

The Bender Gradual Scoring System for Assessment of People with Intellectual Disabilities

Fabián Javier Marin Rueda¹
*Universidade São Francisco,
Campinas-SP, Brazil*

Ana Deyvis Santos Araújo Jesuíno
*Universidade São Francisco,
Campinas-SP, Brazil*

Abstract: The Bender Gradual Scoring System (B-GSS) is one of the main systems of visual-motor assessment used in Brazil. Studies, however, seldom address samples with atypical development. This study tested the psychometric properties of the B-GSS items in a sample of individuals with typical development (Group A; $n = 198$; aged between seven and 10 years old; group application) and in a sample with intellectual disabilities (Group B; $n = 203$; aged between 11 and 30 years old; individual application), in accordance with ethical recommendations. The results reveal figures with differential functioning. Figure 6 was biased toward group A, while figures 1 and 2 were biased toward group B. Test of the information curve indicates low precision for extreme scores. There is a need for further research to better understand how the B-GSS items work across different samples in order to verify whether such findings can be generalized to other samples.

Keywords: psychological assessment, perceptual motor development, Bender-Gestalt Test

O Bender-Sistema de Pontuação Gradual na Avaliação de Pessoas com Deficiência Intelectual

Resumo: O Bender-Sistema de Pontuação Gradual (B-SPG) é um dos principais sistemas de avaliação visomotora utilizados no Brasil. Entretanto, estudos com amostras com desenvolvimento atípico são escassos. Este estudo verificou o funcionamento dos itens do B-SPG em uma amostra com desenvolvimento típico (Grupo A, $n = 198$, idade entre sete e dez anos, aplicação coletiva) e com pessoas com diferentes diagnósticos de deficiência intelectual (Grupo B, $n = 203$, idade entre 11 e 30 anos, aplicação individual) respeitando-se todos os aspectos éticos. Os resultados indicaram figuras com funcionamento diferencial. A figura 6 favoreceu o grupo A, e as figuras 1 e 2 favoreceram o B. A curva de informação do teste indicou baixa precisão nas pontuações extremas. Discute-se a realização de mais estudos, visando melhor compreensão do funcionamento do B-SPG em diferentes amostras de modo a testar o quanto os achados se generalizam a outras amostras.

Palavras-chave: avaliação psicológica, desenvolvimento perceptomotor, teste gestáltico de Bender

El Bender-Sistema de Puntuación Gradual en la Evaluación de Personas con Deficiencia Intelectual

Resumen: El Bender-Sistema de Puntuación Gradual (B-SPG) es uno de los principales sistemas de evaluación visomotora utilizados en Brasil. Sin embargo, los estudios con muestras con desarrollo atípico todavía son escasos. Este estudio verificó el funcionamiento de los ítems del B-SPG en una muestra de personas con desarrollo típico (Grupo A; $n = 198$; edades entre 7 y 10 años, aplicación colectiva) y con personas con diferentes diagnósticos de deficiencia intelectual (Grupo B; $n = 203$; edades entre 11 y 30 años, aplicación individual), respetando todos los aspectos éticos. Los resultados indicaron figuras con funcionamiento diferencial. La figura 6 favoreció al grupo A, mientras que las figuras 1 y 2 al Grupo B. La curva de información del test indicó una baja precisión en las puntuaciones extremas. Se debate la realización de más estudios que pretendan una mejor comprensión del funcionamiento del B-SPG en diferentes muestras, para comprobar si los resultados obtenidos en esta ocasión, se generalizan en otras muestras.

Palabras clave: evaluación psicológica, desarrollo perceptivo-motor, Bender-Gestalt Test

Between 1932 and 1938, based on the principles of Gestalt theory, Lauretta Bender developed a method called

the Bender Visual Motor Gestalt Test, which consists of Gestalt figures individuals are supposed to reproduce and is intended to assess the neurological maturation of children with typical and atypical development. The author used nine out of the 30 figures from Max Wertheimer, one of the founders of the theory, who sought to understand the origin of the perception of form among children (Bender, 1955).

¹Correspondence address: Universidade São Francisco. Faculdade de Ciências Humanas. Rua Waldemar César da Silveira, 105, Jardim Cura D'Ars (SWIFT), Campinas-SP, Brazil. CEP 13.045-510. E-mail: fabian.rueda@usf.edu.br

Perception and reproduction of Gestalt figures are determined by biological factors related to the sensory field and activity of the nervous system, which is reflected in the quality of drawings. Hence, there are differences among individuals given each individual's level of development. Motor development is thought to occur in parallel with mental development, and this integration of the organism is what determines a child's response (Kacero, 2013). Such ability develops better when there is motor stimulation, even in the presence of motor deficit, when compared to individuals who did not receive stimulus (Bonomo & Rossetti, 2010).

Bender's figures present a set of qualities considered to be sensible and organized. These figures are accompanied by a statement and are organized in fixed temporal sequences, requiring the individual to have notions of orientation (right or left; above or below; and vertical or horizontal, among others), contour (straight or curved, continuous or discontinuous), position (peripheral, central, upper or lower, among others), regularity (alignment, progression and equality) and opposition (concave-convex, encompassing-encompassed) (Kacero, 2013).

Visual motor skills are essential in the development of individuals, as these are associated with aspects such as visual perception, motor ability, language, memory, temporal and spatial concepts, and aspects like organization and representation skills (Bender, 1955). Visual motor skills involve four stages, which are characterized by perception of a stimulus, understanding of such stimulus, transformation of what is perceived, and expression through coordination of motor action (Kacero, 2013). This implies that poor performance on the Bender-Gestalt test may be a result of inappropriate functioning of one's visual perception or motor response, or a combination of both (Koppitz, 1989).

Even though the Bender-Gestalt test was disseminated worldwide due to its clinical focus and proposal of assessing one's visual motor skill, it was not conceived with a scoring system in mind. It encouraged researchers to create systems to establish standards to score it (Brannigan & Brunner, 1989; Koppitz, 1989; Pinelli Júnior & Pasquali, 1991-1992; Posada, 2002; Santucci & Pêcheux, 1981). In this sense, it is important to note that the system proposed by Koppitz (1989) was one of the most frequently used around the world and consisted of a dichotomous response (presence or lack) considering the distortion of the form, rotation, or yet, a lack of integration among parts or perseveration (Sisto, Noronha, & Santos, 2005).

Some studies report problems accepting the Koppitz system in Brazil. Among these studies, that of Pinelli Júnior and Pasquali (1991-1992) identified the presence of four factors rather than a single-factor structure as originally proposed. Sisto, Noronha and Santos (2004), in turn, report a lack of differentiation between ages for all the Bender figures, while Bartholomeu, Rueda and Sisto (2005) note a deficiency of the Koppitz system in discriminating between children with and without learning problems when considering written performance.

Given these findings and because the Bender-Gestalt test does not meet the minimum requirements imposed by the SATEPSI (Psychological Testing System), it is no longer used in Brazil. After years of study, the Bender Gradual Scoring System (B-GSS) was presented. Similar to the Koppitz system (1989), it assesses visual motor skills, but differs in regard to the scoring system proposed (Sisto et al., 2005), as it considers only the distortion of the form and establishes gradual scoring for errors instead of the dichotomy proposed by Koppitz.

From the time this new system was first presented, various studies have adopted the B-GSS in Brazil. Such studies include those that relate it to different measures of intelligence (Bartholomeu, Cecato, Montiel, Machado, & Sisto, 2012; Sisto, Bartholomeu, Rueda, Santos, & Noronha, 2008), to attention (Sousa & Rueda, 2017), and to reading comprehension and writing (Carvalho, Noronha, Pinto, & Luca, 2012; Suehiro, Santos, & Rueda, 2015). Studies addressing variables such as age and education (Pinto & Noronha, 2013) were also found, along with studies addressing samples from different Brazilian regions (Noronha, Rueda, & Santos, 2013), including international samples (Santos, Noronha, Rueda, & Segovia, 2014). More recently, a version designed for screening purposes that considers only three figures of the B-GSS was developed (Rueda, Sousa, Santos, & Noronha, 2016).

Note that despite the number of different studies adopting the instrument in Brazil, only two addressed a sample of individuals with intellectual disabilities, which is the focus of this study: those conducted by Noronha, Santos and Rueda (2013) and Pacanaro, Santos and Suehiro (2008). The first study compares the results of 290 intellectually disabled individuals aged between six and 24 years old, with data concerning six-year-old children presented in the normative tables of the B-GSS manual (Sisto et al., 2005). The mean of errors of the individuals with intellectual disability was 15.4, while that of children from the normally-abled sample of the manual was 8.54, while the maximum score is 21 points. Note that the scores of the disabled individuals were higher than the scores of the normally-abled sample in almost all the B-GSS figures (Noronha, Santos, et al., 2013). Therefore, the conclusion is that the performance of intellectually disabled individuals was less than that of children with typical development.

The study by Pacanaro et al. (2008) was intended to assess visual motor and intellectual skills in 51 individuals diagnosed with Down Syndrome aged between six and 24 years old. In regard to the B-GSS, the results show a mean score of 18.7, while 35.3% of the participants scored the maximum of 21 errors. The lowest score was eight, obtained by 1.8% of the individuals, indicating that most participants had compromised visual motor skills. The authors suggest that future studies should verify differential patterns of development in individuals with Down Syndrome using more refined statistical analysis to verify Differential Item Functioning (DIF) in the B-GSS.

An item is considered to display DIF when individuals with the same level of latent trait (skill being assessed) obtain

different scores in the very same item (Kamata & Vaughn, 2004). An item does not present DIF when assessing a given skill in groups of a distinct nature, if the same percentage of success is obtained regardless of the group (Tennant & Pallant, 2007). Note that at least 500 individuals are necessary to ensure greater reliability of data (Valentini & Hauck Filho, 2013).

Only two studies were found that verified DIF in B-GSS. Both addressed children with typical development, namely, the studies conducted by Santos et al. (2014) and Sisto, Santos and Noronha (2010). The first compared 123 Brazilian children with 108 Peruvian children aged between six and 10 years old. Analysis indicated that the goodness of fit indexes were within the expected range. DIF was also verified in figures 1 and 4. Figure 1 was biased toward the Brazilian children and Figure 4 was biased toward the Peruvian children. When, however, DIF Contrast was verified, i.e., if there was difference between the difficulty parameters of the item in each group, the authors verified that the values canceled themselves out and did not affect the total score. It was also verified that the *theta* level in some Peruvian children presented a poor performance ($\theta > 3$), while some children from the Brazilian sample performed better than expected ($\theta > -4$) (Santos et al., 2014).

The study by Sisto et al. (2010), in turn, intended to verify the DIF in the B-GSS according to sex. A total of 1,052 children, 539 girls and 503 boys, aged between six and 10 years old took part in the study. The analyses indicated that figures 5 and 6 worked differently in the groups. Boys scored higher on item 5, while girls scored higher on item 6. The authors stress that cancellation of items reached a value different from zero, indicating that girls scored higher. The conclusion is that the scores obtained may be either a result of specific skills of the participants or due to the functioning of items (Sisto et al., 2010).

Due to the aspects presented here, this study's aim was to verify the B-GSS items and whether they work differently for people with intellectual disabilities when compared to children with typical development. Thus, this study is expected to expand the use of B-GSS and contribute to knowledge concerning visual motor skills in individuals with disabilities.

Method

Participants

A total of 401 people, 198 of whom had no known diagnosis of intellectual disability, (Group A) from four public schools from a city in the interior of Minas Gerais and 203 people with different diagnoses of intellectual disabilities (Group B) from five special education institutions located in the state of São Paulo. Age in group A ranged from seven to 10 years old ($M = 8.31$; $SD = 1.12$), 52.5% of whom were female. In group B, age ranged from 11 to 30 years old ($M = 14.57$; $SD = 3.55$), and 62.6% were males. Therefore, 50.6% of the

total sample had some intellectual disability, 160 of whom (78.81%) presented an intellectual disability, while 43 (28.19%) presented Down syndrome associated with another type of intellectual disability. Note that this sample was not random, so that all those whose legal guardians consented took part in the study. No screening was performed to select the sample with disability, since the individuals had been previously diagnosed by psychologists or psychiatrists, as this is a condition for admission into the institution. Specifically, a psychologist diagnosed more than 90% of the individuals.

Instruments

Bender Visual Motor Gestalt Test: Gradual Scoring System (B-GSS) (Sisto et al., 2005). The B-GSS has nine figures and is used to assess children aged from six to 10 years old. It can be applied individually or collectively. A projector is needed when applying the test in groups and a maximum of 30 children is recommended. Even though there are no restrictions regarding the duration of the test, 15 minutes, as an average, is recommended. The test is applied with a blank sheet and a pencil. The participants are asked to reproduce the figures as similar as possible to the original.

Distortions in the form are a criterion used for scoring so that higher scores represent greater distortions that, consequently, represent worse performance. The total score ranges from 0 to 21, while the score of each figure ranges from zero to two, except for figure six, the score of which ranges from zero to three. The instrument's manual presents studies reporting evidence of validity and precision measures ranging from 0.77 to 0.80.

Procedures

Data collection. The institutions were contacted and, after their authorization was granted, the project was submitted to and approved by the Institutional Review Board. In the following stage, the parents of participants or their legal guardians signed free and informed consent forms. The instrument was applied differently in each group. In the schools, the instrument was applied collectively, with a maximum of 25 students per classroom and lasted approximately 20 minutes. Two volunteers, both psychology students, previously trained in regard to the procedures, aided the application in this group. The data collection in group B was individual and took place in a room on the premises of the special education institution on a date and at a time previously scheduled. Initially, information was collected from the medical files provided by the institution in order to characterize the sample, as described in the section 'Participants', after which the tests were applied with a duration of approximately 40 minutes. Note that instructions were repeated until the participants understood the task at hand.

Data analysis. Data were initially descriptively analyzed and unidimensionality was later verified by analyzing the principal components of the model's residuals, considering a

value less than two (Linacre, 2011). Next, multivariate analysis was performed. The objective of the first stage was to verify whether the items fitted the Partial Credit model (Masters, 1982), a derivation of the Rasch model for polytomous items, which admits different systems of answers for each item. This model is ideal in this case because nine of the 10 Bender figures had a three-point scoring system, while one of them (Figure 6) is scored on a four-point scale. The model enables the estimation of difficulties related to each of the categories of responses to stimuli, information that, in this case, represents level of skill or cognitive load necessary to obtain a given score on the scoring scale. Additionally, the model provides a general estimate of the difficulty of each item or stimulus. For the purpose of model identification, the average *logit* skill scale was set at 0.0 and standard deviation at 1.

The goodness of fit indexes were verified as to whether the participants and items were appropriate for the model. The types most frequently used in this model are *infit* and *outfit*. The first refers to disagreements that occur near the individual's level of latent trait, while an unexpected larger amount of correct or incorrect answers implies a higher index. *Outfit* also refers to unexpected patterns of answers. However such a pattern occurs when the difficulty of items and the individual's *logit* level of skills are very different. It occurs when people with very low skill levels correctly answer very difficult items, as well as when people with very high *logit* wrongly answer items with low difficulty (Nakano, Primi, & Nunes, 2015). For the items to be appropriate to the model, both *infit* and *outfit* should be in an interval between 0.8 - 1.5 (Linacre, 2014).

In addition to the goodness of fit indexes, the thresholds of the items were also verified to better identify the probability of one being classified in the categories of answers according to the variation of the measure (*logit*). The threshold is the point of transition between two categories of answers on a Likert scale and indicates the point at which there is a probability of 50% of individuals with a Y degree of skill choosing a given category of answer (Linacre, 2011).

The differential functioning of stimuli between individuals with and without a diagnosis of some intellectual disability was also verified. The differential functioning of items was assessed using the Mantel test - Winsteps software (Linacre, 2011). This analysis produces a *p*-value that tends to be significant when difficulties of endorsing categories of items vary between groups. Winsteps also provides the DIF contrast information, which is a direct measure of the difference between values of the parameter for difficulty of the item in each group. Outstanding differences are those greater than 0.64 (Linacre, 2011).

The information curves, which refer to the performance of the individuals on the test, were also assessed. This curve indicates for which individuals, when the *logit* skill scale is considered, the test provides the most precise scores, and therefore, fewer errors. The information curve of the total sample was produced using the Psych package (Revelle, 2014) from the R software. Finally, the frequency of *logit* of skill in the sample was also verified.

Ethical Considerations

The project was submitted to the Institutional Review Board at the Universidade São Francisco, in accordance to the Resolution 196/96, National Council of Health (CAAE - 20460113600005514) and, only after its approval, was data collection initiated.

Results

Unidimensionality was initially verified by analyzing the principal components of the model residuals with value equal to 1.8, which is considered to be appropriate (Linacre, 2011). Afterwards, the items' goodness of fit indexes and the probability of each category of answer was verified. The results concerning analysis of the indexes are presented according to the difficulty of items in Table 1.

Table 1 shows that figures 1, 8 and 7B of B-GSS did not present appropriate *infit* and *outfit* indexes. In turn, figures 6, 7A and 2 do not present appropriate *outfit* indexes (Linacre, 2011). The thresholds were appropriate, as individuals with lower levels of motor skills scored higher, while individuals with greater skills scored lower (Andrich, 2013); the Bender test scores errors.

Note that in this sample, figure A was the easiest to duplicate, the one the participants scored lowest, indicating less distortion in the form in its reproduction. In turn, figure 3 was the one that most frequently presented distortions and was considered the most difficult one in the B-GSS. Additionally, five out of the 10 figures were considered difficult based on the value found in Measure.

Afterwards, DIF analysis was performed. The results are presented in Table 2, in which the estimates for the difficulty of each item for individuals, both with a diagnosis and without a diagnosis, are presented.

Table 2 shows that three out of the 10 items (1, 6 and 2) presented a DIF Contrast greater than 0.64, which is considered a value of outstanding magnitude and a probability less than 0.05 (Linacre, 2011). The sum of the DIF contrast of items was considered, which indicates the extent to which the differential functioning of each item contributes to the DIF in the instrument's total score. When zero is the result of this sum, values cancel each other out, indicating that even though the items work differently, the total score is not affected by them, because in this case one item compensates for the other, canceling each other (Stark, Chernyshenko, & Drasgow, 2004; Teresi, 2006). This was not the situation in this study, as the value found was 0.68. Thus, the presence of DIF in the items interfered in the instrument's total score.

Finally, the information curve of the total sample was analyzed. This enables the level of the latent trait for which the instrument is more precise to be identified. The result is presented in Figure 1.

Table 1
Item Adjustment and Difficulty of Items Statistics for the B-GSS

Item	Measure	Error	Infit Mnsq	Outfit Mnsq	Correlation	δ_1	δ_2	δ_3
Figure A	1.74	0.09	1.24	1.51	0.68	1.42	2.07	–
Figure 1	1.06	0.08	1.65	1.97	0.57	0.75	1.38	–
Figure 5	0.40	0.09	1.02	0.98	0.67	-0.51	1.31	–
Figure 6	0.19	0.07	0.83	0.79	0.78	-1.44	0.59	1.44
Figure 8	0.01	0.08	0.67	0.56	0.74	-0.49	0.50	–
Figure 7B	-0.03	0.08	0.66	0.56	0.74	-0.53	0.46	–
Figure 7A	-0.43	0.08	0.81	0.68	0.67	-0.98	0.12	–
Figure 4	-0.53	0.10	0.95	0.94	0.65	-1.86	0.81	–
Figure 2	-0.89	0.09	1.44	1.59	0.43	-1.88	0.09	–
Figure 3	-1.52	0.10	0.97	0.92	0.53	-2.25	-0.78	–

Table 2
Differential Item Functioning and Statistical Comparisons for 10 items

Item	Diagnosis		No Diagnosis		Contrast	Joint SE	Mantel
	DIF Measure	SE	DIF Measure	SE			
Figure A	1.66	0.12	1.87	0.16	-0.21	0.20	0.015
Figure 1	1.66	0.11	0.54	0.10	1.12	0.16	0.002
Figure 2	0.05	0.14	-1.53	0.13	1.59	0.19	0.001
Figure 3	-1.13	0.17	-1.72	0.13	0.59	0.21	0.004
Figure 4	-0.56	0.15	-0.50	0.12	-0.06	0.20	0.364
Figure 5	0.11	0.14	0.61	0.11	-0.49	0.18	0.029
Figure 6	-0.37	0.11	0.61	0.10	-0.97	0.15	0.001
Figure 7A	-0.33	0.14	-0.48	0.10	0.15	0.17	0.032
Figure 7B	-0.43	0.14	0.18	0.10	-0.61	0.18	0.022
Figure 8	-0.29	0.14	0.17	0.10	-0.45	0.17	0.255

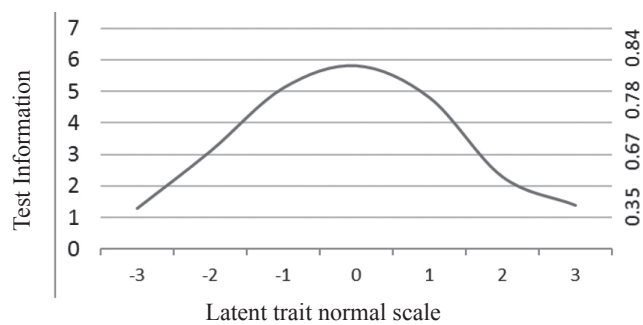


Figure 1. Test information curve in the full sample (logit skill scale = cognitive problems).

Figure 1 shows that the information curve of the total sample is with the apex slightly concentrated to the left (ranging approximately between - 1.0 and 0.5), the precision index of which is approximately 0.84. When extreme scores, whether correct or incorrect, are analyzed, a decline

is verified in the test's precision index. This can be verified by observing, for instance, that at a level of skill -3, precision stays around 0.30; while at a level 3, precision falls below 0.20. When the frequency of the *logit* skill scale is verified in the total sample, a variation from - 4.53 to 4.12 was observed.

Discussion

Initially, data showed unidimensionality, with the most expressive residual being 1.8, indicating the test measures were appropriate (Andrich, 2013). When goodness of fit and difficulty of items of the B-GSS were verified, six out of the 10 items did not present goodness of fit between 0.8 and 1.5, which is the range recommended by Linacre (2014). Nonetheless, as in the study conducted by Sisto et al. (2005), the item-total correlations were greater than 0.30, which is a value considered sufficient to avoid discarding items

(Guilford & Fruchter, 1978). Additionally, figures A and 5 are among the easiest figures, while figures 4, 2 and 3 are the most difficult ones, a fact that is also pointed out in the manual of the B-GSS (Sisto et al., 2005).

DIF analysis showed that three figures (1, 2 and 6) presented a DIF contrast of outstanding magnitude (> 0.64 ; $p < 0.05$), based on the assumptions of Linacre (2011). The results found corroborate those reported by Santos et al. (2014) and Sisto et al. (2010) in regard to figures 1 and 6, though those studies addressed the nationality and sex of the participants, respectively. This may be explained by the fact that 10-year old children have not yet achieved their maximum level of motor development. Such a fact was evidenced by Bender (1955), when she applied the test in 1932 to 800 children aged from 3 to 11 years old and found that most children managed to redraw the figures without errors only after the age of 11. Note that the age of this study's sample of typical development is from seven to 10 years old.

When verifying the values of the DIF contrast of figures 1, 2 and 6, we verified that figures 1 and 2 were biased toward group B, while figure 6 was biased toward group A. A closer look reveals that both figures are horizontally disposed and marked by continuity-discontinuity (Kacero, 2013). The author states that the distance between the points in figure 1, as well as the way children closed the figure, by repeating going over the strokes until the circle collapsed. compressing or extending the figure, and the distance between points, among other factors, may indicate some injury compromised one's motor skills, or even a more primitive situation, such as some problem in the development of the child when growing up.

Figure 2, in turn, is considered a little more complex than Figure 1 due to its inclination and execution, because it is based on the principle of proximity of parts. Kacero (2013) considers that little children (six-year-olds) can make circles, as is the case in Figure A. It is believed that Figures 1 and 2, even though the action used in the drawing is similar to the one used in Figure A, require more visual motor skills that children from group A perhaps had not yet acquired because their motor development was still progressing. Associated with this is the fact that the individuals in the group B, despite their disability, were stimulated through the manual activities performed in the institutions and at home, which may have contributed to a reduced deficit in comparison to group A, as reported by Bonomo and Rossetti (2010); group B may actually have reproduced the figures with fewer errors.

Figure 6, in relation to figures 1 and 2, requires greater ability because it demands a rhythmic movement with undulations, being the first, among the 10 figures, not to demand drawing a figure similar to a circle. The link between the elements of this figure occurs in the format of a cross. The dissociative mechanisms may be conditions that generate notable differentiation and may contribute to a greater presence of errors in the figure (Kacero, 2013). Such a fact may help explain why this figure was better reproduced by group A,

which may include children with sufficient visual motor skills to reproduce the figure compared to those with disabilities, given the demand for rhythmicity in its execution.

The performance of the participants on the B-GSS was verified through the information curve. The total sample presented good precision indexes but it is important to note that the instrument is more precise when applied to individuals with an intermediate level of distortion compared to individuals who have a high number of either correct or incorrect answers. This may be because the test was developed for children with typical development. The instrument does not work properly when high scores are caused by some disability or low scores are obtained because children are older than 11 years old, or have above average development. It means the instrument is able to discriminate better among individuals aged between six and 10 years old with typical development.

The results suggest that B-GSS works in a differentiated manner: two items favor individuals with disabilities, while one item favors six- to 10-year-old children. Note that these results may have been influenced by the fact that these individuals with intellectual disabilities received stimuli and, as noted by Bonomo and Rossetti (2010), stimuli may decrease motor deficits. It is also important to take into account factors such as the age of the participants and sample size. The group of people with disabilities was aged between 11 and 30 years old, while the age recommended by the B-GSS is between six and 10. Still in regard to the sample, note that the number of participants weakens the reliability of results and a sample of at least 500 individuals would be necessary to obtain stable estimates (Valentini & Hauck Filho, 2013). This is a fact to take into account, because the sample of this study was smaller than that indicated by other studies, which may have influenced the DIF results.

There is also a limitation concerning a lack of data regarding the diagnosis of the individuals with intellectual disabilities. More accurate information could have contributed to selecting more homogeneous groups for analysis. This lack of information also impeded verifying potential variations in the sample regarding levels of intelligence and cognitive development.

Considering such limitations, further studies addressing samples composed of individuals with intellectual disabilities are needed, either to corroborate or confirm the results presented here, as well as studies obtaining normative data for the population addressed here. Information curves of the items should also be verified to identify which items influence the test's information curve in order to improve its precision in samples with intellectual disabilities. It is expected to improve the instrument in order to provide more accurate information for those with atypical development.

Finally, in addition to this study's information, there is the possibility of the B-GSS being used to screen children with suspected intellectual disability. This type of diagnosis is usually established at a clinic using more expensive instruments that require more time, as is the case of the Wechsler Scales.

In this sense, future studies can contribute to improved information from the B-GSS in specific samples. Additionally, the screening version proposed by Rueda et al. (2016) can also be addressed in samples with atypical development.

References

- Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous rasch model that dispels any threshold disorder controversy. *Educational and Psychological Measurement*, 73(1), 78-124. doi:10.1177/0013164412450877
- Bartholomeu, D., Cecato, J. F., Montiel, J. M., Machado, A. A., & Sisto, F. F. (2012). Teste de Bender (B-SPG) e DFH-Escala Sisto: Validade por grupos contrastantes [Bender Test (B-SPG) and HFD-Escala Sisto: Validity by contrasting groups]. *Estudos Interdisciplinares em Psicologia*, 3(2), 241-257. doi:10.5433/2236-6407.2012v3n2p241
- Bartholomeu, D., Rueda, F. J. M., & Sisto, F. F. (2005). Teste de Bender e dificuldades de aprendizagem: Quão válido é o sistema Koppitz? [Bender Test and learning disabilities: What is the Koppitz system valid for?]. *Avaliação Psicológica*, 4(1), 13-21.
- Bender, L. (1955). *Test Gestáltico Visomotor: Usos y aplicaciones clinicas* [Visomotor Gestalt Test: Clinical uses y applications] (6th ed.). Buenos Aires, Argentina: Paidós.
- Bonomo, L. M. M., & Rossetti, C. B. (2010). Aspectos percepto-motores e cognitivos do desenvolvimento de crianças com síndrome de Down [Aspects in perceptual-motor development and sensory-motor intelligence in Down syndrome]. *Revista Brasileira de Crescimento e Desenvolvimento Humano*, 20(3), 723-734.
- Brannigan, G. C., & Brunner, N. A. (1989). *The modified version of the Bender-Gestalt Test for preschool and primary school children*. Brandon, VT: Clinical Psychology Publishing.
- Carvalho, L., Noronha, A. P. P., Pinto, L. P., & Luca, L. (2012). Maturidade perceptomotora e reconhecimento de palavras: Estudo correlacional entre o Bender – Sistema de Pontuação Gradual e o Teste de Reconhecimento de Palavras [Perceptive-motor maturity and word recognition: A correlational study between Bender – Scoring System for Grades and Word Recognition test]. *Estudos de Psicologia (Campinas)*, 29(3), 371-377. doi:10.1590/S0103-166X2012000300007
- Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. New York, NY: McGraw-Hill.
- Kamata, A., & Vaughn, B. K. (2004). An introduction to differential item functioning analysis. *Learning Disabilities: A Contemporary Journal*, 2(2), 49-69.
- Kacero, E. (2013). *Test Gestaltico Visomotor de Bender: Uma puesta em espacio de figuras* [Bender visual-motor Gestalt Test: A mise in space figures]. Buenos Aires, Argentina: Lugar Editorial.
- Koppitz, E. M. (1989). *O Teste Gestáltico de Bender para crianças* [The Bender Gestalt Test for children]. Porto Alegre, RS: Artes Médicas.
- Linacre, J. M. (2011). *A user's guide to Winsteps, Program Manual 3.74.0*. Chicago, IL: Winsteps.com. Retrieved from <http://www.winsteps.com/index.htm>
- Linacre, J. M. (2014). *Winsteps Rash measurement computer program*. Beaverton, OR: Winsteps.com. Retrieved from <http://www.winsteps.com/index.htm>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. doi:10.1007/BF02296272
- Nakano, T. C., Primi, R., & Nunes, C. H. S. S. (2015). Análise de itens e Teoria de Resposta ao Item (TRI) [Item analysis and item response theory - IRT]. In C. S. Hutz, D. R. Bandeira, & C. M. Trentini (Orgs.), *Psicometria [Psychometry]* (pp. 97-124). Porto Alegre, RS: Artmed.
- Noronha, A. P. P., Rueda, F. J. M., & Santos, A. A. A. (2013). Bender Gestalt Visual-Motor test - Sistema de Pontuação Gradual (B-SPG): A study with different samples. *Paidéia (Ribeirão Preto)*, 23(55), 179-185. doi:10.1590/1982-43272355201305
- Noronha, A. P. P., Santos, A. A. A., & Rueda, F. J. M. (2013). Habilidade viso-motriz y deficiencia intelectual: Estudio de validez para el Bender-SPG [Visual motor ability and intellectual disabilities: Validity study of the Bender-SPG]. *Acta Colombiana de Psicología*, 16(2), 115-123. doi:10.41718/ACP.2013.16.2.11
- Pacano, S. V., Santos, A. A. A., & Suehiro, A. C. B. (2008). Avaliação das habilidades cognitiva e viso-motora em pessoas com Síndrome de Down [Assessment of cognitive and visuomotor abilities in people with Down Syndrome]. *Revista Brasileira de Educação Especial*, 14(2), 293-310. doi:10.1590/S1413-65382008000200011
- Pinelli Júnior, B., & Pasquali, L. (1991-1992). Parâmetros psicométricos do teste Gestáltico Viso-motor de Bender: Um estudo empírico [Psychometric parameters of the Bender-Gestalt test: An empirical study]. *Revista de Psicologia*, 9-10(1-2), 51-74. Retrieved from http://www.repositorio.ufc.br/ri/bitstream/riufc/10976/1/1992_art_bpinelijunior.pdf
- Pinto, L. P., & Noronha, A. P. P. (2013). Bender–Sistema de Pontuação Gradual (B-SPG): Análise da maturação percepto-motora de crianças [Bender-Gradual Scoring System - B-SPG: Analysis of the maturation of visual-motor perception of children]. *Interação em Psicologia*, 17(3), 281-289. doi:10.5380/psi.v17i3.29683

- Posada, M. C. (2002). Un sistema alternativo para la evaluación del Test Gestáltico Viso-motor de Bender [An alternative system for the evaluation of visual-motor Bender Gestalt Test]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 14(2), 9-31. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=4539931>
- Revelle, W. (2014). *Package psych: Procedures for Personality and Psychological Research*. R package version 1.4.3. CRAN Project. Retrieved from <http://cran.r-project.org/web/packages/psych/psych.pdf>
- Rueda, F. J. M., Sousa, V., Santos, A. A. A., & Noronha, A. P. P. (2016). Bender - Sistema de Pontuação Gradual - B-SPG: Estudo para versão de rastreio [Bender - Gradual Scoring System - B-SPG: Study for screening version]. *Psicologia: Teoria e Prática*, 18(2), 117-128. doi:10.15348/1980-6906/psicologia.v18n2p117-128
- Santos, A. A. A., Noronha, A. P., Rueda, F. J. M., & Segovia, J. L. (2014). Bender-gradual scoring system: Performance of Brazilian and Peruvian children. *Perceptual & Motor Skills*, 118(3), 897-908. doi:10.2466/03.10.PMS.118k25w7
- Santucci, H., & Pêcheux, M. G. (1981). Prova gráfica de organização perceptiva para crianças de 6 a 14 anos [Proof perceptual organization graphic for children 6-14 years]. In R. Zazzo (Org.), *Manual para o exame psicológico da criança* [Manual for the psychological examination of the child] (pp. 291-338). São Paulo, SP: Mestre Jou.
- Sisto, F. F., Bartholomeu, D., Rueda, F. J. M., Santos, A. A. A., & Noronha, A. P. P. (2008). Relações entre os Testes de Bender e Matrizes Progressivas Coloridas de Raven na avaliação da inteligência [Relationship between Bender and Raven's Colored Progressive Matrices in intelligence assessment]. *Interação em Psicologia*, 12(1), 11-19. doi:10.5380/psi.v12i1.10546
- Sisto, F. F., Noronha, A. P. P., & Santos, A. A. A. (2004). Distorção de forma no Teste de Bender: Questionando seu critério de validade [Form distortion of Bender Test: Questioning its validity criterion]. *Revista do Departamento de Psicologia - UFF*, 16(2), 139-153.
- Sisto, F. F., Noronha, A. P., & Santos, A. A. A. (2005). *Manual Bender - Sistema de Pontuação Gradual - B-SPG* [Manual Bender - Gradual Scoring System - B-SPG]. São Paulo, SP: Vetor.
- Sisto, F. F., Santos, A. A. A., & Noronha, A. P. P. (2010). Differential functioning of Bender Visual-Motor Gestalt Test items. *Perceptual and Motor Skills*, 110(1), 313-322. doi:10.2466/PMS.110.1.313-322
- Sousa, V., & Rueda, F. J. M. (2017). A relação entre a habilidade perceptomotora e atenção [The relationship between perceptual motor skills and attention]. *Paidéia (Ribeirão Preto)*, 27(66), 24-32. doi:10.1590/1982-432727662017046201704
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology*, 89(3), 497-508. doi:10.1037/0021-9010.89.3.497
- Suehiro, A. C. B., Santos, A. A. A., & Rueda, F. J. M. (2015). Desenvolvimento perceptomotor e escrita em crianças do ensino fundamental [Perceptual motor development and writing in children of elementary school]. *Psicologia Escolar e Educacional*, 19(2), 369-376. doi:10.1590/2175-3539/2015/0192861
- Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions*, 20(4), 1082-1084. Retrieved from <http://www.rasch.org/rmt/rmt204d.htm>
- Teresi, J. A. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care*, 44(11, Suppl. 3), S152-S170. doi:10.1097/01.mlr.0000245142.74628.ab
- Valentini, F., & Hauck Filho, N. (2013). Estabilidade da estimação dos parâmetros dos itens em diferentes tamanhos amostrais. *CONBRATRI Congresso Brasileiro de Teoria da Resposta ao Item*, (1), 43-48. Retrieved from <http://abave.com.br/ojs/index.php/conbratri/article/view/78>

Fabián Javier Marin Rueda is a Professor of the Universidade São Francisco.

Ana Deyvis Santos Araújo Jesuíno is a Ph.D. candidate from the Graduate Program in Psychology of the Universidade São Francisco.

Received: Jun. 23, 2016

1st Revision: Nov. 7, 2016

Approved: Dec. 2, 2016

How to cite this article:

Rueda, F. J. M., & Jesuíno, A. D. S. A. (2018). The Bender gradual scoring system in assessment of people with intellectual disabilities. *Paidéia (Ribeirão Preto)*, 28, e2807. doi: <http://dx.doi.org/10.1590/1982-4327e2807>