10.33208/PC1980-5438v0031n02A08

The draw-a-person test in the evaluation of child aggression: a pilot study

O desenho da figura humana na avaliação da agressividade infantil: um estudo piloto

El dibujo de la figura humana en la evaluación de la agresividad infantil: un estudio piloto

*Juliane Callegaro Borsa**

Abstract

The aim of this study was to produce a list of indicators of aggression for the draw-a-person test and to investigate evidence of their validity. First, 21 items were developed, which were evaluated by agreement among judges. Next, it was assessed whether these items were able to discriminate groups of children with higher and lower aggression scores according to a self-report measurement. Finally, it was investigated whether these items could identify children with or without a history of aggression. The items "clawed fingers", "presence of teeth" and "exaggerated shoulders in the human figure" were found to be potentially capable of identifying aggression in the sample, corroborating the findings in the literature. The data, however, is inconclusive and further studies are needed to appraise the validity of the proposed items for the evaluation of aggression through the draw-a-person test.

Keywords: aggression; children; draw-a-person test; psychological evaluation.

Resumo

O objetivo deste estudo foi elaborar e investigar evidências de validade de uma lista de indicadores de agressividade no desenho da figura humana. Primeiramente, foram elaborados 21 itens, os quais foram avaliados por meio da concor-

* Professora do Departamento de Psicologia da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brasil.
juliborsa@gmail.com

Este estudo contou com o apoio da FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro) e do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

dância entre juízes. Então, foi avaliado se tais itens eram capazes de discriminar grupos de crianças com maiores e menores escores de agressividade segundo uma medida de autorrelato. Finalmente, investigou-se se esses mesmos itens poderiam discriminar crianças com e sem histórico de agressividade. Os itens "dedos em ponta de garras", "presença de dentes" e "ombros reforçados na figura humana" foram os que se mostraram potencialmente capazes de identificar agressividade na presente amostra, corroborando os achados da literatura. Os dados, porém, são inconclusivos e novos estudos são necessários para investigar a validade dos itens propostos para avaliação da agressividade por meio do desenho da figura humana.

Palavras-chave: agressividade; crianças; desenho da figura humana; avaliação psicológica.

Resumen

El objetivo de este estudio fue elaborar e investigar la evidencia de validez de una lista de indicadores de agresión en el test del dibujo de la figura humana. Primero, se elaboraron 21 ítems y se evaluó el acuerdo entre evaluadores. Posteriormente, se evaluó si estos ítems podían discriminar grupos de niños con puntajes de agresividad más altas y más bajas según una medida de autoinforme. Por fin, se investigó si estos mismos ítems podían identificar a los niños con y sin histórico de agresividad. Los items "dedos en garras", "presencia de dientes" y "hombros reforzados en la figura humana" fueron aquellos potencialmente capaces de identificar la agresividad en la presente muestra, corroborando los hallazgos de la literatura. Sin embargo, los datos no son concluyentes y se necesitan nuevos estudios para investigar la validez de los ítems propuestos para la evaluación de la agresividad por medio del dibujo de la figura humana.

Palabras clave: agresividad; niños; diseño de la figura humana; evaluación psicológica.

Introduction

Aggression is defined in the literature as the intentional action of causing harm to someone or something (Bushman & Anderson, 2001). Aggression may be manifested directly, such as by hitting, pushing, screaming or offending, or indirectly, as in inciting gossip, lying or defaming someone to harm their image (Borsa & Bandeira, 2014). When it occurs in childhood, in an intense and persistent way, aggression can be a risk factor for development, being associated with

social and adaptive problems, such as difficulty of interaction with and rejection by peers, learning difficulties, dropping out of school, and symptoms of depression, anxiety and impulsivity (Evans & Fite, 2019; McQuade, Breaux, Gómez & Zakarian, 2016; Robertson, Forbes & Thyne, 2017; Santo, Bass, Stella-Lopez & Bukowski, 2017).

Studies highlight the importance of interventions to improve the prognosis and reduce the risks to development in aggressive children (Petersen, Souza & Wessel, 2014; Landim & Borsa, 2017). Considering this scenario, it is imperative that there are investments in the evaluation of aggression, so that effective forms of intervention may be considered (Rocha, Emerich & Silvares, 2014). Accordingly, psychological evaluation has an important preventive role, since it allows early identification and comprehension of aggression in children, optimizing the referral for the different types of possible interventions (Borsa & Bandeira, 2014).

Several standardized instruments (scales and inventories) have been developed to assess aggression in childhood (Borsa, 2016). These instruments, however, have some limitations. Self-report instruments, for example, present the important bias of social desirability. Particularly in the evaluation of emotional and personality indicators, the answers to certain items are based on the child's judgment of which response is the most socially appropriate, and not on his or her own behavior (Borsa & Muniz, 2016). Instruments filled by others, on the other hand, also present weaknesses, since they depend on the judgment of a third party regarding the child, which often prevents an impartial evaluation. For example, parents tend to compare their child to their friends' and relatives' children or other children, while teachers tend to assess the children or adolescents by comparing them with other students (Borsa & Muniz, 2016).

Draw-a-person test in the evaluation of aggression in children

The psychological evaluation of children presents peculiarities that sets it apart from evaluations of adults. Given the relevance of non-verbal communication common to this age group, it is necessary to include other data collection tools, in addition to standardized tests such as play and graphic techniques. Specifically, drawings are recommended as "ice breaker" techniques, whether in the psychological assessment or in the psychotherapeutic process with children. They are familiar procedures and, therefore, not very anxiogenic, when compared to standardized instruments (Borsa & Bauermann, 2013).

For the child, creating drawings is a form of communication, similar to gestures and verbalizations. The drawings reveals the emotional and personality aspects of the drawer, just as happens in all artistic work (Buck, 2003; Machover, 1967). Goodenough (1974) was one of the pioneers in suggesting drawing as a useful resource in psychological assessment. According to the author, in drawing the child expresses not what they see, but what they know about the human figure being drawn (Suehiro, Benfica & Cardim, 2016). Currently, the scientific literature encompasses different drawing techniques used in clinical practice, among them the draw-a-person test (Machover, 1967).

The draw-a-person test is a tool that allows for the evaluation of different psychological characteristics, such as cognitive development, creativity, personality and emotions (Suehiro et al., 2016). Cost-effective, easy and quick to apply, this technique has been one of the instruments most used by Brazilian psychologists (Bandeira & Arteche, 2008). Its applicability is mainly with children, due to being a nonverbal task and, in the clinical context, a free technique of graphic expression (Bandeira & Arteche, 2008; Borsa & Bauermann, 2013; Suehiro et al., 2016).

Since Goodenough's contributions, several draw-a-person test assessment systems have been developed and classified according to the objectives of the assessment, from three perspectives: cognitive, projective and emotional (Suehiro et al., 2016). The first considers the drawing as a measure of cognitive development, through the scoring of the existence and quality of the items present in the figures (Goodenough, 1974). The second, however, sees the drawing as a form of manifestation of the unconscious and expressive aspects of the personality (Hammer, 1991, Machover, 1967).

The third and more recent perspective considers the draw-a-person test to be an expressive technique capable of revealing emotional characteristics and aspects of interpersonal relations and interaction with the environment (Koppitz, 1966, 1984; Naglieri, McNeish & Bardos, 1991). This proposal does not present a theory *a priori* to explain the meaning of each item. The idea is to provide an empirically based assessment taking the data emerging from the application of the instrument in a significant number of children as the premise. In this area, the Brazilian study of Segabinazi and Bandeira (2012) stands out, presenting evidence of validity of the global scales of evaluation of the draw-a-person test. This study suggests that the draw-a-person test can be a sensitive instrument to evaluate emotional problems and that its use seems adequate for screening processes or initial clinical evaluation (Bandeira & Arteche, 2008). Aggression, specifically, is among the emotional and behavioral aspects that can possibly be evaluated by

the draw-a-person test (Bauermann, 2012; Bauermann & Bandeira, 2014; Borsa & Bauermann, 2013; Mansur, Alves, Nakano & Ciasca, 2015).

In 1966, Koppitz developed a psychometric approach for the correction of the draw-a-person test, presenting an alternative to the qualitative analysis proposed by the psychoanalytic model of Machover. The author proposed a scale for the evaluation of aggression from indicators such as crossed eyes, presence of teeth and genitals, long arms and large hands.

Van Hutton (1994) developed a scoring system to assess the emotional problems of children victims of sexual abuse. Among the scales created by the author, one of them corresponds to the constructs "Aggression" and "Hostility", which include indicators of aggression, such as thick lines, big drawings, asymmetry between limbs, presence of teeth, clawed fingers, emphasis on facial characteristics, fingers without hands and square shoulders. Hammer (1991), on the other hand, indicated that drawings that were exaggeratedly large and decentralized on the page could also be considered as indicators of aggression.

Although scarce in the international literature, some recent studies have used the draw-a-person test in the evaluation of aggression. For example, the study by Zadeh and Malik (2009), conducted with 75 children who survived an earthquake in Pakistan may be cited. The design was evaluated from a list of 24 emotional indicators of aggression proposed by Gilbert (1980). The results showed the presence of straight lines, clawed fingers, long arms and pressed lines as more frequent in the group of aggressive children.

Sirin and Rogers-Sirin (2015) used the draw-a-person test to evaluate mental health indicators in Syrian refugee children, comparing them to a control group composed of non-refugee children regularly attending school. Indicators of aggression such as presence of blood, tears and guns were found more frequently in the drawings of the first group.

It is worth mentioning the study by Lev-Wiesel and Hershkovitz (2000), which sought to detect indicators of aggression in the draw-a-person test in 60 adult prisoners, which were divided into different groups with and without a history of violence. The results indicated a higher frequency of the following indicators in the drawings of the more aggressive group: shaded, crossed, hollowed or omitted eyes; emphasized eyebrows; thick mustaches and beards; large, pointed, clawed fingers; broad shoulders and strong posture.

In Brazil, initial studies have presented evidence that the draw-a-person test can be a sensitive instrument to evaluate aggression in children. In the study by Bauermann (2012), the drawings of 79 children aged 8 to 12 years, pupils from public schools in the metropolitan region of Porto Alegre, RS,

were evaluated, which were scored based on a list of 143 emotional indicators found in the literature. A grotesque human figure, the presence of background figures, located to the left of the page, arms close to the body, joined legs and the presence of pockets were some of the indicators that allowed groups of aggressive children to be identified. Mansur et al. (2015) compared the performance in the draw-a-person test as a cognitive measure of children with and without complaints of aggression. The results indicated statistically significant differences between the groups in the dressing item of the female figure and in the total score of the female figure, with worst performance in the group of children with complaints of aggression.

The studies of Bauermann (2012) and Mansur et al. (2015) are preliminary and new studies are necessary to investigate, in the Brazilian context, evidence of validity for the indicators of aggression in the draw-a-person test. Considering the above, the present study aimed to investigate indicators of aggression in the draw-a-person test. For this, three sequential steps were performed. In the first, a list of indicators for evaluating aggression was produced. The second step identified which of the 21 indicators were able to identify the groups high and low in aggression from the evaluation through a self-report measure. Finally, the third step investigated whether these indicators were able to distinguish children with or without a history of aggression.

Method

Step I

Participants

Participants in this study were a group of judges composed of four psychologists and four psychology students with experience in clinical psychological assessment. All of the participants were members of the Research Group on Psychological Assessment, APlab, coordinated by the author of this article and linked to the Pontifical Catholic University of Rio de Janeiro (PUC-Rio).

Procedures

Initially, a group of eight psychologists took part in a theoretical and methodological training that went through ten meetings, for a total of 20 hours. In these meetings the literature on the draw-a-person test was discussed, specifically, the theoretical bases and empirical evidence available in the literature for the interpretation of drawings. The participants also evaluated a list of indicators proposed by different authors and discussed whether they were clear and whether there were overlaps, ambiguities or redundancy among them. In addition, they suggested new items, so that, at the end of this stage, a draw-a-person test assessment list composed of 21 indicators was reached (Figure 1).

Figure 1. Items of the Aggression Indicators Scale

Line Characteristics

1. Dashed line
2. Thick line
3. Scribbles
4. Use of eraser

Drawing of the Person

5. Monster or bizarre human figure
6. Figure with theme
7. Action
8. Naked figure
9. Large figure
10. Facial expression
11. Teeth
12. Shoulders
13. Hands in fists
14. Clawed fingers
15. Emphasized arms
16. Incomplete drawing
17. Body marks
18. Crossed eyes

Elements Complementary to the Drawing of the Person

19. Phrases
20. Objects
21. Environment

After the qualitative step, a group of four judges was invited to evaluate 40 drawings elaborated by children aged 7 to 12 years. The drawings originated from a database of a study previously conducted by the author, which was submitted to and approved by the Research Ethics Committee of the State University of Rio de Janeiro (COEP/UERJ) (CAAE: 24367113.0.0000.5282). Initially, the drawings were manually selected, with the "stick" figures (i.e., one-dimensional human body), the incomplete drawings (i.e., that contained only the face or a bust), the drawings that did not identify the presence of a human figure and the drawings containing five or more indistinct human figures (i.e., which did not allow identification of the main human figure) being excluded.

Each of the judges evaluated the 40 drawings, so that all the drawings were evaluated by all the judges. For each drawing, scores were attributed according to the absence (zero points) or presence (one point) of each of the 21 indicators. Afterwards, the data were analyzed with the aim of investigating the agreement between the judges regarding the scoring of the drawings. Considering the particularities of this type of design (the same set of four judges evaluated all the participants in the sample), Fleiss' Kappa Index (κ) (a correction to Cohen's Kappa; Fleiss, 1971) and the Intraclass Coefficient (ICC) were used in a Two-way Mixed Design (Gwet, 2012), both performed in IBM SPSS 23, with a 95% Confidence Interval. The AC_1 statistic (Gwet, 2012) was also calculated in the R programming language (R Core Team, 2013), considering its relevance for comparative studies with Kappa through Monte Carlo simulations (Gwet, 2008). Since it is an analysis that originates from the Kappa statistic, the parameters for interpretation of AC_1 values are the same as those of Kappa, this practice also being endorsed by other authors as a solution in designs of studies for psychological diagnosis (Wongpakaran et al., 2013). For an exploratory analysis of the results, the criteria proposed by Landis and Koch (1977), Altman (1991) and Fleiss (1981) were used for the Kappa/AC_1 indices and Fleiss (1981) to interpret the ICC indices (Table 1).

Table 1. Parameters for interpretation of the Kappa/AC_1 and ICC values.

κ	Landis & Koch (1977)	Altman (1991)	Fleiss (1981)
< 0.0	without agreement	—	—
0.00 – 0.20	low agreement	low agreement	low agreement
0.21 – 0.40	fair agreement	fair agreement	
0.41 – 0.60	moderate agreement	moderate agreement	intermediate – good
0.61 – 0.80	substantial agreement	good agreement	
0.80 – 1.00	almost perfect agreement	very good agreement	excellent
ICC			
< 0.40	—	—	poor
0.40 – 0.75	—	—	satisfactory
≥ 0.75	—	—	excellent

Step II

Participants

A non-probabilistic community sample of 214 children, aged 7 to 13 years (*M* = 9.6; *SD* = 1.4), was used, of whom 117 were boys (54.7%). All participants were enrolled in grades two through five of public (57.9%) and private elementary schools in Rio de Janeiro. The inclusion criteria for participation in the study were: (a) to be in the stipulated age range; (b) to be a resident in the state of Rio de Janeiro; (c) to be duly enrolled in a public or private school in Rio de Janeiro; (d) to have the consent of the parents and/or guardians to participate in the study; and (e) to have the child's agreement to take part in the study.

Instruments and Procedures

All of the children were invited to draw any human figure, male or female (according to the child's preference), with no time limit, using only pencils and erasers. For the scoring of the drawings, the list of 21 indicators of aggression developed in Step I was used. Each item was scored according to its presence

(1 point) or absence (0 points). The sum of the 21 indicators yielded a general score, which could vary from 0 to 21, according to the scale score.

The Peer Aggressive Behavior Scale (PAB-S; Borsa 2016) was used to assess aggression. This is a one-dimensional self-report instrument composed of 25 closed questions to assess physical, verbal and relational aggression. Each item is evaluated by means of a four-point visual analogue scale, varying according to the frequency (never happens to always happens). The PAB-S demonstrated evidence of construct validity through factor analysis for the PAB-S with good indices of fit for a single factor structure (Borsa, 2016; Borsa & DeSousa, 2018). In the present study, the Cronbach's alpha coefficient of internal consistency for the PAB-S was 0.93.

The study was approved by the Research Ethics Committee of the State University of Rio de Janeiro (COEP/UERJ) (CAAE: 57966816.3.0000.5282). The participants were informed about the purpose of the research as well as about the data collection and analysis procedures. Personal contact with the schools was made, at which time a short copy of the research project was presented, a letter of presentation of the study and a letter of acceptance, which was signed and returned to the researchers. For the schools that agreed to participate in the study, the following procedures were performed: (1) Sending the consent form to the parents or guardians of the children. The acceptance of those responsible was a fundamental requirement for the performance of all research procedures. (2) Presentation of the research to the participants and signing of the consent form. (3) Collection of the data with the children who agreed to participate in the study, through the draw-a-person test and PAB-S. The draw-a-person test was the first instrument applied because it is understood that the drawing technique is more familiar and less anxiogenic. Afterwards, a copy of the PAB-S was given to the children. The data collection with the children was collective, in a room provided by the school, and all necessary ethical care was taken so that the procedure did not hinder or hamper the progress of regular activities.

Regarding the data analysis, descriptive statistics were initially calculated to evaluate the level of aggression of the sample according to the PAB-S scores and the draw-a-person test. To evaluate the relationship between aggression and socio-demographic characteristics, Student's t-tests (for sex and school public/private) and Pearson's correlations (for age) were conducted. Student's t-tests were also calculated for the draw-a-person test to assess differences on both sex and type of school. Finally, to evaluate whether the 21 indicators of aggression of the draw-a-person test were able to discriminate children with higher and lower levels of aggression, chi-square tests were performed.

Step III

Participants

The non-probabilistic sample consisted of 142 children, with 79 girls (55.6%), aged 7 to 13 years (M = 9.39, SD = 1.4), living in Rio de Janeiro, divided into two groups:

Clinical Group (G1): composed of 31 children (21.8%) who fulfilled at least one of the following criteria: (1) had been referred to the Psychology Service (SPA) of PUC-Rio due to a complaint of aggression; (2) had been indicated by schools in Rio de Janeiro due to recurrent episodes of aggression; or (3) had a psychiatric diagnosis of Conduct Disorder (CT) or Oppositional Defiant Disorder (ODD).

Control Group (G2): composed of 111 children (78.2%) enrolled in elementary schools, who did not have any indication or complaint of aggression and did not fit any of the criteria used for the composition of the clinical group.

Instruments and Procedures

As in Step II, the participants performed the draw-a-person test and then completed the PAB-S. The application was coordinated by psychologists and conducted together with a team of psychology students. For the Clinical Group, the application of the test was individual in a special room provided by the school or in the treatment rooms of the SPA of PUC-Rio. For the children in the Control Group, the application took place in the schools, following the same procedures mentioned in Step II. This step also was approved by the Research Ethics Committee of the State University of Rio de Janeiro (COEP/UERJ) (CAAE: 57966816.3.0000.5282). All the children had the consent of the parents and/or those responsible to take part in the study and they also agreed to participate.

Regarding the data analysis, firstly, the 142 drawings were randomly distributed to four psychologists with experience in the psychological assessment of children. These, in turn, scored each design according to the same 21 indicators produced in Step I. Next, chi-square tests were conducted to identify possible differences in the score of the indicators between the two groups (clinical and control).

Results

From the qualitative discussion carried out in Step I, a list of 21 indicators for assessing child aggression was produced. It was understood that the items present in the literature consider very specific aspects of the drawings. Therefore, as shown in Figure 1, more general items (i.e., overall analysis of the design as a whole) and more specific items (i.e., analysis of specific parts of the drawings) were incorporated.

In the quantitative stage of concordance among the judges the results indicated that items 3, 6, 7, 8, 9, 17, 19, 20 and 21 could not be estimated due to complete agreement among all the judges for all cases. Only items 2, 11, 12, 14, 15 and 18 reached satisfactory values for the three indices used (Table 2).

Table 2. Fleiss' κ, ICC and AC_1 Values

	Fleiss' κ (95% CI)	**ICC (95% CI)**	**AC_1 (95% CI)**
Item 1	0.52 (0.39 – 0.65)**	0.82 (0.70 – 0.91)**	0.53 (0.36 – 0.70)
Item 2	0.37 (0.25 – 0.50)**	0.71 (0.54 – 0.83)**	0.84 (0.73 – 0.95)**
Item 3	—	—	—
Item 4	0.19 (0.06 – 0.32)**	0.58 (0.28 – 0.76)**	0.19 (0.04 – 0.34)**
Item 5	0.10 (–0.03 – 0.20)	0.34 (–0.02 – 0.61)*	0.89 (0.81 – 0.98)**
Item 6	—	—	—
Item 7	—	—	—
Item 8	—	—	—
Item 9	—	—	—
Item 10	–0.03 (–0.15 – 0.10)	0.00 (–0.54 – 0.40)	0.95 (0.89 – 1.00)**
Item 11	0.21 (0.08 – 0.33)**	0.52 (0.23 – 0.72)**	0.96 (0.92 – 1.00)**
Item 12	0.74 (0.61 – 0.86)**	0.91 (0.87 – 0.95)**	0.95 (0.89 – 1.00)**
Item 13	–0.03 (–0.15 – 0.10)	–0.05 (–0.68 – 0.38)	0.95 (0.89 – 1.00)**
Item 14	0.21 (0.08 – 0.33)**	0.52 (0.23 – 0.72)**	0.97 (0.92 – 1.00)**
Item 15	0.49 (0.36 – 0.61)**	0.80 (0.67 – 0.88)**	0.94 (0.88 – 1.00)**
Item 16	–0.01 (–0.13 – 0.12)	0.00 (–0.62 – 0.42)	0.99 (0.96 – 1.00)**
Item 17	—	—	—
Item 18	0.45 (0.32 – 0.57)**	0.77 (0.63 – 0.86)**	0.92 (0.84 – 0.99)**
Item 19	—	—	—
Item 20	—	—	—
Item 21	—	—	—

Note:

* indicates significance at $p < 0.05$

** indicates significance at $p < 0.01$

Regarding Step II, when considering the results of the PAB-S, the sample presented a mean of 38 points (*SD* = 15.34), which can be considered low since the score could range from 25 to 125 points. The results of Student's t-test indicated that the boys (*M* = 40.8; *SD* = 15.5) presented higher levels of aggression than the girls (*M* = 35.9; *SD* = 14.8; *t* (212) = 0.57, *p* = 0.002, *d* = 0.32, 95%). The public school children (*M* = 40.7; *SD* = 16.1) presented higher levels of aggression than the private school pupils (*M* = 34.6; *SD* = 13.5; *t* (212) = 0.41, *p* = 0.004, *d* = 0.40, 95%).

In relation to the total score for the draw-a-person test, the boys (*M* = 2.18; *SD* = 1.8) presented higher scores than the girls (*M* = 1.27; *SD* = 1.1; *t* (212) = 12.50, *p* = 0.000, *d* = 0.55, 95%). There was no difference in the scores of the children from public and private schools. Finally, to evaluate whether the 21 indicators of aggression for the draw-a-person test were able to distinguish children with higher and lower levels of aggression, chi-square tests were performed. First, the children were classified into two groups according to their PAB-S score, from the following to the quartiles: Non-Aggressive Children (1st quartile, *n* = 111, 51.9%) and Aggressive Children (4th quartile, *n* = 103, 48.1%). The results indicated that only item 14 "clawed fingers" was able to distinguish the aggressive and non-aggressive groups of children (x^2 = 4.39, *df* = 1, *p* = 0.030).

Finally, in Step III, the results showed that four items were able to distinguish the clinical and control groups, that is, they were represented more often in the clinical group than in the control group. The items "presence of scribbles in any part of the drawing" (x^2 = 10.97, *df* = 1, *p* = 0.001), "presence of teeth in the human figure" (x^2 = 9.18, *df* = 1, *p* = 0.002) and "exaggerated shoulders in the human figure" (x^2 = 11.35, *df* = 1, *p* = 0.003) were more frequent in the clinical group. The item "thicker line in any part of the drawing" was more frequent in the control group than in the clinical group (x^2 = 7.36, *df* = 1, *p* = 0.007).

Discussion

In Step I, a group of experts produced a list of 21 items for assessing aggression in the draw-a-person test. This stage provided an understanding of the theoretical and empirical bases that permeate the evaluation of the draw-a-person test. It was understood that the items present in the literature consider very specific aspects of the drawings. Therefore, new items were incorporated, in order to allow for a more global analysis of the drawing, considering not only the human figure itself, but the context in which it is graphically represented.

As far as Step I was concerned, the results of the evaluation of concordance among the judges indicated that nine of the 21 items (3, 6, 7, 8, 9, 17, 19, 20 and 21) could not be estimated because they presented complete agreement among all judges for all the cases. This phenomenon, known as the "Kappa paradox" (Cicchetti & Feinstein, 1990), usually occurs in cases of psychopathological diagnoses, in which the Kappa value is low, even if interpretively there is very high concordance among the judges. Di Eugenio and Glass (2004) also present problems in the use of Kappa statistic: notable situations of abnormal statistical behavior stand out when there is a prevalence problem (with an asymmetrical distribution among the possible categories) and the bias problem (the degree to which the judges disagree with each other).

This is problematic for different reasons: for the Kappa and AC_1 statistics, which measure inter-judge concordance (agreement among different judges) for a sample, there would also have to be intra-judge variability (the same judge would vary in the type of assessment given) and this did not occur for the nine items mentioned. These assumptions are necessary so that the judge is unbiased (the judge is only able to evaluate one category) or in the sample (in the population there are only individuals of one of the categories) and no agreement by chance (there is a small chance that the judges randomly agree, although Kappa and AC_1 make corrections for this phenomenon, some variability is necessary).

For the ICC statistic, there would have to be variation between the categories (the sample would present individuals of all categories and with all types of response to the stimulus), which did not happen in the nine items described. This assumption is important because it ensures that there is no bias in the sample. Thus, the same nine items could not be measured for their reliability by means of concordance estimates, since there was no level of variance and disagreement for the cases. These items should therefore be used with caution because they may have difficulties in describing the phenomenon evaluated throughout its possible continuum, characterized in this case by the absence of a clear trait. In order to be able to verify whether these items are reliable, there must also be individuals whose drawings indicate the presence of aggression and judges who are able to distinguish this presence clearly and objectively.

Following the AC_1 criterion, only item 4 should be excluded. This item is about the use of an eraser, which may not have been adequately assessed, since the drawings were not made in the presence of the judges. Considering the inconclusive results in relation to the parameters evaluated, it was decided to maintain all 21 items for the performance of Step II.

In Step II, low levels of aggression were verified in the total sample. Corroborating the findings of the literature (Björkqvist, 2018), the boys presented higher levels of aggression than the girls; however, the effect size of this difference was small. Differences, with a moderate effect size, were found among students from public and private schools, with the former having higher levels of aggression. This is supported by two previous studies conducted by the author (Borsa & Nunes, 2011; Borsa, Souza & Bandeira, 2011). Specifically regarding the draw-a-person test, the boys presented higher scores than the girls, with the effect size of this difference being moderate, confirming the results obtained through the PAB-S.

In relation to the evaluation of the 21 items of the draw-a-person test, the results indicated that only the item "clawed fingers" was able to distinguish the aggressive and non-aggressive groups of children. This item is reported in the literature (Gilbert, 1980; Lev-Wiesel & Hershkovitz, 2000) as frequently seen in the drawings of aggressive children.

It is important to note that this stage was conducted in a community sample, in which the children presented low levels of aggression according to the score derived from a self-report measure. Therefore, this result may be associated with social desirability, that is, when the answers to certain test items are based on the person's judgment of which answer is the most socially appropriate, and not on their own behavior. It can be inferred that the choice of a self-report measure may not have been the most appropriate to distinguish children with high and low levels of aggression (Borsa & Muniz, 2016).

Finally, in Step III the results presented differences in relation to Step II, since other items were able to distinguish children with high and low levels of aggression. For example, the item "exaggerated shoulders in the human figure" was more frequent in the clinical group than in the control group, corroborating the results of the study by Lev-Wiesel and Hershkovitz (2000). Likewise, the item "presence of teeth in the human figure" was also more frequent in the clinical group, as suggested by Van Hutton (1994).

The item "thick line in any part of the drawing" was more frequent in the control group, which contradicts the data obtained by Zadeh and Malik (2009). In addition, the item "clawed fingers" was not drawn by any child of the sample. It should be mentioned that the majority of the children drew hands without fingers, in the form of gloves, or even human figures with hands omitted (in the pockets or behind the back). This type of drawing pattern is common, given the difficulty of reproducing this item graphically. The omission or simplification of the hands may also be associated with social desirability, since the children showed

concern about making beautiful drawings. Thus, the questions arise again: Is the inconsistency of the results between the two studies explained by the criterion for defining the groups? Or, would social desirability have greatly interfered with the performance of the children in the tests?

Final Considerations

The general aim of the present study was to produce a list of indicators of aggression for the draw-a-person test and to investigate their validity. This sought to fill an important gap in Brazil regarding the psychological evaluation of children with aggression. As previously explained, the main way of evaluating such behavior is through standardized psychometric instruments that present important limitations regarding social desirability and expectations. It is believed, therefore, that the draw-a-person test carried out by the child can be a complementary tool in the identification of aggression.

The results of the present study were inconclusive and only indicate the direction of future studies that may improve on the list of indicators hereby proposed, expanding and revising it. Studies with other samples and with clinical groups constituted through more robust criteria (e.g., with a diagnosis of conduct disorder) could elucidate whether the indicators are, in fact, accurate in identifying aggression in children. Among the limitations, it should also be mentioned that the evaluation of the drawings may have been hampered by the fact that the judges did not take part in the data collection. Observation when the drawing is made can help in the comprehension of indicators such as the use of an eraser. Finally, considering the potential of drawings in the investigation of children's emotional problems, it is understood that the information found in national and international literature still show a dearth of empirical studies that reinforce the validity of indicators of aggression in children.

References

Altman, D. G. (1991). Mathematics for Kappa. In: Altman, D. (ed.). *Practical Statistics for Medical Research*, p. 406-407. London: Chapman & Hall.

Bandeira, D. R.; Arteche, A. X. (2008). Desenho da Figura Humana. In: Villemor-Amara, A. E. de; Werlang, B. S. G. (org.). *Atualizações em Métodos Projetivos para Avaliação Psicológica*, p. 205-119. São Paulo: Casa do Psicólogo.

Bauermann, M. (2012). *Indicadores de agressividade através do Desenho da Figura Humana* (monografia de especialização – Instituto de Psicologia, Universidade Federal do Rio Grande do Sul). Porto Alegre (não publicada).

Bauermann, M.; Bandeira, D. R. (2014). Indicadores de Comportamentos Agressivos através do Desenho. In: Borsa, J. C.; Bandeira, D. R. (org.). *Comportamento agressivo na infância: da teoria à prática*, 1ª ed., p. 361-376. São Paulo: Pearson.

Björkqvist, K. (2018). Gender differences in aggression. *Current Opinion in Psychology*, *19*, 39-42.

Borsa, J. C. (2016). Development and refinement of the Peer Aggressive Behavior Scale – PAB-S. *Psychology: Research and Review*, *29*(19).

Borsa, J. C.; Bandeira, D. R. (2014). Uma breve introdução ao tema dos comportamentos agressivos. In: Borsa, J. C.; Bandeira, D. R. (ed.). *Comportamento agressivo na infância: da teoria à prática*, p. 9-23. São Paulo: Casa do Psicólogo.

Borsa, J. C.; Bauermann, M. (2013). O Desenho da Figura Humana na avaliação da agressividade infantil. *Avaliação Psicológica*, *12*(2), 273-274.

Borsa, J. C.; DeSousa, D. (2018). Invariância de medida e evidências de validade externa da Peer Aggressive Behavior Scale (PAB-S). *Psico*, *49*(2), 178-186.

Borsa, J. C.; Muniz, M. (2016). Testagem Psicológica de Crianças e Adolescentes. In: Hutz, C. S.; Bandeira, D. R.; Trentini, C. M.; Krug, J. S. (org.). *Psicodiagnóstico*, 1ª ed., v. 1, p. 238-246. Porto Alegre: Artmed.

Borsa, J. C.; Nunes, M. L. T. (2011). Prevalência de Problemas de Comportamento em uma amostra de crianças em idade escolar da cidade de Porto Alegre. *Aletheia*, *34*, 32-46.

Borsa, J. C.; Souza, D. S.; Bandeira, D. R. (2011). Prevalência dos Problemas de Comportamento em uma amostra de crianças do Rio Grande do Sul. *Psicologia: Teoria e Prática*, *13*(2), 15-29.

Buck, J. (2003). *HTP: Manual e guia de interpretação*. Cubatão: Vetor.

Bushman, B. J.; Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy?. *Psychological Review*, *108*(1), 273-279.

Cicchetti, D. V.; Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology*, *43*(6), 551-558.

Di Eugenio, B.; Glass, M. (2004). The Kappa Statistic: A second look. *Computational Linguistics*, *30*(1), 95-101.

Evans, S. C.; Fite, P. J. (2019). Dual Pathways from Reactive Aggression to Depressive Symptoms in Children: Further Examination of the Failure Model. *Journal of Abnormal Child Psychology*, *47*(1), 85-97.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, *76*(5), 378.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*, 2nd ed., p. 216-218. New York: John Wiley and Sons.

Gilbert, J. (1980). *Interpreting Psychological Test Data*, vol. II, p. 13-63. New York: Nostrand Reinhold Company.

Goodenough, F. (1974). *Test de inteligência infantil por medio del dibujo de la figura humana*, 7ª ed. Buenos Aires: Paidós.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29-48.

Gwet, K. L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters*. Advanced Analytics, LLC.

Hammer, E. F. (org.) (1991). *Aplicações clínicas dos desenhos projetivos*. Rio de Janeiro: Interamericana.

Koppitz, E. M. (1966). Emotional indicators on Human Figure drawings of children: a validation study. *Journal of Clinical Psychology*, *22*(3), 313-315.

Koppitz, E. M. (1984). *El dibujo de la figura humana en los niños*. Buenos Aires: Guadalupe.

Landim, I.; Borsa, J. C. (2017). Revisão sistemática sobre programas de intervenção para redução de comportamentos agressivos infantis. *Contextos Clínicos*, *10*(1), 110-129.

Landis, J. R.; Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159.

Lev-Wiesel, R.; Hershkovitz, D. (2000). Detecting violent aggressive behavior among male prisoners through the Machover Draw-A-Person Test. *The Arts in Psychotherapy*, *27*(3), 171-177.

Machover, K. (1967). O traçado da Figura Humana: um método para o estudo da personalidade. In: Anderson, H. H.; Anderson, G. L. (org.). *Técnicas projetivas do diagnóstico psicológico*, p. 345-370. São Paulo: Mestre Jou.

Mansur, C. M.; Alves, R. J. R.; Nakano, T. C.; Ciasca, S. M. (2015). O teste do desenho da figura humana em crianças com e sem queixas de agressividade: estudo piloto. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, *15*(1), 8-21.

McQuade, J. D.; Breaux, R. P.; Gómez, A. F.; Zakarian, R. J. (2016). Biased Self-Perceived Social Competence and Engagement in Subtypes of Aggression: Examination of Peer Rejection, Social Dominance Goals, and Sex of the Child as Moderators. *Aggressive Behavior*, *42*(5), 498-509.

Naglieri, J.; McNeish, T.; Bardos, A. (1991). *Draw-a-Person: Screening procedure for emotional disturbance*. Austin, Texas: ProEd.

Petersen, C. S.; Souza, M. A. M.; Wessel, E. (2014). Intervenções baseadas em evidências: Triple P e Programa Anos Incríveis. In: Borsa, J. C.; Bandeira, D. R. (org.). *Comportamento agressivo na infância: da teoria à prática*, 1ª ed., p. 491-506. São Paulo: Casa do Psicólogo.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Robertson, K.; Forbes, S.; Thyne, M. (2017). Perpetration of Alcohol-Related Aggression by Male and Female College Students. *Journal of Interpersonal Violence*. https://doi.org/10.1177/0886260517696872

Rocha, M. M.; Emerich, D. R.; Silvares, E. F. M. (2014). Avaliação de comportamentos agressivos utilizando a perspectiva de diferentes informantes. In: Borsa, J. C.; Bandeira, D. R. (org.). *Comportamento agressivo na infância: da teoria à prática*, 1ª ed. São Paulo: Casa do Psicólogo.

Santo, B. J.; Bass, E. C.; Stella-Lopez, L.; Bukowski, W. M. (2017). Contextual influences on the relations between physical and relational aggression and peer victimization. *School Psychology International*, *38*(1), 42-59.

Segabinazi, J.; Bandeira, D. R. (2012). Desenho da Figura Humana para avaliação emocional de crianças: evidências de validade de escalas globais. In: Wechsler, S. M.; Nakano, T. C. (org.). *O desenho infantil: forma de expressão cognitiva, criativa e emocional*. São Paulo: Casa do Psicólogo.

Sirin, S. R.; Rogers-Sirin, L. (2015). *The Educational and Mental Health Needs of Syrian Refugee Children*. Washington, DC: Migration Policy Institute.

Suehiro, A. C. B.; Benfica, T. de S.; Cardim, N. A. (2016). Produção Científica sobre o Teste Desenho da Figura Humana entre 2002 e 2012. *Psicologia: Ciência e Profissão*, *36*(2), 439-448.

Van Hutton, V. (1994). *House-Tree-Person and Draw-a-Person as measures of abuse in children: A quantitative scoring system*. Florida: Psychological Assessment Resources.

Wongpakaran, N.; Wongpakaran, T.; Wedding, D.; Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology*, *13*(1), 61.

Zadeh, Z. F.; Malik, S. M. (2009). Expression of Aggressive Tendencies in the Drawings of Children and Youth Who Survived the Northern Pakistan Earthquake. *Europe's Journal of Psychology*, *5*(2), 64-81.

Recebido em 07 de abril de 2019
Aceito para publicação em 28 de maio de 2019