Performance of Female Adults on the Southern California Visual Figure-Ground Perception Test

(figure-ground perception, perception, test reliability, tests)

Paul Petersen  Douglas Goar  Julia Van Deusen

This study was conducted to assess adult female performance on the Southern California Visual Figure-Ground Perception Test (FGP) and to obtain an estimate of the test’s reliability. Data were obtained for two scoring formats. The first format was the score for all items on the test (TOTAL), and the second format was the score through the fifth error (FIRST-FIVE). The mean scores were 38.24 (SD = 3.2) for TOTAL and 29.71 (SD = 10.14) for FIRST-FIVE. For each format, mean, standard deviation, kurtosis, and skewness values are presented (N = 124). Both formats were reliable over time (.90 and .71, respectively), and scores of the even-numbered items correlated moderately with the odd-numbered items, indicating split-half reliability (.70 and .87 for both administrations of the test). Although the sample was representative across the socioeconomic variable, moderate correlations were obtained between the performance scores and the age and education level variables. We concluded that the FGP is a reliable assessment tool for use with adults; however, a subject’s age and education must be considered when interpreting the scores.

Visual figure-ground deficit is a condition in which a person has trouble differentiating an object or figure from its competing background. An example is a patient with brain damage who has difficulty locating a necktie on a quilt, a lipstick tube in a drawer of cosmetics, or a ring in a jewelry case. Three measurement procedures purported to assess this deficit in adults have received attention in recent literature. They are the Witkin-Oltman Rod and Frame Test (RFT) (1), the Embedded Figures Test (EFT) (2), and the Southern California Figure-Ground Visual Perception Test (FGP) (3).

The RFT requires that the patient place a rod in a vertical position within a tilted frame. The RFT, with standardized directions, is available in its original version and also in a more portable version (5). However, the research conducted on the RFT indicates that it does not adequately measure visual figure-ground deficit. Allen and others (6) found negative correlations between the RFT and the EFT. In a series of three experiments, Streibel and Ebenholtz (7) demonstrated that the RFT and EFT do not measure a common perceptual construct.

The EFT comprises a series of timed visual items, each consisting of a complex figure in which a simple figure is embedded. The patient is asked to trace the simple figure after viewing both figures in a standardized administrative sequence (4). This measurement procedure has been generally accepted as the standard procedure by the psychometric professional as the best means currently available to assess figure-ground deficit.

The FGP is a series of timed visual items, each consisting of a complex figure within which six simple figures are embedded. The patient is asked to identify a multiple set of three of the simple figures within the complex figure after viewing the complex and multiple figure plates in a standardized

Paul Petersen, PhD, OTR, is Assistant Professor, Occupational Therapy Department, Elizabethtown College, Elizabethtown, PA 17022.
Douglas Goar, PhD, is Assistant Professor, Physical Therapy Department, University of Wisconsin-Madison, Madison, WI 53706.
Julia Van Deusen, PhD, OTR, is Associate Professor, Occupational Therapy Department, University of Florida, Gainesville, FL 32610.
adults. Research using other in-
shown inconsistencies in gender
age group but not by gender. Kap-
lan and Hier (9) recently provided construct val-
ification for the FGP by finding sig-
ficant correlations with several
other relevant measures for 34
stroke patients. Petersen and Wi-
koff (10) demonstrated that a sig-
ificant correlation ($r = .67$) exists
between the results of the EFT and
the FGP. In this paper it was also
pointed out that the motor and
memory skills needed to complete
the EFT are confounded with per-
ceptual skill. The rationale for our
study with the FGP is to identify
an alternative assessment to the
EFT and thus lessen the impact of
potentially confounding variables
in the study of figure-ground skill,
especially in the clinical sample.

Review of the literature pro-
duces essentially no normative data
on the FGP as a test for female
adults. Research using other in-
struments purporting to measure
figure-ground perception has shown inconsistencies in gender
differentiation (6, 7, 9, 11, 12).
Sieve and Freishtat (13) cited lim-
ited normative data for adults by
age group but not by gender. Kap-
lan and Hier (9) reported figure-
ground mean scores separately for
males (10.5) and females (13.5) but
indicated that there were no signif-
ificant gender differences. Sivak and
others (5) reported FGP mean
scores also for disabled and able-
bodied subjects, but subject num-
bers were too small to have any
utility for test standardization.

Because it may not be appropri-
ate to apply the FGP normative
data for males to females, a study
is needed to generate normative
data for the healthy adult female
population. Therefore, the pur-
pose of this study is to establish a
reliable set of normative data in
using the FGP for that female
group.

**Methods**

**Subjects**

The FGP was administered to
124 females from a north-central
state by the first author and several
specially trained research assis-
tants. Thirty subjects were retested
through a random sampling of the
subject pool.

The subjects represented a
range of demographic variables.
The age variable ranged from 18
through 81 years, with a mean of
34.85 years ($SD = 13.46$). To ob-
tain the data for the education level
variable, the highest level of edu-
cation attained was recorded in
years and tenths of years. For ex-
ample, beyond high school, each
college course is equivalent to one-
tenth of a year, unless college was
recorded in whole years (i.e., 13
through 16). The mean education
level of the subjects was 14.66
years ($SD = 2.9$). Socioeconomic
status (SES) was quantified using
the Two-Factor Index Social Posi-
tion (14). This index converts a
subject’s educational and occupa-
tional level to a raw score and
ranges from 11 (high SES) to 77
(low SES). Five SES groupings are
calculated from this raw data. The
mean SES group in this study was
3.02 ($SD = .82$), which repre-
sents a middle range. SES was not cal-
culated for full-time students un-
less they were also working full-
time ($N = 24$). For those subjects
not currently employed, the occu-
pational level of the spouse was
used to calculate the SES raw
scores. Fifty-two subjects were
married at the time of the testing.

No data are available on the status
of those nonmarried subjects (i.e.,
whether they were widowed, di-
vorced, or single). None of the sub-
jects reported any difficulty seeing
the test plates either with or with-
out correction. The distribution of
the subjects’ educational levels and
raw and group SES scores were
within normal limits, based on their
kurtosis and skewness values.

**Instrument**

The FGP contains 18 plates,
each one having a complex picture
on the top and six less complex
pictures below. Following practice
items, the subject has one minute
to identify three of these less com-
plex pictures, which are embed-
ded within the complex picture above.
Subjects earn one point for each
correctly identified picture. Test-
ing is stopped when the fifth error
is made. Any picture not identified
within the time limit, or any picture
improperly designated as part of
the complex figure, is considered
an error. Should the subject’s first
choice for a given plate result in
the fifth error on the test, the next
two choices on that plate, even if
correct, are not included in the test
score. The maximum test score ob-
tainable is 48. Reported reliabili-
ties by the test’s author (3, 15)
range from .36 to .54, with a mean
correlation of .44. Reliabilities
were tested on a sample of over
300 children.

**Procedure**

For an estimate of test-retest re-
liability, 125 adult females were
tested with the full FGP test and
30 were retested after an intertest
interval of four to six weeks. After
analysis, data from one subject
were eliminated from the study.

The resulting number of subjects

526 August 1985, Volume 39, Number 8
tested was 124, including 29 retests. In this study, the scoring procedure of the method discussed above was modified. Based on the recommendation of an earlier paper (10), scoring is not stopped immediately after the fifth error but is stopped after the completion of the plate in which the fifth error was made. This procedure was instituted to increase test reliability.

It was found that adult subjects do not call out answers as they identify the correct pictures as do pediatric subjects. Usually, adults scan the plate and identify their responses all at one time and then call out their three responses in numerical order. It is then impossible to accurately determine where the fifth error was made.

Using this revised procedure, subjects are not penalized for the order in which they called out their response choices. All subjects were administered the full test of 18 plates (including the practice plates). Several scores were examined in this study. TOTAL-1 is the full test of 16 plates taken by all subjects. TOTAL-2 is the full test taken by the 29 subjects who were retested. FIRST-FIVE-1 is the test score through the fifth error obtained by all 124 subjects. FIRST-FIVE-2 is the result of the second testing of 29 subjects. EVEN-1 is the sum of correct items from the even-numbered items. EVEN-2 is the retest (N = 29). ODD-1 is the sum of odd-numbered items taken by all subjects. ODD-2 is the retest of odd-numbered items (N = 29). Distributions, statistics, and reliabilities can be calculated from these data.

Results

Distribution

Table 1 contains the mean, median, standard deviation, skewness, kurtosis, and range for all FIRST-FIVE, TOTAL, EVEN, and ODD scores. In all cases of test-retest data, the later administration resulted in a higher score; this perhaps indicates a memory or practice component during the intertest interval. A perfect score of 48 was obtained by one subject on the initial test and by two subjects on the retest. It did not appear that the scores of this test are subject to a ceiling effect. Based on kurtosis and skewness scores less than +1.00 or -1.00, all TOTAL, EVEN, and ODD scores have distributions within normal limits.

Demographic Variables

Pearson correlation coefficients were calculated to assess the relationships of the demographic variables for the subjects and their FGP test scores (see Table 2). Data indicate significant relationships between educational level socioeconomic status, and age variables.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>29.71</td>
<td>36.41</td>
<td>38.29</td>
<td>41.76</td>
<td>18.24</td>
<td>20.17</td>
<td>20.00</td>
<td>21.55</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>28.00</td>
<td>43.00</td>
<td>39.0</td>
<td>43.0</td>
<td>19.0</td>
<td>21.00</td>
<td>20.00</td>
<td>22.00</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td>10.14</td>
<td>11.39</td>
<td>5.2</td>
<td>5.43</td>
<td>3.38</td>
<td>3.75</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-1.19</td>
<td>-1.18</td>
<td>-1.95</td>
<td>-2.28</td>
<td>-8.93</td>
<td>-9.99</td>
<td>-1.9</td>
<td>-6.44</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>-1.19</td>
<td>-1.18</td>
<td>-1.95</td>
<td>-2.28</td>
<td>-8.93</td>
<td>-9.99</td>
<td>-1.9</td>
<td>-6.44</td>
</tr>
</tbody>
</table>
Table 2
Pearson Correlations, Significance Levels, and Subject Numbers of Demographic Variables and FGP Scores

<table>
<thead>
<tr>
<th></th>
<th>TOTAL-1</th>
<th>TOTAL-2</th>
<th>FIRST-FIVE-1</th>
<th>FIRST-FIVE-2</th>
<th>EVEN-1</th>
<th>EVEN-2</th>
<th>ODD-1</th>
<th>ODD-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.41</td>
<td>-.23</td>
<td>.034</td>
<td>-.18</td>
<td>-.39</td>
<td>-.23</td>
<td>-.35</td>
<td>-.25</td>
</tr>
<tr>
<td>p</td>
<td>.0001</td>
<td>.23</td>
<td>.0001</td>
<td>.36</td>
<td>.0001</td>
<td>.23</td>
<td>.0001</td>
<td>.20</td>
</tr>
<tr>
<td>N</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
</tr>
<tr>
<td>Education level</td>
<td>+.41</td>
<td>+.45</td>
<td>+.36</td>
<td>+.38</td>
<td>+.42</td>
<td>+.43</td>
<td>+.30</td>
<td>+.46</td>
</tr>
<tr>
<td>p</td>
<td>.0001</td>
<td>.0149</td>
<td>.0001</td>
<td>.0428</td>
<td>.0001</td>
<td>.0185</td>
<td>.0006</td>
<td>.0118</td>
</tr>
<tr>
<td>N</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
</tr>
<tr>
<td>SES (raw)</td>
<td>-.43</td>
<td>-.71</td>
<td>-.39</td>
<td>-.55</td>
<td>-.42</td>
<td>-.70</td>
<td>-.33</td>
<td>-.68</td>
</tr>
<tr>
<td>p</td>
<td>.0001</td>
<td>.0149</td>
<td>.0001</td>
<td>.0175</td>
<td>.0001</td>
<td>.0011</td>
<td>.0008</td>
<td>.0020</td>
</tr>
<tr>
<td>N</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
</tr>
<tr>
<td>SES (group)</td>
<td>-.43</td>
<td>-.64</td>
<td>-.43</td>
<td>-.50</td>
<td>-.43</td>
<td>-.60</td>
<td>-.36</td>
<td>-.64</td>
</tr>
<tr>
<td>p</td>
<td>.0001</td>
<td>.0043</td>
<td>.0001</td>
<td>.0342</td>
<td>.0001</td>
<td>.0083</td>
<td>.0003</td>
<td>.0039</td>
</tr>
<tr>
<td>N</td>
<td>99*</td>
<td>18†</td>
<td>99</td>
<td>18</td>
<td>99</td>
<td>18</td>
<td>99†</td>
<td>18†</td>
</tr>
</tbody>
</table>

SES, socioeconomic status. *An N of 99 was obtained because 25 subjects were full-time students and SES was not calculated for them. † SES data were calculated for 18 of the 29 subjects retested.

Table 3
Intercorrelations of FGP Scores

<table>
<thead>
<tr>
<th></th>
<th>TOTAL-1</th>
<th>TOTAL-2</th>
<th>FIRST-FIVE-1</th>
<th>FIRST-FIVE-2</th>
<th>EVEN-1</th>
<th>EVEN-2</th>
<th>ODD-1</th>
<th>ODD-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL-1</td>
<td>.90</td>
<td>.87</td>
<td>.88</td>
<td>.95</td>
<td>.90</td>
<td>.87</td>
<td>.82</td>
<td>.82</td>
</tr>
<tr>
<td>r</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>p</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>TOTAL-2</td>
<td>.72</td>
<td>.95</td>
<td>.90</td>
<td>.98</td>
<td>.98</td>
<td>.64</td>
<td>.94</td>
<td>.94</td>
</tr>
<tr>
<td>r</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0002</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>p</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>FIRST-FIVE-1</td>
<td>.71</td>
<td>.80</td>
<td>.73</td>
<td>.80</td>
<td>.80</td>
<td>.63</td>
<td>.86</td>
<td>.86</td>
</tr>
<tr>
<td>r</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>p</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>EVEN-1</td>
<td>.88</td>
<td>.70</td>
<td>.86</td>
<td>.86</td>
<td>.86</td>
<td>.86</td>
<td>.86</td>
<td>.86</td>
</tr>
<tr>
<td>r</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>p</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>124</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>EVEN-2</td>
<td>.57</td>
<td>.87</td>
<td>.53</td>
<td>.87</td>
<td>.53</td>
<td>.53</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>r</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>p</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 3 Intercorrelations of FGP Scores

be reliable over time. Retesting EVEN items results in a coefficient of .88. The significance is reported in Table 3.

Learning. Witkin and others (4) warned that researchers should be aware of a practice effect as seen in improved scores on the retest. It is not known at this time whether this improvement occurs because the subject remembers specific test figures, has had practice, or both. Data in Table 1 show that in every case scores improved on the retest. This cannot be seen in the correlations because these procedures are yoked to the relative position of subjects to each other's FGP scores. Although the relative positions of subjects are stable, the correlations cannot show that each subject improved her score by several points. A t test (one tailed) revealed that scores on TOTAL-2 were significantly greater than scores on TOTAL-1 (t = 7.0813, df = 28, p < .0005). Thus, the second score of the 29 subjects who completed the FGP twice was significantly higher. However, the
learning that may have occurred during the intertest interval might not apply to brain-injured adults with memory impairment.

Levels of Difficulty. The order of items in the FGP is one of increasing difficulty, beginning with common items and continuing with geometric shapes. Twenty subjects scored higher on the geometric shapes than on the common figures. Perhaps this indicates that the plates containing geometric shapes are not necessarily more difficult than common items but may require a different type of skill than the identification of the common shapes. Seventeen subjects obtained the same score on both halves of the test. This perhaps indicates that a qualitative, rather than a quantitative, difference exists between the groups. The mean score of the eight plates of common objects (items 1–8) was 20.65, whereas the mean for the geometric shapes (items 9–16) was 17.65. The distribution of each set of scores was within normal limits. A t test (one tailed) demonstrated that scores on the common objects were higher than those on the geometric items ($t = 9.15017$, df = 123, $p < .0005$). One explanation of the fact that many subjects obtained higher scores on the geometric plates is that the ability required to disembed geometric shapes is different from that needed to disembed the common objects.

Another item that could affect the difficulty level of the test, and especially the reliability of the shorter administration (FIRST-FIVE), is contained in plate no. 7. Of the 124 subjects tested, only 23 (18.5%) identified plate no. 2 (the second of six response choices) as being contained in the more complex picture. On retest of 29 subjects, only 9 out of 29 were correct. Four subjects were correct both times. Although a complete item analysis was not done, any given item missed by more than half of the subjects certainly does not contribute to the reliability of the test (16) and might account for the lower test-retest reliabilities of the FIRST-FIVE scoring version of the FGP that Ayres (3, 15) obtained.

Discussion/Application

Means for the FGP are presented in Table 1. By applying the standard deviations to these mean scores, the therapist obtains an estimate of a female adult’s current level of functioning, with respect to the mean. Caution is advised in score interpretation because age, sex, and education levels all correlated significantly with the mean scores. There were no clear age increments for the relationship between age and FGP. However, it can be said that in general, performance scores in disembedding type tasks tend to be stable in individuals between the ages of 15 and 45 and that this is followed by a gradual decline in performance as the individual ages.

Because this test has normal distributions for the TOTAL, EVEN, and ODD scores, it is useful in research. The test is reliable over time as evidenced by significant correlations between the two administrations of the FGP. No ceiling effect was identified. Because of the strong relationship between the ODD- and EVEN-numbered items, it is possible to divide the test into two equivalent forms. This provides a tool that may be reliably used before and after a given procedure or treatment. The clinical uses of this test for occupational therapists is discussed in another paper (10). Briefly, these applications include the areas of psychopathology, sensory integration, and physical disabilities.

There are sex differences on figure-ground type performance tests. Typically, males perform significantly higher than females (4). Male norms have been reported (10), and the scores of that male sample and our female sample reported are almost identical. However, caution is indicated when comparing the scores of the sexes, because the male and female FGP data were from different studies. Nonetheless, because of the large representative samples in each study and the use of similar procedures, it is interesting to compare these scores and speculate on the similarity between the sexes in the two FGP normative studies versus the male superiority with the use of other tests (e.g., the Embedded Figures test). The TOTAL mean for males was 38.53, with a SD of 4.84 (10); and the TOTAL mean for females (as found in our study) was 38.29, with a SD of 3.2. The FIRST-FIVE mean for males was 29.72, with a SD of 8.83 (10); and the FIRST-FIVE mean for females was 29.71, with a SD of 10.14.

When comparing the FGP test with other tests (see Reference 10 for a comparison), it is obvious that the FGP test, unlike other tests, requires no motor component. It is perhaps the confounding of the motor skill with the visual perception that has previously accounted for the sex differences. Perhaps visual perception skill needed to disembed figures is similar in males and females. However, when a motor response is needed in a given figure-ground test, differences between the sexes surface.

Conclusion

Normative distribution data, re-
liability coefficients, and other statistics for adult females on the FGP have been presented, providing the therapist with the data needed for comparison. Beyond additional evidence for validity for the FGP as a whole, there is need for an item analysis to assess difficulty levels and internal consistency for each item. The topic of sex differences inferred from this research requires thorough study as does the utility of the FGP with the brain-injured, perhaps looking at the differences between this sample and a sample of non-injured adults. Because our studies indicate the FGP's potential as an evaluative tool, the immediate need is for rigorous study to determine if this instrument is a valid measure for the brain-damaged adult and/or other patients with potential for perceptual dysfunction.

ACKNOWLEDGMENTS
The authors thank David Comello, OTR, Lisa Hansen, OTR, Cindy Nemetz, OTR, Beth Dexter, OTR, and Cindy Johnson, OTR, for their assistance in data collection and the preparation of the data analysis. This research was supported in part by a grant from the Office of the Dean, School of Allied Health Professions, University of Wisconsin-Madison, Madison, WI.

REFERENCES
3. Ayres AJ: Southern California Figure-Ground Visual Perception Test. Los Angeles: Western Psychological Services, 1966