OPEN ACCESS

ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA

PSICO

Psico, Porto Alegre, v. 51, n. 1, p. 1-13, jan.-mar. 2020
e-ISSN: 1980-8623 | ISSN-L: 0103-5371

http://dx.doi.org/10.15448/1980-8623.2020.1.31313

REVISÃO

Desenho da Figura Humana: Sistemas Mais Utilizados na Avaliação Cognitiva de Crianças

Human Figure Drawing: Systems Most Used in Cognitive Assessment of Children

Dibujo de la figura humana: sistemas más utilizados en la evaluación cognitiva de los niños

Fabian Javier Marin Rueda[1]
orcid.org/0000-0001-5173-0802
marinfabian@gmail.com

Ana Paula Porto Noronha[1]
orcid.org/0000-0001-6821-0299
ana.noronha8@gmail.com

Acácia Angeli dos Santos[1]
orcid.org/0000-0002-8599-7465
acacia.angeli@gmail.com

Ana Deyvis Santos Araujo Jesuíno[1]
orcid.org/0000-0002-7031-7682
anadeyvis@gmail.com

Ana Carolina Zuanazzi[1]
orcid.org/0000-0003-1649-2372
anacarolina.zf@gmail.com

Adriana Satico Ferraz[1]
orcid.org/0000-0002-9856-0094
drianasatico.as@gmail.com

Ariela Raissa Lima Costa[1]
orcid.org/0000-0002-5942-6466
arielalima10@gmail.com

Fernanda Otoni[1]
orcid.org/0000-0002-9347-7144
fer_ottoni@hotmail.com

Submetido em: 03 jul. 2018
Aceito em: 11 set. 2019
Publicado em: xx xxx. 2020.

Resumo: Investigou-se sistemas cognitivos do Desenho da Figura Humana (DFH) mais utilizados no âmbito internacional e nacional, e realizou análise qualitativa. A busca foi feita no Periódicos CAPES, SciELO, Redalyc e Pepsic. A partir dos critérios de seleção, foram analisados 33 artigos que permitiram a identificação do sistema Goodenough-Harris, como o mais estudado internacionalmente, e o sistema Wechsler no contexto nacional. O ano de 2005 apresentou maior publicações. Observou-se predomínio da busca de evidências de validade com amostras brasileiras, enquanto que internacionalmente os estudos foram mais heterogêneos. Conclui-se que a maioria dos sistemas apresentam propriedades psicométricas minimamente adequadas sendo, porém, necessário a realização de novos estudos que investiguem o funcionamento dos sistemas, principalmente no âmbito nacional, que possui apenas 10 estudos. Limitações e agenda de pesquisa são apresentadas.

Palavras-chave: DFH; teste de figuras humanas; cognição; crianças; avaliação psicológica

Abstract: This study investigated the most frequently used Human Figure Drawing (HFD) cognitive systems in the international and national scope, and carried out a qualitative analysis of them. A search was conducted through the databases Periódicos CAPES, SciELO, Redalyc, and Pepsic. Based on the selection criteria, a review of 33 articles provided the identification of the Goodenough-Harris system, as the most studied internationally, and the Wechsler system in the national context. The largest number of publications occurred in 2005. Regarding the study types, among those with a Brazilian sample most aimed to find validity evidence, while internationally studies were more heterogeneous. The conclusion is that most systems presented minimally fit psychometric properties, but that new studies are necessary to investigate system function, particularly at the national level, which has only 10 studies. Limitations and research agenda are presented.

Keywords: HFD, human figures tests, cognition, children, psychological evaluation

Resumen: Se investigó los sistemas cognitivos más utilizados del Dibujo de la Figura Humana (DFH) en el ámbito internacional y nacional, y realizó un análisis cualitativo. La búsqueda fue en las bases de datos Periódicos CAPES, SciELO, Redalyc y Pepsic. Un análisis de 33 artículos proporcionó la identificación del sistema Goodenough-Harris como el más estudiado internacionalmente, en el contexto nacional fue el sistema Wechsler. El mayor número de publicaciones se produjo en 2005. Se observó que en los realizados con muestras nacionales la mayoría buscó evidencias de validez y los estudios internacionales eran más heterogéneos. La conclusión es que la mayoría de los sistemas presentan propiedades psicométricas mínimamente ajustadas, pero se necesitan nuevos estudios para investigar la función de cada uno de esos sistema, particularmente a nivel nacional, que fueron encontrados apenas 10 estudios. Se presentan limitaciones y una agenda de investigación.

Palabras Clave: DFH, pruebas de figuras humanas, cognición, niños, evaluación psicológica

[1] Universidade São Francisco (USF)

Human Figure Drawing (HFD) is a relevant instrument for the evaluation of cognitive abilities and is widely used in professional practice. Diverse systems to score and interpret have been developed from Goodenough's initial proposal in 1926. As indicated by Arteche and Bandeira (2006), it is important to investigate which evaluation systems are most studied, considering the indispensability of empirical evidence and sufficient validity evidence to justify its use by psychologists. However, in a more recent publication, Suehiro, Benfica, and Cardim (2016) noted the existence of few papers that aimed to verify the psychometric properties of this instrument, which justifies the present study.

To analyze the results of empirical research on children's drawings, both gender, between the ages of two and 15, Goodenough (1926) identified a relationship between intelligence and the drawings produced. The author postulated that to draw a human figure, young children used more of an intellectual component, rather than plastic/aesthetic intelligence, and noted the existence of characteristics related to cognitive development, rather than manual or visual ability, indicating that the child expresses something that s/he understands cognitively. This means that younger children reproduced what they know about people, producing a drawing with few details, and as they developed, they conceived figures as they see them, that is, characterizing an evolution in drawing concomitant with age.

Thus, for evaluation of cognitive development, Goodenough (1926) proposed the activity named as the Draw-A-Man Test, which was evaluated by 51 graphic elements such as the presence and quality of arms, eyes, and feet, among others. This allowed for an objective weighting of complexity through the presence or absence of items, according to Colom, Flores-Mendoza, and Abad (2007). Such characterization solidified the author's work as pioneering, recognized for its systematization and standardization (Arteche & Bandeira, 2006).

Given the impact of Goodenough's work, new systems were developed, and more information was added to those previously evidenced by the author. In general, systems have common aspects, especially regarding the application and coding process. Similarities include the instruction to draw a human figure and scoring for elements of the drawing. There are, however, important differences that characterize each system, as discussed below.

As highlighted by Arteche and Bandeira (2006), the primary difference between the different systems lies in the evaluation, which is divided into three main domains: cognitive, personality and emotional. The first set encompasses those that understand and analyze the drawing as an indicator of cognitive development, following the original Goodenough's proposal. Examples of such systems are: Goodenough-Harris (Abell, Horkheimer, & Nguyen, 1998; I. Alves, 1981), Naglieri (1988), Koppitz (1968), Wechsler (2003) and Sisto (2005).

Regarding scoring, each system adopted many items according to empirical evidence. For example, Goodenough-Harris presents 73 items for a drawing of a man and 71 items for a drawing of a woman. When the child draws himself/herself, the scale for the child gender is used. Naglieri (1988), in turn, developed a quantitative system, divided into four categories (presence, detail, proportion, and addition) and 14 criteria. Also, some children had five minutes to draw. Both Koppitz's and Sisto's systems are based on 30 items for scoring. Finally, the Wechsler system consists of 58 items and asks the respondent for two drawings, one female and one male.

The second domain evaluates drawing as a form of personality expression through projection. An example is Machover's system (1949), which requests the child to draw a woman and a man. There is no standardization of analysis of the drawings, but an interview is proposed after finish the draw, through which emotional conflicts are evaluated based on the concepts of projection.

Finally, the third domain includes systems that evaluate emotional aspects and are represented by Koppitz (1968) and Naglieri (Naglieri, McNeish, & Bardos, 1991). The Koppitz system, based on Goodenough and Machover, is composed of 30 items considered as emotional indicators. These elements were identified in children in a clinical

sample through empirical studies developed by the author. The items were analyzed and related to potential emotional difficulty. The Naglieri system consists of a score based on 55 items, that summarize the presence or absence of elements, the drawing's size and positioning on the page.

In Brazil, there are two systems available for use, the Wechsler (2003) and the Sisto (2005) (Federal Council of Psychology, 2018). Both evaluate cognitive aspects in children, although some studies have indicated the possibility of emotional and creativity evaluation in the Wechsler system (Comparini, Wechsler, & Machado, 2017; Oliveira & Wechsler, 2016).

Noronha, Beraldo, and Oliveira (2003) indicated HFD as one of the most well-known instruments used by psychology students and professionals in Brazil. More than a decade after this finding, Suehiro, Benfica, and Cardim (2015) indicated that this instrument still in use and study, especially in children's cognitive assessment. Besides, Alves, Rosa, Da Silva, and Sardinha (2016) analyzed the frequency of use of 96 instruments, with HFD being the 5th most used.

In Suehiro et al. (2015) study, 67 articles were classified in terms of a) journals with the highest number of publications on the topic, b) publication frequency according to the region of the country, and c) the primary evaluation contexts. The two journals with the highest number of publications were Psychology: Reflection and Criticism (9) and Psychological Assessment (6). The Brazilians regions with the most papers were Southeast and South (88.8%), with School and Hospital the most frequently contexts of assessment (80.6%). The authors concluded that the number of studies on the subject is low and that more research is needed, especially in concerns about their quality.

The Alves et al. (2016) study investigated which intelligence tests were mostly used in Brazil between 2005 and 2014, and located 96 instruments cited in 72 articles. The authors mapped the instruments in terms of frequency, publication years, and main objectives. Some identified gaps pertained to poor instrument description quality and use of instruments not approved by the SATEPSI (Psychological Testing System), which evaluates the technical-scientific quality of psychological instruments for professional use according to CFP Resolution No. 009/2018 (Federal Council of Psychology, 2018).

Suehiro et al. (2016) investigated the scientific production on HFD between 2002 and 2012. The authors found a total of 39 articles referring to different systems. When analyzing the two systems used in Brazil, only five studies were found, among them, four on the Wechsler system and one on the Sisto system. This data reveals a dearth of studies on HFD systems. However, it is worth noting that the authors limited their searches to only two databases (SciELO and Pepsic). Both databases are well known and used in Brazilians research, but they are not big as other international databases, such as the ones we focus on this study.

In the three review studies cited, a focus was placed on research classification (publication frequency, most used journals, the most researched region of the country, authorship composition, among others) and had less or none focus on the qualitative analysis. Qualitative analysis is understood as the appreciation of the studies based on their main objectives and the results achieved, summarizing the findings, as described by Ramos Vosgerau and Romanowski (2014). This analysis procedure allows for a broader understanding of what has been researched on the topic, its main findings and consequently the possibilities for advances.

Studies which emphasize literature reviews help to explore the most current state of a given topic, to summarize and identify gaps still unanswered, contradictions or inconsistencies, as well as facilitating the proposition of new subjects to be investigated (American Psychological Association [APA], 2012). As pointed out, there is a gap in terms of more detailed research on the use of human figure drawing, and this article aims to verify the most used HFD cognitive systems, presenting not only a greater diversity of consultation bases but also performing a qualitative analysis of the results showed.

Method

Sources

The articles searches were conducted in March 2018 based on four databases – CAPES Periodicals, Scientific Electronic Library Online (SciELO), Network of Scientific Journals of Latin America and the Caribbean, Spain, and Portugal (Redalyc) and Portal of Psychology Electronic Journals (Pepsic). A search period was not determined to retrieve the largest number of studies. In CAPES Periodicals, the search descriptors were *Draw-a-person test* and *children*; in SciELO and Pepsic, *human figure drawing* and *children*; and in Redalyc "Dibujo de la Figura Humana" and "Niño." Except for Redalyc, results were filtered by "all indices" and "peer-reviewed journals." For the Redalyc database a "content filter" was applied, which enables better article tracking in this database.

Selection Procedure

After reading articles titles and abstracts, only those that proposed to evaluate cognitive aspects in children or HFD literature reviews were selected. Then, after thoroughly reading the selected articles, they were categorized according to publication year, research type (theoretical/review, empirical or case study) (APA, 2012), sample characteristics (number of participants, gender, schooling level and type of school – public or private), research objectives, main results, and HFD system used.

Results

The searches resulted in 107 articles from CAPES Periodicals, four articles from SciELO, 10 from Pepsic and 120 from Redalyc, for a total of 241 articles. Three independent judges analyzed the articles and, where divergence occurred, results were discussed until a consensus was reached. From the collection of essays, those who assessed emotional aspects and studies with adult samples were excluded because they are not within the scope of this review, which eliminated 204 articles. A total of 37 articles were read, with three more articles excluded: one due to sample type (adults), one due to the HFD correction and analysis type (evaluation of emotional aspects), and one that cited HFD but did not use it as an instrument. A total of 34 articles remained for further analysis, published between 1955 and 2017.

The data was organized into two sets: the presentation of the descriptive analysis and, then, qualitative analysis. Concerning descriptive analysis, information on publication frequency and year range, research types, sample nationality, and system used was collected. For the qualitative analysis, the main results were considered.

TABLE 1 – *List of national studies*

Authors and Year	System	Objective	N	Principal Results
Wechsler & Schelini (2002)	Wechsler	Validity evidence	255	Correlation between HFD and intelligence ranged from r=0.21 to r=0.27
Flores-Mendoza, Abad & Lelé (2005)	Wechsler	Psychometric analysis of the items that make up the male HFD using mathematical models of the Item Response Theory (IRT).	1,275 (6-12y)	Alpha value of 53 items' fit (0.87). The IRT analysis identified problems in a group of items with difficulty and discrimination.
Rueda, Bartholomeu & Sisto (2006)	Gd	Correlate results obtained in the Bender Test and the HFD	312 (7-10y)	Negative and significant correlations between the tests ranging from 0.21 to 0.39
Bandeira, Costa & Arteche (2008)	Wechsler	Validity evidence	90(6-12y)	Correlations between HFD and Raven (r=0.50) and between HFD and school performance scale (r=0.34)

Authors and Year	System	Objective	N	Principal Results
Flores-Mendoza et al. (2010)	Wechsler Gd Gd-H	Validity evidence	628 (5-11y)	Correlation between systems (Gd, Gd-H, and Wechsler ranging from 0.72 to 0.79); correlation between HFD and R-2 (from 0.32 to 0.38); and HFD and Bender (0.39 to 0.50)
Bartholomeu et al. (2012)	Sisto	Investigate validity evidence through contrasting groups	112 (7-10y)	Correlations with higher maturity on the HFD show less distortion in Bender
Bandeira, Costa & Arteche (2012)	Wechsler Gd-H	Evaluate the Flynn effect in intelligence tests	497 (6-12y)	The results did not show significant generational effects, regardless of the measure used.
Noronha, Santos & Rueda (2013)	Sisto	Check evidence of convergent validity	397 (6-24y)	Correlations between HFD and other tests were moderate, ranging from 0.49 to 0.60.
Zuraban Santos et al. (2013)	Wechsler	Analyze cognition correlations with cancer	6 (7-12y)	The majority of the children presented average performance, two above, and one below average.
Rosa & Alves (2014)	Gd-H	Norms for Goodenough-Harris HFD Testing	1,540 (5-11y)	Nursery School: significant differences between all variables and interaction between gender and type of school. Elementary School: significant difference between age.

Nota: Gd = Goodenough, Gd-H = Goodenough-Harris, NS = Not specified

TABLE 2 – *List of international studies*

Authors and Year	Country	System	Objective	N	Principal Results
Araújo & Fernandes (2015)	AO	Gd	Check the applicability of the test in Angolan correlations.	6 (6-10y)	The test proved to be useful in the intellectual evaluation of correlations.
Carreras, Uriel, Fernández Liporace (2013)	AR	Gd-H	Find validity evidence for HFD	785 (6-12y)	Test fit in the sample. Correlation between Bender and HFD was r=0.519.
Pihl & Nimrod (1976)	CA	Koppitz	Verify reliability.	44	HFD and general academic achievement (r=0.30), reading (r=0.26) and written expression (r=0.29).
Woodbur, Fernández & Boschini (1989)	CS	Gd-H	Estimate relations of the concept of body self-image.	90 (7-9y)	The correlations between PDI-C-UMA and Gd-H were r=0.49; p<0.001

Authors and Year	Country	System	Objective	N	Principal Results
Nielsen (1961)	DK	Gd	Verify differences between correlations with and without cerebral paralysis.	80 (6-14y)	There were no significant differences between the drawings.
Robles et al. (2009)	ES	Gd	Examine mental development in correlations exposed to the television.	130 (5-10y)	There was a statistically significant negative correlation with the use of tv (r=-0.2, p=0.02)
Garaigordobil & Amigo (2010)	ES	Koppitz	Analyze the relationship between intelligence and self-concept.	74 (5y)	HFD and verbal intelligence (r=0.53); nonverbal intelligence (r=0.27) and total intelligence (r=0.52)
Picard (2015)	FR	Gd	Verify the difference between boys and girls in HFD and the interference of graphic skills.	336 (5-12y)	There was a significant difference between the sexes at the younger ages, favoring the girls. There was an increase in graphic skills as the age increased. No difference between sexes.
Hagood (2003)	GB	NS	Verify the difference between sexes in the three HFD figures.	34 (5-10y)	Drawings of correlations from each age group showed that details tend to increase with age. Lack of significant differences between boys and girls.
Lange-Küttner, Küttner & Chromekova (2014)	GB	Naglieri	Investigate effects of repetition on HFD.	80 (6-12y)	Repetition did not improve the correlations score, but there were less detail and deterioration of the correlations score up to 10 years after repetition.
Dorazco-Valdes (1968)	MX	Gd	Verify performance in epileptic correlations.	19 (5-13y)	Patients with more severe cases had worse results in HFD
Casillas (2012)	MX	Gd	Compare the level of development of students in educational centers.	50 families (4-6y)	The HFD results are only used for screening and are not explored in the study.
Ter Laak et al. (2005)	NL	Gd–H	Verify reliability and validity.	115 (7-9y)	The reliability was moderate (0.64). Drawing skill estimation by judges and teachers correlated significantly (r=0.59)
Merino (2013)	PE	R&H	Verify the validity of two visuomotor measures.	154 (4-8y)	Age-related to intellectual ability. HFD results are consistent with expectation.

Authors and Year	Country	System	Objective	N	Principal Results
Swensen & Newton (1955)	US	NS	Verify the development of correlation differences between sexes in HFD.	185	Difference between younger and older correlations in indicating sex differences in drawings. From age 8, there were no significant differences.
Datta (1967)	US	Gd-H	Verify performance in correlations of low-income families.	956 (3-6y)	The average performance of HFD participants when compared to normative sample. There were no differences between boys and girls in HFD.
Adams & Lieb (1973)	US	Gd-H	Check the performance of black and white correlations.	162 (4-6y)	Test performance is not changed due to ethnicity.
Ritter, Duffey & Fischman (1974)	US	Gd-H	Validity evidence.	31 (4-6y)	PPVT and HFD were not reliable to estimate the intelligence of correlations in kindergarten.
Oakland & Dowling(1983)	US	Gd	Psychometric properties of HFD in three different ethnic groups.	188 (8-10y)	Correlations between HFD and WISC-R scales were significant (r=0.55). Correlations were greater for age 10 than for younger ones.
Stromme & Smith (1973)	US	Gd-H	Validity evidence.	150 (5-8y)	Adequate internal consistency for all ages (0.63 to 0.82), with no differences between sexes.
Prewett, Bardos & Naglieri, (1988)	US	Naglieri	Validity evidence.	77	The group with difficulty had significantly lower scores on HFD.
Prewett et al. (1989)	US	Naglieri	Verify use of HFD for mental deficiency screening.	85	Not useful in tracking correlations with deficiency.
Haddad & Juliano (1991)	US	Naglieri	Verify the relation between cognitive tests and HFD in correlations with low socioeconomic level.	82 (8-10y)	Correlation between Matrix Analogies and HFD of 0.32 and HFD and Iowa was 0.47.
Olatunya et al. (2017)	NG	Ziler	Verify the relation between intellectual indicators and blood diseases.	101 (4-12y)	Correlation of 0.86 between HFD and school performance.

Note. US – United States, DK – Denmark, CA – Canada, CS – Costa Rica, AR – Argentina, AO – Angola, ES – Spain, FR – France, MX – Mexico, GB – Great Britain, PE – Peru, NL – Netherlands, NG – Nigeria.
Gd = Goodenough, GD-H = Goodenough-Harris, NS = Not specified, R&H = Reynold & Hickman

Tables 1 and 2 present data for national and international articles, respectively. These tables reveal a higher frequency of publications after 2005, with the highest concentration between 2010 and 2014. Also, most of the studies were empirically delineated, with only one bibliographic

review and one case study, which were selected from the proposed criteria. Regarding nationality on the selection, a predominance of studies was developed by Brazilians (10), followed by the US (9). The most used systems were Goodenough-Harris (10) and Goodenough (8). In Brazil, the Wechsler system was the most researched (6), while the Sisto system was used in two surveys.

After the descriptive analysis, a qualitative evaluation of the selected studies was performed. The principal results concerning the HFD cognitive aspects evaluation are also included in tables 1 and 2. For this, only empirical studies are considered, as our objective was to verify the research that used HFD as a form of cognitive evaluation or as an instrument used in comparison with others.

The sample sizes used in the studies ranged from 6 to 1,540 children. The majority of studies aimed to find validity evidence for HFD or other instruments. Some studies found performance differences based on gender, but the results were divergent. Some found differences between boys and girls, while others could not replicate this finding (e.g., Picard, 2015; Rosa & Alves, 2014).

In general, all studies indicated HFD as a suitable instrument for assessing cognitive abilities in children, relating positively to other non-verbal intelligence tests such as Raven (e.g., Bandeira, Arteche, & Costa, 2008), and Bender (eg, Noronha, Santos, & Rueda, 2013), and verbal intelligence tests such as the Wechsler Intelligence Scale for Children (WISC; e.g. Oakland & Dowling, 1983). Significant correlations were mostly positive and with moderate magnitude (Cohen, 1992) and added convergence validity evidence for HFD in their different cognitive assessment systems.

Comparing national and international studies in terms of objectives and main results, in Brazil, a predominance of research focus on validity evidence. In international research there was more heterogeneity of goals, which included, for example, the use of HFD to differentiate between gender and ethnic groups, as reported below. About the year of publications, international studies are older than the Brazilian ones.

The work in Brazil, initiated in the 2000s with the Wechsler system, presented mostly validity evidence with the HFD proposal. Correlations were moderate with other cognitive assessment instruments such as Bender, Raven, and measures of school performance. Regarding the systems used, Wechsler (6) was predominant, followed by Goodenough-Harris (3). The Goodenough system and the Sisto Scale were each used in two studies. The sample size ranged from 6 to 1,540 children from different age groups.

Regarding the main findings divided by system type, most Wechsler system studies presented evidence of reliability, except for two. Of the latter, the first study showed weak correlations with the Nonverbal Test of Child Reasoning (TNVRI) (Wechsler & Schelini, 2002). According to the authors, this result may be explained by the fact that the first test is related to analogical reasoning, a construct that encompasses different dimensions than HFD. The other study presented difficulty in discriminating one from a group of items, through IRT, with the conclusion that the instrument needed calibration (Flores-Mendoza, Abad, & Lelé, 2005).

In the Goodenough system, the study (Rueda, Bartholomeu, & Sisto, 2006) found correlations of 0.21 to 0.39 between the HFD and the Bender test. In the Goodenough-Harris system, normalization was performed with 1,540 children and cognitive development reliability indicators were identified according to age and educational level (Rosa & Alves, 2014), as well as positive and moderate correlations with Bender (Carreras, Uriel, & Fernández Liporace, 2013). Finally, there are two studies that, using the Sisto system, aimed to analyze the validity evidence comparing HFD with other nonverbal cognitive performance tests (Bartholomeu, Cecato, Montiel, Machado, & Sisto, 2012; Noronha et al., 2013). Correlations were positive and moderate, varying from 0.49 to 0.60.

Two national studies had objectives different from the previously cited. The first used the Wechsler and Goodenough-Harris systems to analyze the Flynn effect and found no such time effect in the sample of 497 children (Bandeira, Costa, & Arteche, 2012). The second study used only the Wechsler system to check

cognitive performance in a sample of six children with cancer. The authors obtained results of performance within the normative average, finding one child with inferior performance, and two above-average (Zubaran Santos, Sardá Junior, Menezes, & Thieme, 2013).

About international research, these studies come from several continents, including South America, North America, Africa, and Europe, and date back to 1955. The sample sizes ranged from six to 956 children in different age groups. The most used systems were Goodenough and Goodenough-Harris, followed by the Naglieri, Koppitz, Reynolds and Hickman, and the Ziler system. Two studies did not specify the system used.

Analyzing the work developed within each system, results were divergent in nearly all systems. In Goodenough specifically, we highlight four studies that indicated the system's reliability. One example, Dorazco-Valdes (1968), found a positive association between low HFD performance and symptomatic severity in epileptic patients, and Oakland and Dowling (1983) presented a correlation of 0.55 between WISC-R and HFD. The study by Araújo and Fernandes (2015) analyzed the adequacy of the instrument for the evaluation of Angolan children, finding satisfactory results. On the other hand, Nielsen's (1961) research indicated the system was imprecise in differentiating the performance of children with and without cerebral palsy. Another study carried out by Picard (2015), aimed at verifying differences between boys and girls but did not find data to support this hypothesis.

The Goodenough-Harris system presented the most consistent results, in psychometric terms, with internal consistency ranging from 0.63 to 0.82 and correlations with cognitive measures varying from 0.49 to 0.59 (Carreras et al., 2013; Woodburn & Boschini, 1989). On the other hand, a study by Ritter, Duffey, and Fischman (1974) identified that HFD was not reliable in estimating the intelligence of children in early grades, such as kindergarten.

The Naglieri system, though it is associated with emotional assessment, it was also used to verify cognitive indicators. However, in analyzing the results, the system was not adequate in discerning cognitive differences between children with and without an cognitive delay diagnosis (Prewett, Bardos, & Naglieri, 1989). Despite this, correlations of 0.32 and 0.47 were found with the Matrix Analogies Test and Iowa cognitive instruments (Haddad & Juliano, 1991). Regarding the effect of repeating details when drawing, Lange-Küttner, Küttner, & Chromekova (2014) verified that there was neither performance improvement nor a reduction in the number of details in drawings resulting from re-test procedure.

Psychometric studies with the Koppitz system found adequate reliability and validity evidence based on correlations with cognitive instruments, ranging from 0.27 to 0.52 (Garaigordobil & Amigo, 2010; Pihl & Nimrod, 1976). One study was based on the system developed by Reynolds and Hickman and correlated the HFD with a visomotricity measure, verifying a correlation value of 0.75 with the Visuomotor Integration Test-version 4 (Merino, 2013). Finally, Olatunya et al. (2017) found a correlation of r = 0.86 with school performance.

Discussion

The objective of this study was to map research developed with samples of children using application and correction systems of Human Figure Drawing, specifically the cognitive aspects. To do so, databases were searched using specific keywords to enhance search comprehensiveness. After analysis and selection, the data were arranged descriptively and qualitatively, organized by system.

From the data presented, some observations can be made. With regard to publication frequency, there was an increase in studies focusing on cognitive aspects since the 2000s. There is a greater tendency to publish empirical methodologies to the detriment of bibliographical/theoretical reviews or case studies. These findings, also present in other reviews (Alves et al., 2016; Suehiro et al., 2015), reflect the present moment of psychological evaluation in Brazil, where the number of publications increased in the early 2000s, and those began to adopt replicable

methodologies that allow further generalizations.

Regarding the nationality of research, the selection of databases to gather national and international studies justifies the results found: three databases, predominantly Latin American, likely contributed to the recovery of a higher number of studies in these countries. Nevertheless, it was possible to incorporate many US and European studies, which indicates the growing internationalization and accessibility of science.

In association, it is noteworthy that the instrument is researched in different cultures. The Goodenough-Harris and Goodenough systems are the most investigated in the international and national contexts, even though there are two systems adapted to/developed for Brazil. The mean sample size ranged from 200 to 300 children. To generalize the data, and even for the development of normative samples, a more significant number of children would be useful (Valentini & Hauck Filho, 2014).

Based on the qualitative analysis of the studies in general, the results reveal that the systems included here present psychometric properties that are minimally adequate for use in children cognitive assessment. However, all systems need new studies, especially those that analyze the discriminative capacity of children with different cognitive conditions. Most research that proposed this objective were not successful in discerning children with cognitive difficulties from those without complaints of cognitive delay.

Specific to the national studies, it was expected that the first surveys would start in the 2000s, since both HFD systems, approved by SATESPI, date back to 2003 (Wechsler) and 2005 (Sisto). It is noteworthy that both systems require revision, as pointed out by Flores-Mendoza, Abad, and Lelé (2005). Although studies with the Sisto Scale did not reveal psychometric problems, it should be taken into account that they were performed in smaller numbers (only two) and with relatively small samples. This suggests the need for further research, especially concerning the psychometric qualities of the system. Finally, about Brazilian studies, SATEPSI-approved systems are completing, respectively, 15 and 13 years of approval for professional use, which by Resolution 002/2003 indicates the importance of new psychometric studies that should be presented soon, to continue with SATEPSI approval.

Some limitations of this review may be listed and considered in the conduct of new studies. Specifically, with regard to databases, it would be interesting to search databases that include broader international coverage, such as PUBMED and PsycINFO. Although CAPES Periodicals intends to seek such databases, by accessing a specific database directly, it is possible to shape the search conditions to each, to reach a higher number of studies. Also, no gray literature texts (theses and dissertations) were consulted, which may contribute to amplification of the analyzes carried out herein.

Another possibility that was not contemplated in the present study is the revision encompassing personality and emotional evaluation systems. Although it was not an objective of this study, through the descriptive and qualitative analysis and mapping of systems that focus on other dimensions of psychological functioning, it would be possible to compare findings, contributing to verification of which systems allow greater inferences in certain populations, as well as associations with other tests. This review permits the identification of the empirical relations already performed with the cognitive version of HFD, allowing the replication of previous studies and the realization of new ones. Additionally, with regard to its practical applications, it has contributed to the professionals who can opt for the version that best suits their needs.

References

Abell, S., Horkheimer, R., & Nguyen, S. (1998). Intellectual evaluations of adolescents via Human Figure Drawings: An empirical comparison of two methods. *Journal of Clinical Psychology*, *54*(6), 811-815. https://doi.org/10.1002/(sici)1097-4679(199810)54:6%3C811::aid-jclp8%3E3.3.co;2-h

Alves, I. (1981). O Teste Goodenough-Harris em pré-escolares paulistanos. *Boletim de Psicologia*, *80*(33), 40-52.

Alves, I. C. B., Rosa, H. R., Da Silva, M. A., & Sardinha, L. S. (2016). Avaliação da inteligência: revisão de literatura de 2005 a 2014. *Avaliação Psicológica*, *15*(esp.), 89-97.

American Psychological Association. (2012). *Manual of the American Psychological Association* (6th ed). Washington, DC: White Plains.

*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. https://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.

*Bandeira, D. R., Arteche, A. X., & Costa, A. (2008). Estudo de validade do HFD como medida de desenvolvimento cognitivo infantil. *Psicologia: Reflexão e Crítica*, *21*(2). https://doi.org/10.1590/s0102-79722008000200020

*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 (São Paulo)*, *44*(3), 9-18.

*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

*Carreras, M. A., Uriel, F., & Fernández Liporace, M. (2013). Actualizaciones en el análisis de ítemes 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

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Colom, R., Flores-Mendoza, C. E., & Abad, F. J. (2007). Generational changes on the Draw-a-Man test: A comparison of Brazilian urban and rural children tested in 1930, 2002 and 2004. *Journal of Biosocial Science*, *39*(1), 79-89. https://doi.org/10.1017/s0021932005001173

Comparini, I. P., Wechsler, S. M., & Machado, W. de 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

*Dorazco-Valdes, J. (1968). Estudio Comparativo entre los Aspectos Clinicos, Electroencefalogrficos y la Prueba del Dibujo de la Figura Humana en el Nifio Epilrptico. *Journal of the Neurological Sciences*, *6*, 373-380. https://doi.org/10.1016/0022-510x(68)90103-2

*Flores-Mendoza, C. E., Abad, F. J., & Lelé, Á. J. (2005). Análise de itens do desenho da figura humana: Aplicação de IRT. *Psicologia: Teoria e Pesquisa*, *21*(2), 243-254. https://doi.org/10.1590/s0102-37722005000200015

*Garaigordobil, M., & Amigo, R. (2010). Inteligencia: diferencias de género y relaciones con factores psicomotrices, conductuales y emocionales en niños de 5 años. *Interdisciplinaria*, *27*(2), 229-245.

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. https://doi.org/10.1080/08856559.1926.10532353

*Haddad, F., & Juliano, J. M. (1991). Relations among scores on Matrix Analogies Test, Draw-A-Person, and the IOWA tests of basic skills for low socioeconomic children. *Psychological Reports*, *69*, 299-302. https://doi.org/10.2466/pr0.69.5.299-302

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

*Lange-Küttner, C., Küttner, E., & Chromekova, M. (2014). Deterioration and Recovery of DAP IQ Scores in the Repeated Assessment of the Naglieri Draw-A-Person (DAP) Test in 6- to 12-Year-Old Children. *Psychological Assessment*, *26*(1), 297-306. https://doi.org/10.1037/a0034581

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

*Merino, C. (2013). Test gestáltico Bender modificado y VMI-4: comparación de la validez incremental. *Psicoperspectivas*, *12*(1), 183-204. https://doi.org/10.5027/psicoperspectivas-vol12-issue1-fulltext-201

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

Naglieri, J. A., McNeish, T. J., & Bardos, A. N. (1991). *Darw A Person: Screening Procedure for Emotional Distrubance*. Austin: Pro-ed.

*Nielsen, H. H. (1961). Human figure drawings by normal and physically handicapped children: Draw-A-Person Test. *Scandinavian Journal of Psychology*, *2*. https://doi.org/10.1111/j.1467-9450.1961.tb01230.x

Noronha, A. P. P., Beraldo, F. N. M., & Oliveira, K. L. (2003). Instrumentos psicológicos mais conhecidos e utilizados por estudantes e profissionais de psicologia. *Psicologia Escolar e Educacional*, *7*(1), 47-56. https://doi.org/10.1590/s1413-85572003000100005

*Noronha, A. P. P., Santos, A. A. A., & Rueda, F. J. M. (2013). Habilidad Viso-motriz y Deficiencia Intelectual: Validez para el Bender-SPG. *Acta Colombiana de Psicología*, *16*(2), 115-123. https://doi.org/10.14718/acp.2013.16.2.11

*Oakland, T., & Dowling, L. (1983). The Draw-A-Person Test: Validity Properties for Nonbiased Assessment. *Learning Disability Qarterly*, *6*(4), 526-534. https://doi.org/10.2307/1510541

*Olatunya, O. S., Oke, O. J., Kuti, B. P., Ajayi, I. A., Olajuyin, O., Faboya, O. A., & Ajibola, A. (2017). Factors Influencing the Academic Performance of Children with Sickle Cell Anaemia in Ekiti, South West Nigeria. *Journal of Tropical Pediatrics*, 1-8. https://doi.org/10.1093/tropej/fmx034

Oliveira, K. da S., & Wechsler, S. M. (2016). Indicadores de criatividade no Desenho da Figura Humana. *Psicologia: Ciência e Profissão*, *36*(1), 6-19. https://doi.org/10.1590/1982-3703001682014

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

*Pihl, R. O., & Nimrod, G. (1976). The reliability and validity of the Draw-A-Person Test in iQ and personality assessment. *Journal of Clinical Psychology*, *32*(2). https://doi.org/10.1002/1097-4679(197604)32:2%3C470::aid-jclp2270320257%3E3.0.co;2-i

*Prewett, P. N., Bardos, A. N., & Naglieri, J. A. (1989). Assessment of mentally retarded children with the matrix analogies test-short form, draw a person: a quantitative scoring system, and the kaufman test of educational achievement. *Psychology in the Schools*, *26*. https://doi.org/10.1002/1520-6807(198907)26:3%3C254::aid-pits2310260306%3E3.0.co;2-c

Ramos Vosgerau, D. S., & Romanowski, J. P. (2014). Estudos de revisão: implicações conceituais e metodológicas. *Revista Diálogo Educacional*, *14*(42), 165-189. https://doi.org/10.7213/dialogo.educ.14.041.ds08

*Ritter, D. R., Duffey, J. B., & Fischman, R. (1974). Comparisons of the intellectual estimates of the Draw-A-Person Test, Peabody Picture Vocabulary Test, and Stanfard-Binet (L-M) for kindergarten chilgren. *Psychology in the Schools*. https://doi.org/10.1002/1520-6807(197410)11:4%3C412::aid-pits2310110408%3E3.0.co;2-b

*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. https://doi.org/10.11606/t.47.2006.tde-01102006-162329

*Rueda, F. J. M., Bartholomeu, D., & Sisto, F. F. (2006). Maturidade perceptual e inteligência. *Psicologia: Ciência e Profissão*, *26*(3), 490-503. https://doi.org/10.1590/s1414-98932006000300012

Sisto, F. F. (2005). *Desenho da Figura Humana – Escala Sisto*. São Paulo: Vetor.

Suehiro, A. C. B., Benfica, T. de S., & Cardim, N. A. (2015). Avaliação cognitiva infanti nos periódicos científicos brasileiros. *Psicologia: Teoria e Pesquisa*, *31*(1), 25-32. https://doi.org/10.1590/0102-37722015011755025032

Suehiro, A. C. B., Benfica, T. de S., & Cardim, N. A. (2016). Produção Cientí ca sobre o Teste Desenho da Figura Humana entre 2002 e 2012. *Psicologia: Ciência e Profissão*, *36*(2), 439-448. https://doi.org/10.1590/1982-3703000822014

Valentini, F., & Hauck Filho, N. (2014). Estabilidade da Estimação dos Parâmetros dos Itens em Diferentes Tamanhos Amostrais. In *CONBRATRI Congresso Brasileiro de Teoria da Resposta ao Item,* 1, 43-48. https://doi.org/10.26512/2015.02.d.18036

Wechsler, S. M. (2003). *HFD III. O Desenho da Figura Humana: Avaliação do desenvolvimento cognitivo de crianças brasileiras.* Campinas: Editora da Pontifícia Universidade Católica de Campinas. https://doi.org/10.17143/ciaed/xxiilciaed.2017.00248

*Wechsler, S. M., & Schelini, P. W. (2002). Validade do desenho da figura humana para avaliação cognitiva infantil. *Avaliação Psicológica*, *1*, 29-38.

*Woodburn, S. S., & Boschini, H. F. (1989). Relationships between the Goodenough-Harris Draw-A-Person Test and a body-image Test in Costa Rican children. *Perceptual and Motor Skills*, *68*, 1131-1136.

*Zubaran Santos, M., Sardá Junior, J. J., Menezes, M., & Thieme, A. L. (2013). Avaliação do desenvolvimento cognitivo de crianças com câncer por meio do HFD III. *Avaliação Psicológica*, *12*(3), 325-332.

Endereço para correspondência:

Ana Carolina Zuanazzi (Rua Waldemar César da Silveira, 105, Campinas, São Paulo, 13045-510; (43) 98841-6655; anacarolina.zf@gmail.com).

Autor I

Nome: Fabian Javier Marin Rueda

Titulação Acadêmica: Doutor

Afiliação Institucional: Universidade São Francisco (USF).

Autor II

Nome: Ana Paula Porto Noronha

Titulação Acadêmica: Doutora

Afiliação Institucional: Universidade São Francisco (USF).

Autor III

Nome: Acácia Angeli dos Santos

Titulação Acadêmica: Doutora

Afiliação Institucional: Universidade São Francisco (USF).

Autor IV

Nome: Ana Deyvis Santos Araujo Jesuíno

Titulação Acadêmica: Mestre

Afiliação Institucional: Universidade São Francisco (USF).

Autor V

Nome: Ana Carolina Zuanazzi

Titulação Acadêmica: Mestre

Afiliação Institucional: Universidade São Francisco (USF).

Autor VI

Nome: Adriana Satico Ferraz

Titulação Acadêmica: Graduada em Psicologia

Afiliação Institucional: Universidade São Francisco (USF).

Autor VII

Nome: Ariela Raissa Lima Costa

Titulação Acadêmica: Mestre

Afiliação Institucional: Universidade São Francisco (USF).

Autor VIII

Nome: Fernanda Otoni

Titulação Acadêmica: Graduada em Psicologia

Afiliação Institucional: Universidade São Francisco (USF).