BRAIN DYSFUNCTION INDICATION ON THE BENDER-GESTALT TEST:
A VALIDATION OF THE EMBREE/BUTLER SCORING SYSTEM

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Louise Henderson, B.S.
Denton, Texas
December, 1981
Henderson, Louise, Brain Dysfunction Indication on the Bender-Gestalt Test: A Validation of the Embree/Butler Scoring System. Master of Science (Clinical Psychology), December, 1981, 19 pp., 1 figure, references, 18 titles.

The Embree/Butler scoring system served as criterion for ascertaining brain dysfunction on the protocols of 100 subjects--50 had been diagnosed by health professionals as having brain dysfunction, and 50 had been diagnosed as having no brain dysfunction. In comparing the hospital's diagnoses with those of the Embree/Butler method, the data strongly supported the hypothesis that the Embree/Butler scoring system did effectively discriminate (chi square of 77.99 < .01) between those with organic brain syndrome (or cerebral dysfunction) and those with psychiatric classification.

A point-biserial correlation was used to distinguish the relationship between diagnosis and the score. A cutoff score of above 14 produced the least false-negative or false-positive evaluations.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>iv</td>
</tr>
<tr>
<td>Thesis</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Method</td>
<td>5</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td></td>
</tr>
<tr>
<td>Procedure</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>7</td>
</tr>
<tr>
<td>Discussion</td>
<td>8</td>
</tr>
<tr>
<td>Appendices</td>
<td>13</td>
</tr>
<tr>
<td>References</td>
<td>17</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Error Scores on the Bender-Gestalt Test Utilizing the Embree/Butler Scoring System</td>
</tr>
</tbody>
</table>
BRAIN DYSFUNCTION INDICATION ON THE BENDER-GESTALT TEST:
A VALIDATION OF THE EMBREE/BUTLER SCORING SYSTEM

There has been a need for better techniques in discriminating between brain dysfunction, or organic brain syndrome, and functional psychiatric disorders. An instrument that has become widely recognized as a valuable tool in the diagnosis of neurological impairment has been the Visual Motor Gestalt Test (Bender, 1938), which was developed at Phipps Psychiatric Clinic of the Johns Hopkins Hospital. This study utilized subjects diagnosed as schizophrenic. The Bender-Gestalt Test has been reported to be the third or fourth most frequently used psychological instrument available to the diagnostician (Sundberg, 1961). Since then, a plethora of research has been done with this test to determine if there were indeed limitations to its effectiveness.

Billingslea (1948) was one of the first to devise an objective scoring system for the Bender-Gestalt Test. Using 39 factors and 137 indices, his scoring system was used to test neurotic males. Although this system was altogether unmanageable, it stimulated others, such as Peek and Quast (1951), to attempt different methods of scoring. Their system used 17 attributes to discriminate between the performance of a suspected brain-damaged group and a suspected emotionally disturbed group. Pascal and Suttell...
(1951) proposed a methodology which yielded a total score for each of eight designs for each subject. Since the method was restricted to subjects in the age range 15-50, with at least 1 year of high school education, the system has proven inapplicable to many psychiatric patients. However, the Pascal and Suttell system did survive cross-validation in differentiating organic from nonorganic patients (Ascough & Dana, 1962).

In 1953, Gobetz was interested in distinguishing between neurotics and a normal population and was successful in cross-validating his system, although his work was done on a specific set of subgroups which limited generalizability. Hain (1964) was the first to develop a scoring system with the main purpose of objectively distinguishing between groups of brain-damaged, psychiatric, and normal subjects. Hain scored protocols by the use of discriminative weights of signs derived from an original sample, which he found to discriminate between brain-damaged and non-brain-damaged subjects.

A survey of published research indicated that Bender records of groups of brain-damaged patients, regardless of age and intelligence, differed significantly from those of non-brain-damaged subjects who were not psychiatric patients. Koppitz (1962) proposed a set of brain-injury indicators comprised of those items from the Developmental Scoring System on which brain-injured and control subjects differed.
significantly. She also presented a series of emotional indicators for assessing emotional problems. McConnell (1967) compared Bender protocols of subjects who were simultaneously emotionally disturbed and brain-damaged to varying degrees. Mehlman and Vatovec (1956) utilized three clinical judges who were considered authorities, using the Bender-Gestalt Test. Their results showed identification of organic and nonorganic to be only just slightly better than chance. Goldberg (1959) had correct diagnosis from 57% to 77% of the time, using Goldberg's Objective Index with an optimal cutting score.

Brilliant and Gynther (1963) compared the accuracy of predicting organicity using the Bender-Gestalt Test, Benton Visual Retention Test, and Graham-Kendall Memory-for-Designs. Results for 120 patients classified as Chronic Brain Syndrome, Acute Brain Syndrome, Psychosis, Personality Disorder, Chronic Alcoholism, and Other showed that test performance was not related to race or sex, but was significantly aligned to age, IQ, and education. The best single measure was the Bender-Gestalt Test, which was scored by the Hutt-Briskin (1960) method, which correctly identified 82% of all patients. This system rates subjects on the number of errors made in performance of the Bender-Gestalt Test, which includes rotation, overlap difficulty, closure difficulty, cohesion, perseveration, retrogression, angulation difficulty, fragmentation, collision, simplification, impotence, and
motor incoordination. The Hutt-Briskin method had been chosen for its simplicity, because it requires 3 minutes per protocol.

Russell (1976) described the Bender-Gestalt Test as a fairly accurate, gross, and quick test for diffuse, slowly progressive types of brain damage, such as Alzheimer's disease. The question posed by Russell was: if figural tests for brain damage, such as the Bender-Gestalt Test, have been related more or less directly to one area of the brain, then why have we obtained the accuracy in testing that we seemingly have? He had a three-part explanation which credited the Bender in testing an area of the brain not usually covered in standard neurological and psychiatric examination. He pointed to the fact that a standard drawing test was necessary to examine right-hemisphere functioning. Secondly, Russell further states that the Bender served as a "don't hold" test, i.e., test performance on the Bender would be more readily impaired by brain damage of any kind and in any part of the brain. Because "don't hold" tests can be affected by diffuse conditions, the Bender-Gestalt Test can be affected by slowly progressive diseases. Lastly, the Bender-Gestalt Test, picking up on slowly progressive types of diseases, allows for an earlier diagnosis than that of other batteries such as the Halstead-Reitan, which is more sensitive to mental deterioration than gross neurological examination.
Embree (1967) attempted to distinguish a brain-damaged population from a chronic schizophrenic population by utilizing the Bender-Gestalt Test. The Embree/Butler method of scoring assigned weights to nine signs (Hain, 1964; Hutt & Briskin, 1960; Pascal & Suttell, 1951) according to the extent to which they discriminated between the organic and psychiatric patients. It was determined by means of chi square analysis that a cutoff score of 14 and above was indicative of the presence of brain damage. An additional comparison was made with the scoring system in which different brain-damaged loci were compared with each other. Embree's results failed to show statistical difference in comparing various brain-damaged loci with each other--but the Embree/Butler scoring system did significantly differentiate the organic group from the psychiatric group.

The present study was undertaken to distinguish between organic brain syndrome (or cerebral dysfunction) and psychiatric classification through the utilization of cross-validation of the Embree/Butler scoring system. It was hypothesized that this system could distinguish organic brain syndrome and discriminate between brain dysfunction and psychiatric classification.

Method

Subjects

The subjects were 100 patients (45 males, 55 females) who had completed the Bender-Gestalt Test as part of a
standard diagnostic battery used at a psychiatric inpatient hospital in Dallas, Texas. The age range of the group was 13-77, with a mean of 34.07 years. Of these subjects, 50 had been diagnosed as having organic brain syndrome, according to hospital records, and the remaining 50 subjects were diagnosed as having functional psychiatric disorders. Hospital staff professionals had diagnosed 71 subjects as psychotic, and 21 were diagnosed as neurotic. Education of the subjects ranged from 4 to 17 years, with a mean of 10.3 years.

**Instruments**

The Bender-Gestalt Test is a visual construction test which simply involves the reproduction of nine geometric figures visually presented to the subject, who is asked to copy the designs with a pencil, one at a time, onto a standard 8½x11-inch piece of paper. The Embree/Butler system was used to score the Bender (see Appendix A). Subjects were classified as having brain dysfunction according to the criteria described by Embree, using a weighted scoring system. A cutoff score of above 14 indicated the presence of brain damage. The system of error included partial rotation (5 points), omission of angles (5 points), added angles (4 points), overlap difficulty (4 points), distortion (4 points), tremor (4 points), embellishments (3 points), lack of closure (3 points), and angles flattening (3 points). This scoring procedure was
selected over other methods because of its simplicity and brevity, and because of the lack of previously published research on the total effectiveness of the technique.

Procedure

Permission was granted by the hospital administration for the review of the subjects' records, which included personal data and the results of the Bender-Gestalt Test given upon admission to the hospital by one of five different clinicians under highly similar conditions.

Tabulation was made of the hospital diagnoses for all of the subjects. The same protocols were scored by the experimenter and two other psychologists, utilizing the Embree/Butler scoring system.

A scoring reliability check on the first 20 protocols with the three raters yielded .97 consistency. Additionally, each rater's exact score on each protocol was compared with those of the other two raters, producing 85% absolute agreement. The experimenter scored the remaining 80 protocols, using the Embree/Butler scoring method.

Results

Data obtained from the 100 protocols were subjected to chi square analysis (77.99 < .01 level), showing differentiation between the organic brain syndrome (or cerebral dysfunction) and psychiatric classification as determined by the Embree/Butler scoring system. Brain impairment was also
indicated by error scores when an organic brain syndrome was not necessarily the primary diagnosis, and, as such, was considered to be discriminatory in classification.

A point-biserial correlation was used to distinguish relationship between diagnosis and the score. A cutoff score of 15 produced the fewest false-negative or false-positive evaluations (see Figure 1, Appendix B). A false-negative indicated no brain dysfunction (using this particular scoring system) when previous diagnosis by hospital personnel had indicated brain dysfunction. A false-positive implied brain dysfunction (using the Embree/Butler system) when hospital personnel had diagnosed no brain dysfunction.

**Discussion**

The data strongly support the hypothesis that the Embree/Butler scoring system discriminates between classification of organic brain syndrome (or cerebral impairment) and psychiatric diagnosis. The Embree/Butler system is exceptionally reliable between test interpreters or raters, proving to be simple to score, with objective results. The scoring system does not appear to require the time and effort of an experienced clinician in administration and scoring.

The cutoff score of above 14 was established as an appropriate one and marked a departure of a half score from the Embree study (1967); that cutoff point was set at 14 and
above, and the present study denotes a cutoff point above 14. This diversion produced an exceptionally low minimum (one in this study) number of false-negatives and a very small number (five) of false-positive indicators of brain impairment when measured against the diagnosis of hospital clinicians. It is expected that no single instrument can serve as a complete and accurate screen for any type of pathology, including brain impairment. It is also possible, particularly in the case of the false-positives, that the Embree/Butler system added precision and accuracy to the hospital clinician's diagnostic criteria.

Age and sex do not appear to be influential variables differentiating error scores on the Bender-Gestalt Test, using the Embree/Butler scoring technique. In the present study, both groups (those with organic brain syndrome and those without) were almost equally divided between men and women, but with a wide age range.

The impairment consequent to brain damage is a joint function of several factors, including etiological background, physiological locus of damage, and the extent of damage. Hain (1964) and others have suggested that it is, at present, unrealistic to expect any single psychological test to isolate impairment associated with all types of brain damage. The Bender-Gestalt Test appears to focus on that impairment which is by nature diffuse and slowly progressive; such impairment characterizes arteriosclerotic and cerebrovascular
insufficiencies. Test scores do not identify particular types of brain damage.

Garron and Chiefetz (1967) suggest that the issues of locus and extent of intercranial pathology are critical in assessing the efficacy of the Bender-Gestalt Test. Cogency of the test is inevitably a function of its ability to demonstrate impairment in visual-motor behavior. It should be stated that the present study was unable to assess, through neurological examination, the extent to which the subjects exhibit parietal lobe dysfunction. Garron and Chiefetz further argue that an acceptable Bender-Gestalt protocol does not necessarily rule out organic brain pathology; however, it does raise questions as to possible parietal lobe involvement. Conversely, an inadequate protocol does not perforce limit the extent of the brain pathology to the parietal areas. Billingslea (1963) observed that the Bender-Gestalt Test is constructed on the premise that accurate visual-motor perceptual behavior is a skillful act involving sensory reception, central neural interpretation, and motor reproduction. Further implied is the notion that this perceptual process may be distorted by a neural injury, by variation in the cognitive integrity, and by dysfunctions in the emotional processes of the perceiving subject. Assuming the validity of this premise, the problem must lie in the quantitative evaluation of the degree of distortion in the figure reproduction.
Medical science has long noted configurations of signs and symptoms, termed syndromes, which may be taken as indices of pathology or injury.

What is actually being measured by the Bender-Gestalt Test, using the Embree/Butler scoring system or any other, remains an object of speculation from a neurological perspective. To perform well, the subject's perceptual motor processes must function with minimal deficiency or impairment. To score with the smallest number of errors, the subject must analyze the figure, synthesize a whole gestalt comprising component parts, and finally reproduce the figure by accomplishing a motoric response. Because this entire process is more complicated than the act of perceiving a spatial relationship, a poor score represents a level of brain functioning which surpasses the functional capacity of the right parietal lobe, an area of the brain generally associated with perceptual spatial reproductions.

Regardless of the dimension which is actually being quantified by the Bender-Gestalt Test, the Embree/Butler scoring system has demonstrated a high level of accuracy. As an example, consider Russell's (1976) single-case study of a 53-year-old white male who received a severe brain injury and was subject to surgery which involved the repair of fracture lines extending through the frontal parietal bones to the frontal bone. This subject was administered the Bender-Gestalt Test some years later, using the Hutt
scoring system. Russell presents the results of this subject's score in part as a demonstration of the point that the Bender-Gestalt Test may be well within the normal range while the subject may yet have severe brain damage. The subject evidenced a normal protocol following the Hutt scoring system. Remarkably, this same protocol, when scored by the Embree/Butler system, identified the subject as one who should be considered strongly for further investigation of brain damage.

In summary, the Embree/Butler scoring system of the Bender-Gestalt Test appears to produce a reliable and valid indication of brain impairment as differentiated from primary psychiatric classification. In the present study, this scoring system correctly identified all protocols recognized as dysfunctional by the hospital personnel, with the exception of a single case. In five cases, this system identified dysfunction which had not been ascertained by the hospital staff. Further, this scoring system is able to present an objective method, as evidenced by the 85% absolute agreement of the three raters in the present study. These results may be best applied for screening purposes in clinical settings in which the patient population is diagnostically heterogeneous, as in this research.
Appendix A

Embree/Butler's Bender Scoring Sheet

<table>
<thead>
<tr>
<th>Name</th>
<th>Date tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error Scores</th>
<th>1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 points</td>
<td></td>
</tr>
<tr>
<td>Partial rotation (A, 4, 5, 6, 7, 8)</td>
<td></td>
</tr>
<tr>
<td>Omission of angles (A, 4, 7, 8)</td>
<td></td>
</tr>
<tr>
<td>4 points</td>
<td></td>
</tr>
<tr>
<td>Added angles (A, 4, 7, 8)</td>
<td></td>
</tr>
<tr>
<td>Overlap difficulty (7)</td>
<td></td>
</tr>
<tr>
<td>Distortion (all)</td>
<td></td>
</tr>
<tr>
<td>Tremor (all)</td>
<td></td>
</tr>
<tr>
<td>3 points</td>
<td></td>
</tr>
<tr>
<td>Embellishments (A, 4, 6, 7, 8)</td>
<td></td>
</tr>
<tr>
<td>Lack of closure (A, 4, 7, 8)</td>
<td></td>
</tr>
<tr>
<td>Angles flattening (3)</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL __________________
Figure 1. Error scores on the Bender-Gestalt Test utilizing the Embree/Butler scoring system. False-negative indicates no brain dysfunction when the previous diagnosis had indicated brain dysfunction. A false-positive implies brain dysfunction when hospital personnel had diagnosed no brain dysfunction.
Appendix C

Scoring Instructions

Each of the weighted signs is only scored once per record. For example, a Partial Rotation might occur on Designs A, 4, and 7 on one record, but the maximum number of points scored for partial rotations would be five points. When a sign is noted on any of the designs indicated, the weight given on the Score Sheet is recorded in the blank space. The sum of all weighted scores recorded is the Total Score.

Scoring Criteria

1. Partial Rotation (Designs A, 4, 5, 6, 7, 8):

   When only one subpart of the design is rotated 21 degrees or more. Card or paper rotations of the whole, but correctly reproduced, design do not score.

2. Omission of Angles (Designs A, 4, 7, 8):

   An angle is omitted.

3. Added Angles (Designs A, 4, 7, 8):

   An extra angle is added. Lines forming the extra angle should be approximately straight and form a definite angle. Dog-ears do score. Curves or arcs caused by a change in direction of a line do not score. A jagged line which results from a gross tremor does not score.

4. Overlap Difficulty (Design 7):

   Difficulty in reproducing the overlapping angles on Design 7.

5. Distortion (All Designs):

   Reproduction exhibits a basic destruction of the gestalt of the design although all of the separate elements may be present.

6. Tremor (All Designs):

   Noticeable fine or gross waviness is evident in reproduction of lines.
7. Embellishments (Designs A, 4, 6, 7, 8):

An extra meaningless line is included in the design. Extra lines that are not integrated into a design are scored. Such lines are usually in an opposite direction from the lines to which they are near or attached. These are often small and lightly drawn and are easily overlooked in scoring.

8. Lack of Closure (Designs A, 4, 7, 8):

Any one design has two or more angles which are not closed.

9. Angle Flattening (Design 3):

Angled dots are flattened to the extent that no angle or apex of angle is evident.
References


Bender, L. A visual-motor gestalt test and its clinical use. *American Orthopsychiatric Association, research monograph*, 1938, No. 3.


Gobetz, W. A quantification, standardization, and validation of the Bender-Gestalt Test on normal and neurotic adults. *Psychology Monograph, 1953, 67(6, Whole No. 609).*


