) attending the 6th grade in both public and private schools in the city of Rio de Janeiro and metropolitan region, RJ, Brazil participated in this stage.

Table 2 Differential item functioning of the PAB-S

DIF analysis	Mantel-Haenszel		DIF contrast
Sex (Boys versus girls)	Size	P = value	Magnitude
Easier for girls			
Item 4. When someone provokes me, I yell at him/her	.11	.021	.33
Item 16. I ask my friends not to be friends with people I don't like	.08	.016	.37
Item 20. I leave those I don't like out of my group	-.51	.003	.44
Item 32. When someone makes me mad, I tell lies about him/her	-.78	.017	.19
Easier for boys			
Item 8. When someone hits me, I hit him/her back	-.81	.000	-.33
Item 11. When someone makes me mad, I hit him/her	-.61	.010	-.26
Item 28. I call those I don't like bad names	-.98	.001	-.47
Age (7–10 years old versus 11–13 years old)	Size	P = value	Magnitude
Easier for younger children			
Item 17. When someone does something bad to me, I ask people to stop talking to her/him	-.43	.048	- .28

Instruments

In this study, the children responded to the PAB-S's initial version (39 items), however, both the initial and refined (with 25 items) versions were used according to the refining procedures described in the previous studies.

Procedures of data collection

Data collection was part of the activities performed in the study titled: "Prevalence of aggressive behavior among school-aged children in the city of Rio de Janeiro", which was submitted to and approved by the Institutional Review Board at the State University of Rio de Janeiro (Process number: 24367113.0.0000.5282). Similar to the previous study, we first contacted the schools and presented a summary of the study project, a letter clarifying the nature of the study and a consent letter. The schools that agreed to participate signed the letter of consent and, afterwards, an informed consent form was sent to the children's parents or legal guardians. Data were collected in groups in either classrooms or in areas established by the schools. The study was in compliance with all ethical guidelines for such studies.

Procedures of data analysis

Two ESEM were performed, one with the PAB-S's initial version (39 items) and another with the refined version (25 items), using the same criteria presented in Study I.

Results

The goodness-of-fit indexes for the model with 39 items were: χ^2 df = 1258.01 (702); χ^2/df = 1.79; RMSEA (90 % CI) = .047 (.043–.051); CFI = .95; TLI = .94; SRMR = .072 and for the refined version (25 items) were: χ^2 df = 508.475 (275); χ^2/df = 1.85; RMSEA (90 % CI) = .049 (.042–.056); CFI = .96; TLI = .96; SRMR = .064.

The goodness-of-fit indexes of both versions were satisfactory and very similar. According to the previous study, the results presented appropriate goodness-of-fit for both the initial and the refined versions. This study shows that χ^2/df and RMSEA favor the initial version, while the CFI, TLI and SRMR favor the refined version.

Discussion

In these studies we sought to present evidences of content and internal structure validity of the PAB-S, beyond presenting the refinement procedures for the scale. Evidence of content validity was gathered through a thorough qualitative investigation, employing external judges and a pilot study. After having an adequate and comprehensive measure, we employed it in a large Brazilian non-representative sample, investigated its psychometric properties and refined the scale from a 39-item to a 25-item version.

Results of the exploratory factor analysis did not distinguish the theoretical difference proposed by Dodge and Coie (1987) between proactive aggression and reactive aggression. Even though some of the literature reports proactive and reactive aggression to have distinct etiologies, distinct neurophysiological processes and to be manifested in distinct conditions (Hubbard et al. 2010; Raine et al. 2006), other studies indicate a high correlation between proactive and reactive aggression (Price & Dodge 1989; Rodkin & Roisman 2010), suggesting these behaviors coexist and are parts of a single dimension (Bushman & Anderson 2001; Hubbard et al. 2010; Poulin & Boivin 2000). Thus, albeit a distinction of the proactive and reactive aggression. factors was not found for the PAB-S, its unidimensionality can be also easily comprehended.

In terms of the refining processes, the pairs 33–37, 24–34 and 13–19 presented important residual correlation. According to Brown (2006), two main hypotheses can be proposed in regard to the presence of correlated errors. The first hypothesis refers to a potential non-modeled latent variable explaining the residual variance and co-variance of items. The second hypothesis refers to an overlap of content among items so that the parts not explained by the modeled latent variable (residue) correlates with each other. A qualitative analysis of these items suggests there is an overlap of content among items. Based on this hypothesis, we opted to exclude from each pair of items (33–37, 24–34 and 13–19) the one that presented the lowest factor loading. Considering that items were similar in their content, the exclusion of these items did not diminished content validity of the scale.

Rasch's analysis, specifically infit analysis, shows that the 36 remaining items were adequate. In regard to outfit, however, three items (2, 21 and 27) presented unexpected pattern of answers, indicating a discrepancy in the answers related to the sample's *theta* level in comparison to the item's level of difficulty (i.e., children with *theta* levels above or below these item's level of difficulty were not answering as expected). Considering this pattern of answers to be inappropriate, the items were excluded.

DIF analysis presented response bias for sex (items 4, 8, 11, 16, 20, 28 and 32) and age (item 17) and, for this reason, they were excluded. Table 2 shows that the items concerning relational verbal aggressive behavior were more easily endorsed by girls, while physical aggressive behavior was more easily endorsed by boys. This information corroborates evidence reported in the literature, which shows that boys and girls manifest aggressive behavior in different ways (Leff et al. 2014; Nivette et al. 2014; Tapper & Boulton 2004). While boys more frequently present physically aggressive behavior, girls tend

to present verbally and relationally aggressive behaviors (Borsa & Bandeira 2014b; Card et al. 2008; Leff et al. 2014; Lim & Ang 2009). Interestingly, item 28 was more easily endorsed by boys. Even though this item represents a type of verbally and relationally aggressive behavior, it represents a more intense and confrontational (direct) behavior if compared to the items that assess verbal aggressive behavior. The literature shows that boys tend to present more confrontational behavior than girls (Card et al. 2008; Lim & Ang 2009; Tapper & Boulton 2004).

Finally, in regard to age, item 17 (relational aggressive behavior) was more easily endorsed by younger children. Even though aggressive behavior of the relational type is more frequently reported by older children (Borsa 2012), the literature shows that this type of behavior is more stable over the course of childhood, only changing the ways in which it manifests (Leff et al. 2014). The age groups considered in this study were very close to each other so that it was not possible to explain this difference. Further analyses using larger and diversified samples can enable better inferences about the level of difficulty of each item for children belonging to different age groups and living in different regions of Brazil.

In regard to the psychometric properties of the PAB-S's initial and refined versions, analyses presented in Study I show that all the goodness-of-fit indexes improved in the refined version. To verify this information, a second study was conducted to compare both versions using an independent sample. The results show that the goodness-of-fit indexes of both versions were appropriate and virtually the same, so that the initial version (39 items) and the refined version (25 items) were considered acceptable. Considering the similarity of both versions, we chose the refined version as the most adequate based on its brevity and initial objectives proposed for this study.

The study has some important limitations, such as the use of a non-probabilistic and limited sample, given the cultural diversity and geographical dimension of Brazil. There is also the fact that the age groups are too similar to each other and, therefore, verifying each item's level of difficulty was not possible. Nevertheless, this study met its objectives and provides an easy-to-understand new instrument for the academic community to assess aggressive behavior among children. It is a starting point for further studies to assess other evidence of validity for the PAB-S in the Brazilian context. It is important to note that the validity of an instrument does not cease with a single study and a broader source of evidence of validity is needed to consider an instrument valid in a given context (Urbina 2007). Validity is seen as the degree to which all evidence gathered corroborates the interpretation intended for the scores obtained by a given test considering its purpose (AERA, APA, NCME, 2014). Therefore, further studies are needed to confirm and/or add evidence of validity for the PAB-S in the Brazilian context.

Further studies should assess the extent to which the scale's initial (39 items) and the refined (25 items) versions are equivalent in terms of external indicators. For instance, the performance of both versions should be verified in terms of correlation with external indicators (convergent and criterion validity). Additionally, further studies with clinical and non-clinical groups could attest to the scale's sensitivity and specificity and compare the performance of both versions in light of these two indicators.

References
American Educational Research Association [AERA], the American Psychological Association [APA] and the National Council on Measurement in Education [NCME]. Standards for educational and psychological testing. Washington: American Educational Research Association; 2014.
Andrich D. A rating formulation for ordered response categories. Psychometrika. 1978;43(4):561–73. doi:10.1007/BF02293814.
Bandura A. Aggression: a social learning analysis. Englewood Cliffs: Prentice-Hall; 1973.
Berkowitz L. Aggression: its causes, consequences, and control. New York: McGraw-Hill; 1993.
Bond T, Fox C. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah: LEA; 2007.
Borsa JC. Adaptação e validação transcultural do Questionário de Comportamentos Agressivos e Reativos entre Pares (Q-CARP), Tese de Doutorado não publicada. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2012.
Borsa JC, Bandeira DR. Uso de instrumentos psicológicos de avaliação do comportamento agressivo infantil: Análise da produção científica brasileira. Avaliação Psicológica. 2011;10(2):193–203. Retrieved from: http://pepsic.bvsalud.org/pdf/avp/v10n2/v10n2a10.pdf.
Borsa JC, Bandeira DR. Comportamentos Agressivos na Infância: da teoria à prática. São Paulo: Casa do Psicólogo; 2014a.
Borsa JC, Bandeira DR. The Peer Aggressive and Reactive Behaviors Questionnaire (PARB-Q): evidence of validity in the Brazilian context. Trends Psychiatr Psychother. 2014b;36(2):89–100. doi:10.1590/2237-6089-2013-0051.
Brown TA. Confirmatory factor analysis for applied research. New York: The Guilford Press; 2006.
Brown K, Atkins MS, Osborne ML, Milnamow M. A revised teacher rating scale for reactive and proactive aggression. J Abnorm Child Psychol. 1996;24(4):473–80. doi:10.1007/BF01441569.
Bushman BJ, Anderson CA. Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? Psychol Rev. 2001;108:273–9. doi:10.1037/0033-295X.108.1.273.
Card NA, Stucky BD, Sawalani GM, Little TD. Direct and indirect aggression during childhood and adolescence: a meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. Child Dev. 2008;79(5):1185–229. doi:10.1111/j.1467-8624.2008.01184.x.
Crick NR, Dodge KA. Social information-processing mechanisms in reactive and proactive aggression. Child Dev. 1996;67(3):993–1002. doi:10.1111/j.1467-8624.1996.tb01778.x.
Dodge KA, Coie JD. Social-information-processing factors in reactive and proactive aggression in children's peer groups. J Pers Soc Psychol. 1987;53(6):1146–58. doi:10.1037/0022-3514.53.6.1146.

Dodge KA, Crick NR. A review and reformulation of social information-processing mechanisms in children's social adjustment. Psychol Bull. 1994;115(1):74–101. doi:10.1037/0033-2909.115.1.74.

Horn JL. A rationale and technique for estimating the number of factors in factor analysis. Psychometrika. 1965;30(1):179–85. doi:10.1007/BF02289447.

Hubbard JA, McAuliffe MD, Morrow MT, Romano LJ. Reactive and proactive aggression in childhood and adolescence: precursors, outcomes, processes, experiences, and measurement. J Pers. 2010;78(1):95–118. doi:10.1111/j.1467-6494.2009.00610.x.

Kempes M, Matthys W, Maassen G, van Goozen S, van Engeland H. A parent questionnaire for distinguishing between reactive and proactive aggression in children. Eur Child Adolesc Psychiatry. 2006;15(1):38–45. doi:10.1007/s00787-006-0502-2.

Leff SS, Waasdorp TE, Waanders C, Paskewich BS. Better understanding and intervening to prevent relational aggression. In: Weist MD, Lever NA, Bradshaw CP, Owens JS, editors. Handbook of school mental health. New York: Springer; 2014. p. 171–82. doi:10.1007/978-1-4614-7624-5_13.

Lim SH, Ang RP. Relationship between boys' normative beliefs about aggression and their physical, verbal, and indirect aggressive behaviors. Adolescence. 2009;44(175):635–50. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/19950874.

Linacre JM. A user's guide to WINSTEPS/MINISTEP Rasch-model computer programs. Chicago: MESA Press; 2011.

Linacre JM. Winsteps® Rasch measurement computer program user's guide. Beaverton: Winsteps.com; 2013.

Little TD, Jones SM, Henrich CC, Hawley PH. Disentangling the 'whys' from the 'whats' of aggressive behavior. Int J Behav Dev. 2003;27(2):122–33. doi:10.1080/01650250244000128.

Lorenzo-Seva U, Ferrando PJ. FACTOR: a computer program to fit the exploratory factor analysis model. Behav Res Methods. 2006;38(1):88–91. doi:10.3758/BF03192753.

Marsee MA, Frick PJ. Exploring the cognitive and emotional correlates to proactive and reactive aggression in a sample of detained girls. J Abnorm Child Psychol. 2007;35:969–81. doi:10.1007/s10802-007-9147-y.

Miller JD, Lynam DR. Reactive and proactive aggression: Similarities and differences. Personal Individ Diff. 2006;41(8):1469–80. doi:10.1016/j.paid.2006.06.004.

Muthén LK, Muthén BO. Mplus: statistical analysis with latent variables. User's guide. Los Angeles: Muthén & Muthén; 2012.

Nivette AE, Eisner M, Malti T, Ribeaud D. Sex differences in aggression among children of low and high gender inequality backgrounds: A comparison of gender role and sexual selection theories. Aggress Behav. 2014;40(5):451–64. doi:10.1002/ab.21530.

Papalia ED, Olds SW, Feldman RD. Desenvolvimento Humano. 12th ed. Porto Alegre: Artmed; 2013.

Poulin F, Boivin M. Reactive and proactive aggression: evidence of a two-factor model. Psychol Assess. 2000;12(2):115–22. doi:10.1037/1040-3590.12.2.115.

Price JM, Dodge KA. Reactive and proactive aggression in childhood: relations to peer status and social context dimensions. J Abnorm Child Psychol. 1989;17(4):455–71. doi:10.1007/BF00915038.

Raine A, Dodge K, Loeber R, Kopp L, Lynam D, Reynolds C, et al. The reactive-proative aggression (RPQ) questionnaire: differential correlates of reactive and proactive in adolescent boys. Aggress Behav. 2006;32(2):159–71. doi:10.1002/ab.20115.

Rodkin PC, Roisman GI. Antecedents and correlates of the popular-aggressive phenomenon in elementary school. Child Dev. 2010;81(3):837–50. doi:10.1111/j.1467-8624.2010.01437.x.

Smith PK, Smith C, Osborn R, Samara M. A content analysis of school anti-bullying policies: progress and limitations. Educ Psychol Pract. 2008;24(1):1–12. doi:10.1080/02667360701661165.

Tapper K, Boulton MJ. Sex differences in levels of physical, verbal and indirect aggression amongst primary school children and their associations with beliefs about aggression. Aggress Behav. 2004;30(2):123–45. doi:10.1002/ab.20010.

Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychol Methods. 2011;16:209–20. doi:10.1037/a0023353.

Tremblay RE, Gervais J, Petitclerc A. Prévenir la violence par l'apprentissage à la petite enfance. Montreal: Centre D'Excellence pour Le Développement dês Jeunes Enfants; 2008.

Urbina S. Fundamentos da Testagem Psicológica. Porto Alegre: Artmed; 2007.

Vitaro F, Barker ED, Boivin M, Brendgen M, Tremblay RE. Do early difficult temperament and harsh parenting differentially predict reactive and proactive aggression? J Abnorm Child Psychol. 2006;34(5):681–91. doi:10.1007/s10802-006-9055-6.