OPEN ACCESS

ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA
PUCRS

PSICO

Psico, Porto Alegre, v. 54, n. 1, p. 1-12, jan.-dez. 2024
e-ISSN: 1980-8623 | **ISSN-L:** 0103-5371

http://dx.doi.org/10.15448/1980-8623.2024.1.41815

SEÇÃO: ARTIGO

Draw-a-Person Test: correction Using Reduced Criteria

Desenho da Figura Humana: correção por meio de Critérios Reduzidos
Dibujo de la Figura Humana: corrección por medio de Criterios Reducidos

Fernanda Otoni[1]
orcid.org/0000-0002-9347-7144
fernanda.otoni@utp.br

Acácia A. Angeli dos Santos[2]
orcid.org/0000-0002-8599-7465
acacia.angeli@gmail.com

Fabian Javier Marin Rueda[3]
orcid.org/0000-0001-5173-0802
marin.rueda@gmail.com

Recebido em: Sep, 22, 2021.
Aprovado em: Sep, 11, 2022.
Publicado em: Oct, 15. 2024.

Abstract: The Draw-a-Person test is a screening instrument used to assess children's non-verbal intelligence. In the assessment of children, brief criteria of DAP's could be promising to identify possible learning difficulties. Therefore, our aim was to propose a reduced number of criteria to evaluate DAP. To do so, 1403 children aged from 6 to 10 years (M = 8.16; SD = 1.40) participated. Through the mirt package in the R Studio program, we proposed a 2-parameter logistic model, extracting 12 indicators that were able to predict 83% of non-verbal intelligence. In this correction proposal, we selected indicators that maintained easy, medium, and difficult levels, which allowed children's performance to differentiate according to age. The results were positive, indicating that the proposal is promising for use in research, clinical practice, psychoeducation, and educational psychological assessments.
Keywords: item response theory, psychological assessment, school performance, screening instrument

Resumo: O Desenho da Figura Humana é um instrumento de rastreio utilizado para avaliar a inteligência não verbal infantil. Acredita-se que no processo de avaliação critérios breves de correção possibilitariam prever possíveis déficits intelectuais. Assim, objetivou-se propor um número de critérios reduzidos para avaliar tal instrumento. Para tanto, participaram 1403 crianças com idade entre seis e 10 anos (M = 8,16; DP = 1,40). Por meio do pacote mirt no programa R Studio, utilizou-se um modelo logístico de dois parâmetros, extraíram-se 12 indicadores que demonstraram ser capazes de predizer 83% do construto avaliado. Nesta proposta de correção, foram selecionados indicadores que mantiveram os níveis de dificuldade fácil, médio e difícil, que foram capazes de diferenciar o desempenho das crianças em função da idade. Os resultados foram positivos e apresentaram bons índices psicométricos, indicando que a proposta é promissora para ser utilizada não apenas na pesquisa, mas também na prática clínica e na avaliação psicológica escolar educacional.
Palavras-chave: teoria de resposta ao item, avaliação psicológica, desempenho escolar, instrumento de rastreio

Resumen: El dibujo de la figura humana es un instrumento de rastreo que evalua la inteligencia no verbal de los niños. En el proceso de evaluación, una cantidad menor de criterios de corrección permitirían predecir posibles déficits intelectuales. El objetivo fue proponer un número reducido de criterios para evaluar este instrumento. Participaron 1403 niños con seis y 10 años (M = 8,16; DP = 1,40). A través del paquete mirt del programa R Studio, utilizamos el modelo logístico de 2 parámetros, extrayendo 12 indicadores que demostraron ser capaces de predecir 83% del constructo evaluado. En esta propuesta de corrección se seleccionaron indicadores que mantuvieron los niveles de dificultad fácil, medio y difícil, los cuales fueron capaces de diferenciar el desempeño de los niños según la edad. Los resultados fueron positivos y presentaron buenos índices psicométricos, lo que indica que la propuesta es prometedora para la práctica clínica y evaluación psicológica educativa.
Palabras-clave: teoría de la respuesta al ítem, evaluación psicológica, el rendimiento escolar, instrumento de rastreo

1 Universidade Tuiuti do Paraná (UTP), Curitiba, PR, Brasil.
2 Pesquisador Autonômo, Campinas, SP, Brasil.
3 Universidade São Franciso (USF), Campinas, SP, Brasil.

During child development, drawing tends to reflect the neurological maturational process through which children express their social, affective, and intellectual experiences (Tükel et al., 2018). As it is a basic and universal language, it came to be used as an important tool in the psychological assessment process. In addition to preceding verbal language, it indicates the ability to communicate and helps develop cognitive and emotional skills. Drawing is also one of the playful resources that favor social interaction, exchange of experiences, and the possibility of learning, as it is a spontaneous way of expressing and representing what transits between the real and the imaginary (Silva et al., 2015).

Based on different theoretical perceptions and distinct areas of psychological processes, drawing has been used as a systematic tool that allows us to identify and comprehend the structure and functioning of human thought (Arteche & Bandeira, 2006). Since the beginning of the 20th century, some researchers have sought to analyze the relationship between the aptitude for drawing and intellectual capacity (Claparède, 1907; Rouma, 1913; Luquet, 1927). Psychologist Florence Goodenough (1926) pioneered the development of a systematic method with correction criteria to assess a drawing of a human figure.

Goodenough (1926) asked children to reproduce only a drawing of a man. Therefore her correction system came to be known as the Draw-a-Man Test. Goodenough (1926) chose this type of task considering that the human figure would be equally familiar to all children, as well as presenting the least possible variability in its essential characteristics. Based on the drawings of 900 children aged between 4 and 10 years, enrolled between the first and fourth grades, Goodenough (1926) created 51 correction indicators, which were scored through the absence or presence of the representation of the eyes, mouth, nose, arms and legs, neck, and hair, among other aspects. The author found that older children produced better quality drawings, which indicated that the instrument would be useful in childhood cognitive assessments.

In a recent study, Rueda et al. (2020) carried out a systematic literature review in order to qualitatively analyze the results of studies that used the Draw-a-Person Test (DAP) correction systems as a cognitive measure. Based on the reading of 33 articles, the authors found no studies that refuted the evaluation and correction criteria proposed by Goodenough. Furthermore, scholars have found that, almost 100 years later, this system remains capable of assessing intelligence and considers the evolutionary aspect of the conceptual repertoire (Araújo & Fernandes, 2015; Bandeira et al., 2012; Picard, 2015). However, by understanding that drawing allows multiple interpretations, different methods have been developed to assess emotional aspects (Koppitz, 1968; Naglieri, 1988), personality (Machover, 1949), intelligence (Harris, 1991; Koppitz, 1968; 1984; Naglieri, 1988; Sisto, 2005; Wechsler, 2003) and creativity (Oliveira & Wechsler, 2016).

Regarding the assessment of intelligence, specifically, several application and correction systems have been developed. Some authors started to request the reproduction of two (Koppitz, 1968; Wechsler, 2003) or even three figures (Harris 1963; Naglieri, 1988), while others requested only one human figure (Sisto, 2005). Regarding the correction criteria, these also differed in the number of indicators, ranging between 30 and 73 for each figure requested. After these different correction proposals for the DAP, some studies aimed to verify whether there was, among these different forms of application, one with greater sensitivity to differentiate childhood development or more correlated with instruments that assess intelligence (Campbell & Bond, 2017; Flores-Mendoza et al., 2010; Rosa, 2008).

Rosa (2008) used the Harris and Koppitz methods to verify whether both, in fact, would correspond to a measure of intelligence. For this purpose, the author applied the DAP and the R-2 test to 1540 children between 5 and 11 years of age. The results indicated significant correlations (r = 0.80 to r = 0.90) between the measures. Furthermore, the two correction methods could differentiate performance according to age. In another study, Flores-Mendoza et al. (2010) asked 107 children aged 7, 9, and 11 years to reproduce a human drawing, which was corrected using the criteria of Goodenough, Harris, and Wechsler. The authors found significant correlations, with coefficients ranging from $r = 0.52$ to $r = 0.85$, indicating that there is a strong similarity between these correction criteria.

Campbell and Bond (2017), in turn, used the analysis of the Rasch model to assess the psychometric properties of the figures reproduced by 246 children aged between 4 and 10 years. The authors found that the errors made in one drawing were repeated in the others, indicating that the creation of the three figures was unnecessary and that only one would be sufficient for the assessment of intellectual maturity. It should be noted that Koppitz (1973) herself indicated that the use of three drawings adds little information about the child's intellectual capacity, and therefore, it would not justify the child's time and effort.

The need to propose different correction methods has also been considered, aiming to decrease the time of application of the instrument, as well as the degree of complexity of the correction criteria, since the instrument must be used as a tool that aims to facilitate and simplify the measurement of a given construct. Furthermore, both national and international studies that have used the DAP have indicated that this is an instrument for screening child cognitive development and that it should not be considered a single measure to assess intelligence but to hypothesize possible intellectual difficulties (Bartholomeu et al., 2012; Flores-Mendoza et al., 2010; Panesi & Morra, 2017; Rehrig & Stromswold, 2017; Rosa & Alves, 2014; Rosa-Neto et al., 2013; Tükel et al., 2018).

In line with Streiner's (2003) considerations that the psychological tests used for screening different cognitive skills should be simple and accurate, this study aimed to propose a reduced number of indicators for correcting the DAP. It is believed that a system with fewer correction criteria could serve as a useful screening tool to be used in assessments with many children, in addition to contributing to the prompt collection of results. In addition, it could assist in the identification of learning difficulties at the beginning of schooling and enable the creation of intervention strategies that minimize possible intellectual impairments.

Method

Participants

Study participants were 1403 children, 713 girls (51%) and 690 boys (49%), aged between 6 and 10 years, $M = 8.16$ and $SD = 1.40$. All are regularly enrolled in the public-school network between the 1st and the 5th year of Elementary Education.

Instruments

Draw-a-Person Test - Sisto Scale (Sisto, 2005). To carry out this study, it was decided to apply and correct the DAP based on the Sisto system. The choice was guided by the fact that it is specific to the Brazilian context and because it requires the production of only one human figure. Although this system is the simplest one, it can be identified that some of the 30 indicators are more complex and require more time for correction. In the application instructions, the child is asked to draw a human figure with as much detail as possible. Sisto (2005) determined that the drawing should only be corrected if it has the head, arms, and legs, and once it meets this criterion, 30 indicators will be observed that follow a pattern of difficulty. For the correction, one point is attributed to the presence of each of these items and zero to the absence, and the general score is obtained by the total sum of the items.

Procedures

Data collection. After the delivery and return of the parent's consent form and the signing of the consent form by the student, data collection was carried out collectively in a single application of approximately 15 minutes. The classrooms had an average of 25 children who received a blank sheet of paper and a black pencil. Initially, they were asked to write their name, age, and school year on the sheet and then to start producing the drawing. The drawings of students who were not aged between 6 and 10 years were excluded.

Data analysis. Through the R Studio program, we used the dplyr package to perform descriptive and inferential sample analysis. The Rasch, 2 PL, 3 PL, and 4 PL models were evaluated using the mirt package (Chalmers, 2012). The comparison of these models considered the Chi-Square (^{2}) and degrees of freedom <0.05; Root-Mean-Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) <0.10; Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) ≥0.90; Akaike Criterion Index (AIC) and Bayesian Information Criterion (BIC), in which lower values are considered ideal (Marôco, 2014). Using the Kuder-Richardson alpha, the reliability indices of these sets of indicators were evaluated, then linear regression analysis was used, considering the 30 indicators of the DAP as an independent variable and the sets of 15 and 12 indicators, which we obtained through Item Response Theory, as dependent variables. The analysis of variance (ANOVA) made it possible to verify whether the reduced correction criteria were able to differentiate the children's performance according to age.

Ethical Considerations

After authorization from the schools, we submitted the study to the Research Ethics Committee of **HIDDEN** in accordance with Resolution 196/96 of the National Health Council (CAAE: **HIDDEN),** and after approval, we began the data collection.

Results

Initially, we found that the total score of the DAP - Sisto Scale presented a minimum score was 0 and a maximum of 30 points, with a mean of 12.08, *SD* = 5.98. Sisto (2005) proposed that the indicators used to correct the DAP followed a hierarchical order of difficulty. Accordingly, for those that were part of the easy reproduction level, the score ranged between 0 and 11 points, *M* = 6.64, *SD* = 2.40. For the indicators with medium difficulty, the minimum score was 0 and the maximum 10 points, with a mean of 3.09 and a standard deviation of 2.37. Finally, for the indicators with a higher difficulty level, the score also ranged between 0 and 10 points, with *M* = 2.34, *SD* = 2.01. Subsequently, it was confirmed, through the variance in the first contrast less than 2, that the DAP has one dimension (*F* = 1.84).

Based on the suggestions of Primi et al. (2018) that studies carried out with the DAP should freely estimate the discriminative capacity of the items, we decided to verify whether there would be differences between the fit indices due to the analyses of the Rasch model and the 2, 3 and 4 parameter (PL) logistic models. We present these results in Table 1.

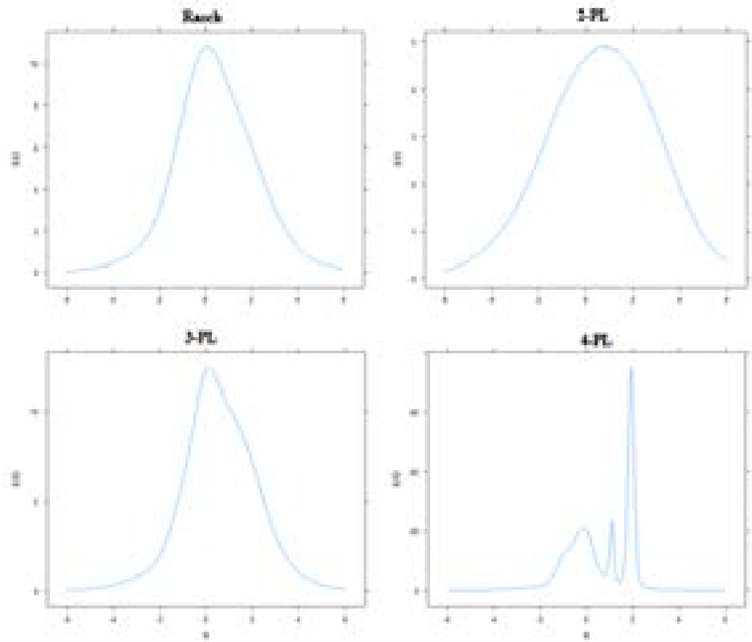
Table 1 – Fit indices of Item Response Theory models

Model	M^2	*df*	*p*	RMSEA	SRMSR	CFI	TLI
Rasch	4339.108	435	0	0.08	0.08	0.86	0.85
2-PL	3692.367	405	0	0.07	0.06	0.87	0.88
3-PL	3154.519	375	0	0.07	0.06	0.88	0.90
4-PL	2581.519	345	0	0.06	0.06	0.90	0.92
Differences between models			**AIC**	**BIC**	X^2	*df*	*p*
Rasch x 2-PL			38584.39	38899.17	581.55	29	0.01
2-PL x 3-PL			38588.09	39060.93	56.30	30	0.03
2-PL x 4-PL			38392.36	30021.93	312.03	60	0.01

It should be noted that in the Rasch Model, the discriminative capacity of the item/indicator is the same for all people that respond to the test. The 2PL model, in turn, considers how much the item/indicator can discriminate the respondents at different skill levels. The 3PL model also assesses random correct responses, while errors due to carelessness are considered in the 4PL model (De Ayala, 2008). The results presented in Table 1 indicated that the fit of the RMSEA and SRMSR presented acceptable values for all models, while only the 4PL model was adequate in relation to the CFI and TLI (Marôco, 2014). Furthermore, the M^2 and degrees of freedom decreased from one model to another, suggesting a slight improvement along these models' fit indexes. However, it should be considered that in the DAP, there is no correct response by chance or due to carelessness since the proposed task requires a series of indicators that underlie a specific part of the human body, for example, to add an earring is expected that the child reproduces the ear.

In this case, a simple logistic model would be the most suitable for identifying the internal structure of the DAP (Primi et al., 2018; Streiner, 2003). In line with this consideration, the comparison between the estimated models indicated that the 2PL model is superior to the others, suggesting that it is the most explanatory for the type of activity proposed by the DAP. When comparing one model with the other, in the 2PL model, the ACI and BIC indices showed a significant decrease in relation to the others. Furthermore, the information curve presented in Figure 1 demonstrated that the 2PL model tended to be more informative.

Figure 1 – Test information curves for each of the observed models



Based on the results obtained up to this point, we decided to use the 2PL model to propose the reduced criteria for correcting the DAP. Table 2 presents the levels of difficulty and discrimination of the DAP-Sisto Scale indicators, as well as the quality of the fit indices.

Psico, Porto Alegre, v. 54, n. 1, p. 1-12, jan.-dez. 2024 | e-41815

Table 2 – Parameter of the difficulty of endorsement of the Sisto Scale indicators and fit indices

Indicators	Difficulty	Discrimination	S-X²	*df*	*p - X²*	*r*item-theta
1 Mouth	-2.78	1.29	13.05	14	0.52	0.26
2 Nose	-1.16	0.57	31.42	23	0.12	0.34
3 Arms/Legs	**-0.69**	**1.61**	**12.96**	**19**	**0.84**	**0.55**
4 Clothes	-1.13	2.24	58.31	14	0.02	0.50
5 Neck	**-0.59**	**1.24**	**30.68**	**20**	**0.59**	**0.52**
6 Trunk	-0.70	1.49	14.71	19	0.74	0.52
7 Motor coordination	-0.98	1.49	56.02	18	0.01	0.45
8 Two/three garments	**-0.35**	**2.66**	**47.62**	**16**	**0.54**	**0.63**
9 Legs	**-0.32**	**0.89**	**14.60**	**21**	**0.84**	**0.42**
10 Feet	**0.89**	**1.34**	**25.14**	**21**	**0.24**	**0.51**
11 Integrated neck	**0.85**	**0.84**	**93.02**	**23**	**0.21**	**0.40**
12 Legs/Arms position	0.06	1.75	20.21	19	0.38	0.60
13 Pupils	0.51	0.93	39.72	23	0.01	0.45
14 Shoulders	**1.03**	**1.89**	**28.61**	**19**	**0.72**	**0.54**
15 Hair	0.06	1.01	36.42	23	0.03	0.35
16 Arms	0.13	1.21	16.67	21	0.73	0.49
17 Fingers	**1.77**	**0.93**	**27.72**	**23**	**0.22**	**0.37**
18 Eyes	**1.18**	**1.45**	**27.53**	**21**	**0.15**	**0.51**
19 Arm/elbow/shoulder	**1.94**	**1.18**	**43.08**	**22**	**0.46**	**0.64**
20 Arm/Leg Contour	0.23	2.13	30.55	19	0.04	0.35
21 Trunk contour	0.34	2.51	28.10	16	0.03	0.61
22 Physiognomic features	**2.50**	**1.34**	**33.08**	**19**	**0.23**	**0.64**
23 No transparency	0.75	0.90	30.83	23	0.12	0.40
24 Coherent dress	-0.55	1.82	29.29	18	0.04	0.48
25 Head contour	0.87	1.53	36.02	21	0.02	0.50
26 Mouth and nose	**3.02**	**0.97**	**19.36**	**21**	**0.56**	**0.27**
27 Four pieces of clothing	1.57	1.53	42.41	20	0.01	0.42
28 Thumb	**1.69**	**1.48**	**44.96**	**20**	**0.11**	**0.42**
29 Chin	**1.92**	**1.59**	**30.70**	**18**	**0.31**	**0.38**
30 Ears	**1.85**	**1.29**	**8.87**	**21**	**0.90**	**0.39**

We observed that 9 of the 30 correction indicators (5, 7, 13, 15, 20, 21, 24, 25, and 27) presented *p* values less than 0.05 for the S-X^2 index, suggesting that these did not fit the model. Regarding the discrimination parameters, Nakano et al. (2015) suggest that the variation in values ranges between 0 and 3; in the present study, the values ranged from 0.57 (low) to 2.51 (very high). Since the aim of the present study was to propose a version with reduced correction criteria for the DAP, it was decided to select the 15 indicators with greater difficulty to be reproduced (half of what was proposed by Sisto, 2005) with 5 of each hierarchical level. In this selection process, only the indicators that presented adequate S-X^2 values and item-theta correlations greater than 0.30 were considered.

Bond and Fox (2015) suggest that the difficulty values of the items should be located between – 3 and + 3, and the higher the value, the better the difficulty level of the indicators. From this perspective, we found that the easy level indicators showed a trend towards negative values. Therefore, those with values closer to 1 were prioritized, selecting arms/legs in two positions, neck, two/three garments, legs, and feet (the latter was the only positive value indicator). For the indicators of medium level of difficulty, the ones with the highest value were chosen (integrated neck, shoulders, fingers, eyes, and arms with elbow and shoulder). For the difficult level indicators, the physiognomic features, mouth, and nose in two dimensions, thumb, chin, and ears, were considered.

In the DAP correction process, some indicators had overlapping criteria, such as arms and legs in two positions that would already be evaluated by indicators 9 and 19 (legs and arms, elbows and shoulders, respectively), and the neck, which appears in the easy level (integrated neck; medium difficulty level). As well as the mouth and nose in two dimensions that make up the physiognomic features, a point is assigned if the child reproduces the eyes, mouth, and nose in two dimensions. Therefore, arms and legs in two dimensions (-0.70) were replaced by the trunk (-0.70) since the values were the same. Furthermore, we removed three indicators from each hierarchical level, namely, neck, shoulders, and mouth and nose since these indicators are assessed by physiognomic features, and arm/elbow/shoulder. After these changes, the proposal to correct the DAP through reduced criteria had a set of 12 indicators. To verify whether the suggested indicators presented good fit indices, we compared between 30, 15, and 12 indicators (Table 3).

Table 3 – Comparison of scales according to the number of indicators and fit indices

Model (2-PL)	M^2	*df*	*p*	RMSEA	SRMSR	CFI	TLI
30 indicators	3692.37	405	0	0.076	0.06	0.87	0.88
15 indicators	1024.66	90	0	0.086	0.08	0.85	0.82
12 indicators	332.23	54	0	0.060	0.05	0.91	0.89

Indicators	Difficulty	Discrimination	S-X^2	*df*	*p* - X^2	$r^{item\text{-}theta}$
1 Trunk	-0.75	1.34	7.12	8	0.52	0.60
2 Two/three garments	-0.42	1.71	20.25	9	0.16	0.66
3 Legs	0.31	0.92	3.83	8	0.87	0.51
4 Feet	0.82	1.56	30.58	7	0.07	0.55
5 Integrated neck	0.91	0.77	23.64	8	0.26	0.44
6 Fingers	1.48	1.21	13.21	8	0.10	0.42
7 Eyes	1.12	1.58	18.05	9	0.34	0.53
8 Arm/elbow/shoulder	2.38	0.89	34.34	7	0.14	0.35
9 Physiognomic features	2.25	1.59	13.53	9	0.13	0.47
10 Thumb	1.50	1.88	5.99	7	0.54	0.45
11 Chin	1.84	1.72	37.78	8	0.08	0.45
12 Ears	1.79	1.37	9.53	8	0.29	0.44

The results presented in Table 3 indicated that the proposal to maintain the DAP with 12 criteria had an internal structure with a better fit. The values of the S-X^2, *df* and *p* for the S-X^2 parameter demonstrated that all the indicators fit within the model (Bond & Fox, 2015). When analyzing the distribution of the indicators in the theta scale, we verified that many children had a lower theta level of ability to reproduce arms, elbows and shoulders, physiognomic features, and ears. Most of the children found it easier to correctly reproduce the indicators with easy and medium levels of difficulty.

To verify how much the reduced indicators could predict the variance of the total score of the DPA-Sisto Scale, we performed a linear regression analysis, in which it was shown that the set of 15 indicators predicts 85% (adjusted R^2 = 0.852), while the set of 12 predicts 83% (adjusted R^2 = 0.832). After verifying that the difference in the predictive power of both proposals was minimal, we used the set of 12 indicators to calculate the internal consistency value through the Kuder-Richardson coefficient. We observed a $KR\text{-}_{20}$ = 0.77 for this reduced version and $KR\text{-}_{20}$ = 0.87 for the total Sisto Scale. We considered such values acceptable, suggesting good reliability for the two scales (Dunn et al., 2014).

When constructing the DAP scale with 30 indicators, Sisto (2005) suggested that only those with an item-theta correlation above 0.20 should be maintained. For the author, the higher the correlation value, the more the indicator contributes to differentiating the age of the participants since it is an instrument that has a developmental character. The set of 12 indicators showed item-theta correlations above 0.32, suggesting that this choice has good assumptions for assessing children's performance in terms of age. From this perspective, through the analysis of variance, the set of reduced correction indicators presented statistically significant differences between and among the groups [F 7769.379) = 96.654, p <0.001]. Tukey's *post hoc* test indicated the groups that differed (Table 4).

Table 4 – Tukey's test according to age differences

		Subgroup for significance <0.005				
Age	*n*	1	2	3	4	5
6	219	1.96				
7	291		2.47			
8	272			3.33		
9	285				4.12	
10	336					5.01
p		1.000	1.000	1.000	1.000	1.000

In general, the proposal to correct the DAP using 12 indicators is a measure capable of discriminating evolutionary levels of children's conceptual repertoire since all ages showed differences between each other.

Discussion

The aim of this study was to verify whether the correction of the DAP through a reduced set of criteria would be a possible measure to assess the non-verbal intelligence of children. Since it was defined as a systematic method for assessing non-verbal intelligence, several researchers have developed different proposals for the application, correction, and interpretation of the Draw-a-Person Test (Goodenough, 1926; Harris, 1963; Koppitz, 1968; Sisto, 2005; Wechsler, 2003). Sisto (2005), the author that presented the application and correction system closest to Goodenough's original suggestion, proposed his correction criteria based on the Rasch model, understanding that when representing a human drawing, children tend to manifest a certain amount of a particular ability. Therefore, the author identified three difficulty levels for his scale, believing that people may have a greater or lesser possibility of endorsing an indicator, depending on their cognitive capacity.

In this study, unlike the Rasch model that fixes the item discrimination parameter, it was found that the 2PL model that freely estimates this parameter was the most suitable for evaluating the DAP, suggesting that children's responses for all indicators depend on the same latent trait, that is, on the same ability (Bond & Fox, 2015). The comparison of the Rasch, 2, 3, and 4 parameter models through the AIC and BIC results, indices that measure the quality of statistical models and visualize their simplicity, reaffirmed that it is most appropriate to evaluate the indicators that compose the DAP from two parameters. After excluding the overlapping items, there was a set of 12 correction indicators, namely, trunk, two or three garments, legs, feet, integrated neck, fingers, eyes, arm, elbow and shoulders, physiognomic features, chin and ears, and thumb. Although the eyes are considered in the correction of physiognomic features, the latter evaluates the concept of integrated details, while the first refers to the capacity for proportion. The same happens with the fingers, in which the correct amount is observed while drawing the thumb refers to laterality (Sisto, 2005).

The distribution of the 12 indicators followed a hierarchy capable of favoring the assessment of the level of childhood conceptual repertoire, allowing the observation that the children's level of abilities remained constant throughout the instrument. Through the four criteria of the easy, medium, and difficult levels, it was possible to consider that the most skillful children will be able to draw the most difficult details while the less skillful ones will not. Although the choice of indicators for this proposed DAP correction through reduced criteria was specifically based on the two-parameter model, a brief review of the criteria suggested by the different correction systems indicated that, except for the chin and ears details (which are not considered by Koppitz's and Naglieri's systems), the other indicators are repeated in all systems.

Regarding the reliability verification, it was observed that the $KR\text{-}_{20}$ coefficient for the 12 correction criteria of the DAP – Sisto Scale was 0.77. We considered this value satisfactory since the alpha coefficient of the total scale was 0.87. According to Dunn et al. (2014), the value of the Kuder-Richardson coefficient is affected by the number of indicators that compose a scale, which justifies the difference in coefficient values from one correction proposal to another since 18 correction criteria were excluded. Regression analysis showed that these indicators predict 83% of children's conceptual repertoire. This evidences even more, the capacity of this proposal, which, despite being composed of only 12 indicators, showed a very satisfactory value to measure the evaluated construct.

As the DAP is an instrument used to assess non-verbal intelligence, this technique is related to several cognitive skills, such as perceptual-motor maturity, planning, and immediate memory (Carreras et al., 2013; Flores-Mendoza et al., 2010; Otoni & Rueda, 2020; Silva et al., 2015). As the DAP is an instrument that demands the reproduction of a drawing, children can add as much detail as they want. In this way, there are no errors, only the complete realization or not of these details, in which the absence or presence of these is scored. From this perspective, Sisto (2005) stated that the DAP differs from other instruments in which the manifestation of an *a priori* reality is expected, for, in this test, the reality is constructed by the child.

Goodenough (1929) and Sisto (2005) stated that over time, the ability to perceive and differentiate stimuli tends to improve as the child develops. Current studies prove that the DAP can differentiate performance based on age (Clarke et al., 2018; Comparini et al., 2017; Lilienfeld et al., 2015). The findings of this study corroborate the statement of these authors when verifying that a shorter, 12-item set of indicators has the same function, suggesting its use in clinical practice and psychoeducational contexts.

Our results suggest that DAP's correction with fewer criteria does not affect the main characteristics of the instrument, the assessment of intelligence, since when reproducing the drawing, the child still presents information regarding their ability for an association, the memory of details, discrimination, spatial orientation, planning, abstraction, and visual-motor coordination. Therefore, these reduced correction criteria to assess the DAP may allow the clinical and/or school psychologist to have an instrument that is simpler to apply and correct. This, would allow the screening of the maturity level of the child's intellectual knowledge and the formulation of diagnostic hypotheses and more comprehensive psychological assessment protocols.

Finally, we considered that this version with only 12 correction indicators can contribute to collective assessments with a large volume of children, in both the school/educational context and in the clinical practice. The identification of possible difficulties at the beginning of schooling allows for pedagogical intervention strategies to be carried out, minimizing possible intellectual impairments. As the name implies, the 12 indicators are only for screening in relation to intelligence. In cases where the child presents a low performance in the test, a correction that encompasses the 30 indicators of Sisto (2005) is suggested, in addition to a broad assessment that provides data on cognitive functioning.

Regarding limitations of this study, it can be highlighted that the 12 correction indicators of this screening version need to be tested in different samples, aiming to investigate whether they would continue corroborating the findings of this study. In addition, in this new proposal, it was not verified whether the selected indicators present a differential functioning of the item due to age, sex, and cultural region. We suggest that further studies, in addition to remedying these gaps presented, also seek evidence of predictive criterion validity, such as academic performance and achievement. Furthermore, we recommend that new validity evidence is sought and investigated with measures that assess related constructs such as perceptual-motor maturity, attention, and executive functions, among others. Research of this nature can contribute to the practical and scientific use of the instrument and to obtaining good psychometric indices.

References

Araújo, P., & Fernandes, R. I. (2015). O teste do Desenho da Figura Humana em crianças angolanas: contribuições à perspectiva da psicologia positiva. *Psicologia: Ciência e Profissão, 35*(3), 855-869. http://dx.doi.org/10.1590/1982-3703002132013

Arteche, A. X., & Bandeira, D. R. (2006). O desenho da figura humana: revisando mais de um século de controvérsias. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica, 2*(22), 133-155. http://www.redalyc.org/articulo.oa?id=459645449008

Bartholomeu, D., Cecato, J. F., Montiel, J. M., Machado, A. A., & Sisto, F. F. (2012). Teste de Bender (B-SPG) e HFD-Escala Sisto: Validade por grupos contrastantes. *Estudos Interdisciplinares em Psicologia*, *3*(2), 241-257. https://doi.org/10.5433/2236-6407.2012v3n2p241

Bandeira, D. R., Costa, A., & Arteche, A. X. (2012). The Flynn effect in Brazil: Examining generational changes in the Draw-a-Person and in the Raven's Coloured Progressive Matrices. *Revista Latinoamericana de Psicopatologia*, *44*(3), 9-18. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-05342012000300001

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.). Routledge.

Campbell, C. & Bond, T. (2017). Investing Young children's human figure drawings using Rasch analysis. *Educational Psychology, 37*(7), 888-906. https://doi.org/10.1080/01443410.2017.1287882

Carreras, M. A., Uriel, F., & Liporace, M. F. (2013). Actualizaciones en el análisis de ítem es madurativos del dibujo de la figura humana en niños escolarizados de buenos aires. *Interdisciplinaria, 30*(1), 101-118. https://doi.org/10.16888/interd.2013.30.1.6

Chalmers, M. P. (2012). *Package 'mirt'*. Citeseer. https://cran.r-project.org/web/packages/mirt/mirt.pdf

Clarke, M., McAneney, H., Chan, F., & Maguire, L. (2018). Inconsistencies in the drawing and interpretation of smiley faces: an observational study. *BMC Research Notes, 11*(1), 2-5. https://doi.org/10.1186/s13104-018-3185-0

Claparède, E. (1907). Plans dexperiences collectives sur le dessin des enfantes. *Archives de Psychologie, 6*, 276-278.

Comparini, I. P., Wechsler, S. M., & Machado, W. L. (2017). Indicadores emocionais no Desenho da Figura Humana: investigando evidências de validade. *Revista Psicologia: Teoria e Prática*, *19*(3), 256-269. https://doi.org/10.5935/1980-6906/psicologia.v19n3p256-269

De Ayala, R. J. (2008). *The theory and practice of item response item theory*. The Gilford Press.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Flores-Mendoza, C. E., Abad, F. J., Lelé A. J. & Mansur-Alves, M. (2010). O que mede o Desenho da Figura Humana? Estudos de validade convergente e discriminante. *Boletim de Psicologia, LX*(132), 73-84. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0006-59432010000100007

Goodenough, F. (1926). A new approach to the measurement of the intelligence of young children. *The Pedagogical Seminary and Journal of Genetic Psychology, 33*(2), 185-211. http://dx.doi.org/10.1080/08856559.1926.10532353

Harris, D. B. (1991). *El test de Goodenough. Revisión, ampliación y actualización*. Ediciones Paidós.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3. ed.). The Guilford Press.

Koppitz, E. (1973). *El dibujo de la figura humana en los niños*. Editorial Guadalupe.

Koppitz, E. (1968). *Psychological Evaluation of Children's Human Figure Drawings.* Grune & Stratton.

Koppitz, E. (1984). *Psychological evaluation of human figure drawings by middle school pupils*. Grune and Stratton.

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2015). *Science and pseudoscience in clinical psychology* (2. ed.). Guilford Press.

Luquet, G. H. (1927). *O desenho infantil*. Ed. Minho.

Machover, K. (1949). *Personality projection in the drawing of The Human Figure: A method of personality investigation*. Charles C Thomas Publisher.

Marôco, J. (2014). *Análise de equações estruturais: Fundamentos teóricos, software & aplicações*. Report Number.

Naglieri, J. A. (1988). *Draw A Person: A Quantitative Scoring System*. Psychological Corporation.

Nakano, T. C., Primi, R., & Nunes, C. H. S. S. (2015). *Análise de itens e Teoria de Resposta ao Item (TRI).* In C. S. Hutz, D. R. Bandeira, & C. M. Trentini (Orgs.), Psicometria [Psychometry] (pp. 97-124). Artmed.

Oliveira, K. S., & Wechsler, S. M. (2016). Indicadores de criatividade no desenho da figura humana. *Psicologia: Ciência e Profissão, 36*(1),6-19. https://doi.org/10.1590/1982-370301682014

Otoni, F., & Rueda, F. J. M. (2020). Perceptive-Motor Maturity and its Relations with Planning, Memory Immediate and Non-Verbal Intelligence. *Paidéia, 30*(e3031). http://dx.doi.org/10.1590/1982-4327e3031

Panesi, S., & Morra, S. (2016). Drawing a dog: The role of working memory and executive function. *Journal of Experimental Child Psychology, 152*, 1–11. http://dx.doi.org/10.1016/j.jecp.2016.06.015

Picard, D. (2015). Sex differences in scores on the draw-a-person test across childhood: do they relate to graphic fluency? *Perceptual & Motor Skills, 120*(1), 273-287. http://dx.doi.org/10.2466/10.27.PMS.120v10x1

Primi, R., Nakano, T. C., & Wechsler, S. M. (2018). Using four-parameter item response theory to model Human Figure Drawings. *Avaliação Psicológica, 17*(4), 473-483. http://dx.doi.org/10.15689/ap.2018.1704.7.07

Rehrig, G., & Stromswold, K. (2017): What Does the DAP:IQ Measure? Drawing Comparisons between Drawing Performance and Developmental Assessments. *The Journal of Genetic Psychology, 179*(1), 9-18. https://doi.org/10.1080/00221325.2017.1392281

Rosa, H. R. (2008). Validade do desenho da figura humana na avaliação de Goodenough-Harris e nos indicadores maturacionais de Koppitz em crianças da cidade de São Paulo. *Boletim de Psicologia, LVIII*, 1-4. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0006-59432008000100002

Rouma, G. (1913). *La language graphique de l'enfant*. Misch et Thron.

Rosa, H. R., & Alves, I. C. B. (2014). Estudo normativo do Teste Goodenough-Harris em crianças da cidade de São Paulo. *Boletim Academia Paulista de Psicologia, 34*(87), 336-351. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S-1415-711X2014000200004

Rosa-Neto, F., Xavier, R. F. C., Santos, A. P. M., Amaro, K. N., Florêncio, R., & Poeta, L. S. (2013). A lateralidade cruzada e o desempenho da leitura e escrita em escolares. *Revista CEFAC, 15*(4), 864-872. https://doi.org/10.1590/S1516-18462013000400015

Rueda, F. J. M., Noronha, A. P. P., Santos, A. A. A., Jesuíno, A. D. S. A., Zuanazzi, A. C., Ferraz, A. S., Costa, A. R. L., & Otoni, F. (2020). Human Figure Drawing: Systems Most Used in Cognitive Assessment of Children. *Psico, 51*(1), 1-13. http://dx.doi.org/10.15448/1980-8623.2020.1.31313

Silva, M. C. R., Montiel J. M., Jr. Fiamengi, G. A., & Bartholomeu, D. (2015). *Técnicas gráficas aplicadas à educação e à saúde*. Mennon.

Sisto, F. F. (2005). *Desenho da Figura Humana - Escala Sisto*. Vetor.

Tükel, E., Eliasson, A. C., Böhm, B., & Smedler, A. C. (2018). Simple Categorization of Human Figure Drawings at 5 Years of Age as an Indicator of Developmental Delay, Developmental Neurorehabilitation, 1-8. https://doi.org/10.1080/17518423.2018.1532969

Wechsler, S. M. (2003). *HFD III. O Desenho da Figura Humana: avaliação do desenvolvimento cognitivo de crianças brasileiras.* Editora da Pontifícia Universidade Católica de Campinas.

Fernanda Otoni

Doutora em Psicologia em Universidade do Paraná (UTP), em Campinas, SP, Brasil; mestre em Psicologia em Universidade do Paraná (UTP), em Campinas, SP, Brasil. Professora do Programa de Pós-Graduação em Psicologia Forense da Universidade Tuiuti do Paraná (UTP), Curitiba, PR, Brasil.

Acácia A. Angeli dos Santos

Doutora em Psicologia Escolar e do Desenvolvimento Humano pela Universidade de São Paulo (USP), em São Paulo, SP, Brasil; mestre em Psicologia Clínica pela Pontifícia Universidade Católica de Campinas (PUC-Campinas), em Campinas, SP, Brasil.

Fabian Javier Marin Rueda

Doutor em Psicologia em Universidade do Paraná (UTP), em Campinas, SP, Brasil; mestre em Psicologia em Universidade do Paraná (UTP), em Campinas, SP, Brasil. Professor da Universidade São Francisco (USF), Campinas, SP, Brasil.

Psico, Porto Alegre, v. 54, n. 1, p. 1-12, jan.-dez. 2024 | e-41815

Mailing address

Fernanda Otoni

Rua Padre Ladislau Kula, 395

Santo Inácio, 82010-210

Curitiba, PR, Brasil

Os textos deste artigo foram revisados pela SK Revisões Acadêmicas e submetidos para validação do(s) autor(es) antes da publicação.