Scandinavian Journal of Psychology, 2006, 47, 1–8

Cognition and Neurosciences

Neuropsychological assessment with the Visual Gestalt Test: Psychometric properties and differential diagnostic probabilities

PETER LA COUR[1] and RUTH ANDERSEN[2]

[1]*Department of Health Psychology, Institute of Public Health, University of Copenhagen, Denmark*
[2]*Psychiatric Clinic and Department of Psychology, Rigshopitalet, Copenhagen, Denmark*

la Cour, P. & Andersen, R. (2006). Neuropsychological assessment with the Visual Gestalt Test: Psychometric properties and differential diagnostic probabilities. *Scandinavian Journal of Psychology*, *47*, 1–8.

The Visual Gestalt Test is a neuropsychological instrument developed for evaluation of learning and memory of visuo-spatial material. A revised strategy of scoring has motivated the present study, where data from 153 normal persons, 99 epilepsy patients, and 24 depressed patients are presented and compared. The Visual Gestalt Test is observed to discriminate between normal and diagnosed groups in several ways. Additionally it is found to discriminate between depressed and brain damage subgroups of patients. Data are presented in order to supplement previously published ways of scoring and norms. Practical guidelines for the clinical applications of the test are suggested as perspectives.

Key words: Neuropsychological tests, brain damage, depression, visual perception, test.

Peter la Cour, Department of Health Psychology, Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, P.O.B. 2099, 1014 Copenhagen K., Denmark. Tel: +45 3532 7929; fax: +45 3532 7748; e-mail: p.lacour@pubhealth.ku.dk

INTRODUCTION

The visuo-spatial memory test "Visual Gestalts" (Andersen, 1968) appeared in the late 1960s. The test was designed due to the growing interest in the various side effects of unilateral and bilateral electroconvulsive treatment (ECT; Cronholm & Molander, 1957; Cronholm & Ottosen, 1961) and the observations of the different patterns of amnesia caused by temporal lobe dysfunction (Penfield & Milner, 1958; Milner, 1968). The primary incentive for the construction of the test was the preparation of methods to be used in a research project on unilateral vs. bilateral ECT (Hesche, Röder & Theilgaard, 1978). The purpose was to develop a test to supplement the totally dominating verbal memory tests that represented memory function at the time (Willanger, 1970), and furthermore to provide a more exact procedure in non-verbal tests for assessing different variables in learning and memory processes, e.g. immediate recall, learning, and delayed production.

There was a special focus on delayed memory, as a graded measure for evaluation of this function was missing in the visuo-spatial tests available at that time. Regarding delayed reproduction, the commonly used tests most often had "all or nothing" measures (Wechsler, 1945; Benton, 1963; Rey, 1941; Österreith, 1946).

Description of the test

The Visual Gestalt Test consists of four complex designs, each circumscribed by a well-known geometric figure: circle, square, triangle, and semi-circle (example in Fig. 1). A model card with the first complex design (in a circle) is presented to the person for 10 seconds, then turned away, and the person is asked to draw the full figure on a response sheet with the circumscribed figure (an empty circle) outlined.

If the figure is filled in correctly, the next design (a square) is presented, but if the drawing is incorrect or incomplete, a new 10-second presentation of the figure is followed by a new response sheet (with the preprinted circle). To prevent over-learning, the tester has already filled in what the person has drawn correctly in the previous sheet. The tested person has only to complete or correct missing or faulty parts and will not have to repeat drawing the parts of the figure that have already been reproduced correctly. This procedure is repeated until the whole figure has been reproduced correctly. The test is continued with the same procedure for all four designs (circle, square, triangle, semi-circle). The number of erratic sub-patterns (subgestalts) is counted as a testscore (error count).

When the learning task has been completed, the person is told that he will later be asked to redraw as much of the figures as he can remember. After one hour, the person – without having seen the model cards in the meantime – is again given a sheet from the pad with the empty circle outline, and is asked to fill in as much of the patterns as he remembers. If the empty circle is filled in correctly, the empty square sheet is given, and the reproduction phase proceeds in the same order as the learning process.

If the person does not succeed, that is, shows either total or partial failure, the person is given an outline on which the

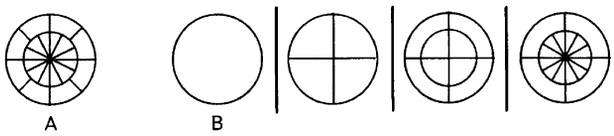

Fig. 1. Example from the Visual Gestalt Test; total figure and subgestalts.

tester has redrawn only the parts (subgestalts) already reproduced correctly, plus one more. Thus, the person progresses with increasing quantities of the figure given as prompts by the tester. The procedure is repeated until the figure is completed.

The test is constructed both to prevent over-learning and to give a graded measure of delayed reproduction.

The *traditional* scoring procedure, given by the original publication of the test (Andersen, 1968) counted the number of errors on each trial sheet used according to the score key of perceptual units, named subgestalts, as illustrated in Fig. 1.

The scores were summarized to a total error score for the learning and the reproduction phase separately.

Until now the traditional way of scoring has been used in many investigations (for instance Andersen, 1976, 1978; Bruhn, 1987; Drejer, Theilgaard, Teasdale, Schulsinger & Goodwin, 1985; Gade, Mortensen & Bruhn, 1988; Hansen, Andersen, Theilgaard & Lunn, 1982; Hesche *et al.*, 1978; Mikkelsen, Jorgensen, Browne & Gyldensted, 1988; Rasmussen, Jeppesen & Sabroe, 1993; Rosenberg & Andersen, 1990; Sørensen, Hansen, Andersen, Hogenhaven, Allerup & Bolwig, 1989; Teasdale, Hansen & Gade, 1997; and Theilgaard, 1984).

A *revised* scoring system was proposed (Andersen, 1989) to supplement the traditional counting of faulty subgestalts. First, separate counts of *error sheets* used in the learning phase and reproduction was suggested. Second, each drawn figure was suggested to be evaluated as a qualitative whole, and a scoring system was generated to describe and classify the *types of error*, that is, incompleteness of the figure, organizational level, distortion of the figure, fragmentation, rotation, perseveration and fusion of several figures. Each error sheet was given a score for the total qualitative impression. The qualitative scoring system was elaborated empirically from the data set also used in the present study and by years of practical experience with evaluation of test results (see Andersen, 1989 for details). The *types of error* categories were based on descriptions of perceptual patterns in normal and brain damaged persons, and on developmental issues in children's perception (Piotrowski, 1937; Bender, 1946; Hamby, Wilkins & Barry, 1993).

Furthermore, a system of categorization of the *pencil-line quality* of the tested person was presented (see below). The categorization was based on clinical observations of differences in patient behavior during testing.

The *aim of the present study* is to investigate the differential diagnostic possibilities of the Visual Gestalt Test by:

(1) Utilizing the supplementary quantitative scores.
(2) Analyzing the qualitative aspects of the figure drawing and quantifying the *different types of* errors.
(3) Comparing the results from the different diagnostic groups according to the revised scoring system.
(4) Presenting detailed data for clinical work, future research and suggestions for new clinical guidelines.

METHOD

Samples

The total sample in this study is $N = 279$. The sample consists of a number of subsamples from previously published studies on normal persons and brain damaged patients and from a new subsample of depressive patients. The tests of the samples were all administered by trained psychologists in the years of 1974–98. Information on gender, age, years in school, and hand preference was also recorded for each subject. All accessible material (score sheets) were re-examined for suitability for further analysis and rescored by the revised guidelines by two persons separately (second author and different assistant colleagues).

Subjects older than 59 years were excluded.

The normal control group

The normal control group ($N = 153$) consists of 138 subjects from a study on verbal and visual memory. Further information on the sample and procedures may be found in Andersen (1976). The original study comprised 165 subjects, but score sheets for 27 tests were not suitable for a revised scoring. The supplementary 15 normal control subjects originate from a study on personality characteristics and epilepsy (Sørensen *et al.*, 1989).

All potential subjects were questioned about head trauma, neurological diseases, and any subject suspected of suffering from a brain disorder was excluded.

The normal control group consisted of 82 females and 71 males. Age range was 15–59, age mean 33.6 (*SD* 13.8). Although data from persons older than 60 years would have been useful, the age limits from the earliest publication have been maintained.

Brain damage groups

The total group ($N = 99$) consisted of patients suffering from epilepsy; 39 of these were non-resected temporal lobe epilepsy patients with duration of illness of at least 15 years. This subsample was divided after careful examination of clinical data and EEG measures (Sørensen *et al.*, 1989). Twelve subjects had predominantly right temporal lobe epilepsy (*right-sided epilepsy*) (females: 9; males: 3, age range 26–52, mean 37.0, *SD* 8.2). Twelve subjects had predominantly *left-sided epilepsy* (females: 7; males: 5, age range 38–54, mean 43.1, *SD* 5.4). Fifteen subjects had *primarily generalized epilepsy* (females: 8; males: 7, age range 24–50, mean 35.5, *SD* 6.7). Four subjects with bilateral EEG-findings were excluded in the present study due to small *N*. All the epilepsy patients were tested under their usual anti-epileptic medication. Detailed information may be found in Sørensen *et al.* (1989).

Sixty subjects had undergone unilateral temporal lobe resection. This subsample consisted of 34 right-side resected and 26 left-side resected and originated from a retrospective study of resected patients (Jensen, 1977). Right-side resected (females: 15; males: 19, age range 16–53, mean 31.8, *SD* 10.9). Left-side resected (females: 10; males:

16, age range 17–57, mean 31.7, *SD* 9.4). The follow-up investigations were performed after a minimum postoperative period of 14 months.

The depression group

The depression subsample ($N = 24$) consists of 14 subjects diagnosed as suffering from major depression according to DSM III-criteria (Rosenberg & Andersen, 1990). The Visual Gestalt Test was administered before onset of anti-depressant medication. The remaining 10 subjects either had a well-known diagnosis of depression and were tested for the purpose of this depression sample, or they were referred to psychological testing in order to have their diagnosis clarified. Only patients in whom the diagnosis was confirmed by the later course of the illness were included in the study. The depression diagnosis was according to either DSM III, IV, or ICD10 (due to the time span of the period of data collection). All other subjects were excluded.

Three subjects were without medication, and 7 were medicated with usual anti-depressants at the time of assessment. Subjects treated with ECT or suspected of brain damage were excluded. In the depression group there were 13 females and 11 males, age range 23–58, mean 44.6 (*SD* 9.7).

Scoring procedures

The scoring followed guidelines in the revised scoring manual (Andersen, 1989). In addition to the scoring of errors (i.e. faulty subgestalts on answer sheets) in immediate recall and total errors in the learning and reproduction phases, error sheets were counted separately in the two phases, and difference in types of error were evaluated. The differentiation in error types included: omission of whole figure, omission of subgestalt, reversal or rotation, fragmentation, unspecified lower level processing, disorganization, fusion of figures, perseveration of single elements, unspecific stereotypy, disharmonic stereotypy, and unilateral spatial neglect. A few more categories were suggested in Andersen (1989), but omitted in this study, because they were not used in practice.

A registration of the pencil-line quality was also assigned, dividing pencil-line quality into five categories according to the revised manual:

Category 0: Unremarkable, reasonably precise, firm lines, which may contain small inaccuracies arising from the lack of ruler and compass (the drawings are to be made freehand).
Category 1: Exceptionally well-controlled, effortless, precise and competent drawing.
Category 2: Cautious, sketchy, and possibly weak pencil lines.
Category 3: An impression of an insufficient control of the fine motor coordination predominates. The person appears motivated to do his best, but unable to do so. The pencil lines are clumsy, rigid, awkward, squeezed, and possibly with varying degrees of pencil pressure. The lines may be shaky or endings inexact. Performance may vary from sheet to sheet during the reproduction of the figures.
Category 4: Impulsive, superficial or careless performance. The person appears motorically capable of achieving a good drawing, but does not make the effort.

RESULTS

The normal control group

For the normal group there were no significant gender differences with regard to age, education (years in school < 10 yrs >), or any types of errors.

The number of errors increased with age, especially in the reproduction sequence, where nearly all types of error counts were associated with age in some way. Age specific values of the error counts are shown in Table 1. Since length of education and age were found significantly related, interactions between education and number of errors were analyzed with age as a covariant in a multivariate model. Three error counts showed significance, all in the learning phase: immediate recall, total errors, number of learning sheets. In the reproduction phase, differences were still found, but they were not significant. The count of errors within education groups are listed in Table 2.

Differences between diagnosed and normal groups

Analysed with the *t*-test, there was no significant difference in education level between any of the diagnosed groups and the normal control group, but age differed significantly between the normal group and two other groups: the left temporal lobe epilepsy group (mean 43.0 *SD* 5.3) and the depressed group (mean 44.3 *SD* 9.7).

Table 3 presents the error count comparisons between the normal and the diagnosed groups. Significances are analyzed by one-way ANOVA with LSD (least significant difference) post hoc correction for the multiple comparisons. As may be seen, all error counts are higher in the diagnosed groups

Table 1. *Normal control group, main error counts and age – means and (SD)*

	Age group		
	15–39 $N = 95$	40–49 $N = 30$	50–59 $N = 28$
Learning, immediate recall	1.29 (1.47)*	2.23 (1.96)	2.64 (2.54)
Learning, total errors	1.41 (1.67)*	2.63 (2.72)	2.89 (2.91)
Learning, number of error sheets	0.97 (1.07)*	2.03 (1.79)	1.79 (1.60)
Reproduction, immediate recall	1.37 (2.22)*	3.73 (3.19)	4.50 (3.97)
Reproduction, total errors	1.78 (3.30)*	5.97 (6.29)	8.07 (7.66)
Reproduction, number of error sheets	0.95 (1.56)*	2.97 (2.59)	3.75 (3.28)

* Age group 15–39 differs significantly from both other age groups on all variables ($p < 0.5$). Differences between 40–49 and 50–59 are all not significant.

Table 2. *Normal control group, main error counts and years in school – means and (SD)*

	Less than 10 years $N = 64$	More than 10 years $N = 89$
Learning, immediate recall	2.36 (2.18)**	1.27 (1.49)
Learning, total errors	2.69 (2.70)**	1.37 (1.68)
Learning, number of error sheets	1.81 (1.62)**	0.98 (1.12)
Reproduction, immediate retention	2.91 (3.46)	2.04 (2.78)
Reproduction, total errors	4.64 (6.30)	3.11 (5.00)
Reproduction, number of error sheets	2.27 (2.77)	1.56 (2.21)

** Difference significant $p < 0.01$.

than in the normal group. With the LSD correction for multiple comparisons applied, the most robust count of difference between all the diagnosed groups and the normal group is the total count of error sheets, where the differences all meet statistical significance below the $p = 0.05$ level.

The total number of error sheets from both the learning and the reproduction sequence were (mean (*SD*)) 3.18 (3.30) for the normal and 9.03 (5.94) for the joint diagnosed groups, $p < 0.001$.

Table 4 shows the types of errors made by normal and diagnosed groups during the whole test. The detailed table is presented for clinical use. Significances are corrected by LSD post hoc procedure. The differentiations of error types showed patterns of differences between diagnosed and normal groups. As the only group the group of right-resected epilepsy showed highly significant differences on the error types of *fragmentation*, *disorganization*, and *transmission of error*. The groups of depressed and left-resected epilepsy shared highly significant differences from the normal group concerning the error types of *omission* of both *whole figure* and *subgestalts*.

Regarding *pencil-line quality* the most important differences are found between the categories 0 and 3, as may be seen in cross-tabulation in Table 5. When these two categories were computed separately, all separate diagnosed groups differed significantly from the normal group ($p < 0.001$ in Chi-square tests). The relative probability of having a diagnosis if pencil-line quality was in category 3 was between 5.9 and 11.5.

Differences between brain damaged and depressed

Age differences between the two groups were significant, therefore all analyses of variables were analyzed with age as a covariant in a multivariate general linear model with LSD-adjustment for multiple comparisons ($N = 123$).

The comparisons of the groups can be seen in Table 6. Significant differences were found for the error type of *fusion of figures*, where the brain damage groups scored higher ($p < 0.000$). The *omission of whole figure* also showed difference between the groups, tending to be significant ($p = 0.06$), but with the depressed group making the most errors.

On the error summary variables, age corrected significant differences were found on *reproduction*, *total errors* ($p = 0.04$), and a high significance level was found on the *learning sheet*, *learning minus reproduction variable* ($p = 0.003$). The value of this variable was negative in the brain damage group, indicating that the number of error sheets is higher in the reproduction sequence than in the learning sequence. With estimated effect size (age correction to the statistical mean age (age = 36.3), the mean was –1.58 (*SD* 0.39).

For the depressed group the value of this variable was positive, indicating more sheets used in the learning sequence than in the reproduction sequence. The estimated age corrected mean for the depressed group increased this finding (1.33, *SD* 0.84), suggesting that the mean for the variable given in Table 3 is largely determined by the higher age of the depressed group and would be of a more positive value (more different from the brain damage group), if groups were of the same age.

In the present sample the relative risk of brain damage compared to depression is 2.03 (CI 1.07 – 3.87), when *learning sheet*, *learning minus reproduction* variable is <0, while it is 0.58 (CI 0.41 – 0.82) when *learning sheet*, *learning minus reproduction* variable is ≥0.

Sensitivity and specificity

The sensitivity and specificity of the Visual Gestalt Test, *sum of learning plus reproduction error sheets* are shown in Table 7. For a cut-off score of 3 and above, 83 of 100 diagnosed can be found within this limit, while 33 of 100 normal subjects would be falsely labelled. At cut-off 5 and above, 71 of 100 diagnosed would still be found, while 18 of 100 normal subjects would be falsely labelled. As may be seen in Table 7, these two cut-off points can be chosen as logical cut-offs, due to their relatively greater distance to the next values.

For the discrimination between brain damage and other aetiology for dysfunction, a cut-off point of zero or above can be chosen for the *number of error sheets in learning phase minus the number of error sheets in reproduction phase.* If so, the sensitivity for correct labelling of brain damage will be 59.2%, while the specificity with regard to correct labelling of depression will be 70.8%.

Table 3. *Visual Gestalts Test, error counts – means and (SD)*

	Normals $N = 153$	Right temporal lobe epilepsy $N = 12$	Left temporal lobe epilepsy $N = 12$	Primary generalized epilepsy $N = 15$	Right resected $N = 34$	Left resected $N = 26$	Depressed $N = 24$
Learning phase							
Immediate recall	1.73 (1.88)	3.67 (1.88)*	3.17 (2.82)	4.87 (2.67)***	6.29 (4.32)***	4.35 (3.35)***	5.29 (3.68)***
Total errors	1.92 (2.26)	4.00 (2.17)	3.92 (3.99)	5.93 (4.32)**	10.15 (10.49)***	6.04 (6.40)***	8.54 (8.02)***
Number of error sheets	1.33 (1.40)	2.33 (1.07)	2.42 (2.19)	3.80 (2.33)***	5.24 (4.63)***	3.46 (2.63)***	4.47 (3.41)***
Reproduction phase							
Immediate retention	2.41 (3.10)	5.08 (4.48)*	6.83 (4.35)***	4.43 (3.96)	8.53 (4.17)***	6.23 (5.04)***	5.71 (4.31)***
Total errors	3.75 (5.61)	7.83 (7.84)	11.50 (8.66)***	7.57 (8.46)	14.29 (8.96)***	11.04 (10.51)***	8.75 (7.99)**
Number of error sheets	1.86 (2.48)	3.58 (3.42)	5.58 (3.48)***	3.71 (3.45)**	6.59 (3.42)***	5.00 (4.17)***	4.00 (3.09)***
Error sheets							
Learning plus reproduction	3.18 (3.30)	5.92 (4.01)*	8.00 (4.84)***	7.43 (4.43)**	11.82 (6.74)***	8.46 (5.97)***	8.67 (5.50)***
Learning minus reproduction	–0.53 (2.30)	–1.25 (3.10)	–3.17 (3.21)**	0 (3.96)	–1.35 (4.57)	–1.54 (3.58)	0.67 (3.48)

Notes: Significance between normal and diagnosed groups, LSD-corrected for multiple comparisons: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4. *Types of error in different diagnostic groups – means and (SD)*

Independent samples *t*-tests	Normals $N = 153$	Right temporal lobe epilepsy $N = 12$	Left temporal lobe epilepsy $N = 12$	Primary generalized epilepsy $N = 15$	Right resected $N = 34$	Left resected $N = 26$	Depressed $N = 24$
Omission, whole figure	0.10 (0.52)	0.17 (0.39)	0.75 (1.14)*	0.50 (0.86)	0.59 (0.78)**	0.77 (1.66)***	1.29 (1.97)***
Omission, subgestalt	0.62 (0.97)	1.25 (1.44)	1.17 (0.84)	1.79 (1.80)**	1.91 (1.87)***	1.62 (1.94)***	1.71 (1.73)***
Reversal rotation	0.27 (0.50)	0.33 (0.49)	1.00 (1.48)***	0.29 (0.61)	0.65 (0.77)**	0.54 (0.76)	0.46 (0.72)
Fragmentation	0.25 (0.60)	0.33 (0.49)	0.08 (0.29)	0.86 (0.95)**	1.09 (1.36)***	0.62 (0.94)*	0.71 (1.08)*
Unspecific lower level processing	0.41 (0.77)	1.00 (1.13)	1.42 (1.38)*	0.86 (1.10)	2.47 (2.56)***	1.54 (1.88)***	1.29 (1.43)**
Disorganisation	0.04 (0.20)	0.50 (0.91)	0.25 (0.45)	0.64 (1.73)	1.85 (3.26)***	0.62 (1.44)*	0.33 (1.05)
Fusion of figures	0.68 (1.13)	1.17 (1.64)	1.67 (1.50)*	1.07 (1.73)	1.44 (1.69)**	1.23 (1.45)*	0.63 (1.01)
Perseveration of single elements	0.39 (0.67)	0.33 (0.65)	0.50 (0.80)	0.79 (1.19)	0.62 (0.89)	0.42 (0.88)	0.92 (1.14)**
Stereotypy, unspec.	0.27 (0.60)	0.16 (0.39)	0.50 (0.67)	0.36 (0.50)	0.61 (0.92)*	0.58 (0.75)*	0.54 (1.10)
Transmission of error	0.14 (0.44)	0.58 (0.90)	0.33 (0.65)	0.14 (0.36)	0.88 (1.77)***	0.62 (1.13)*	0.61 (1.03)*
Other	0.16 (0.51)	0.66 (0.89)	0.67 (0.65)	0.28 (0.47)	0.79 (1.07)***	0.54 (0.90)*	0.88 (1.85)***

Notes: Significance between normal and diagnosed groups, LSD-corrected for multiple comparisons: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5. *Cross-tabulation of normals and diagnosed groups with percentage of pencil line quality score in categories*

	Category 0 % (N)	Category 1 % (N)	Category 2 % (N)	Category 3 % (N)	Category 4 % (N)
Normal control	81.7 (125)	3.9 (6)	5.2 (8)	9.2 (14)	
Right temporal Lobe epilepsy	41.7 (5)	50.0 (6)	8.3 (1)		
Left temporal Lobe epilepsy	66.7 (8)	33.3 (4)			
Primary generalized Epilepsy	26.7 (4)	60.0 (9)	13.3 (2)		
Right-resected epilepsy	52.9 (18)	2.9 (1)	8.8 (3)	29.4 (10)	5.9 (2)
Left-resected epilepsy	53.8 (14)	3.8 (1)	34.6 (9)	7.7 (2)	
Depression	45.8 (11)	4.2 (1)	25 (6)	25 (6)	

Table 6. *Comparisons between joint brain damage groups and depressed group – mean (SD)*

Error counts	Brain damage groups $N = 99$	Depressed group $N = 24$
Learning phase		
Immediate recall	4.87 (3.57)	5.29 (3.68)
Total errors	6.93 (7.65)	8.54 (8.02)
Number of error sheets	3.86 (3.42)	4.47 (3.41)
Reproduction phase		
Immediate retention	6.70 (4.61)	5.71 (4.31)
Total errors	11.34 (9.35)*	8.75 (7.99)*
Number of error sheets	5.27 (3.75)	4.00 (3.09)
Error sheets		
Learning plus reproduction	9.11 (6.02)	8.67 (5.50)
Learning minus reproduction	–1.42 (3.93)*	0.67 (3.48)*
Error types		
Omission, whole figure	0.59 (1.10)	1.29 (1.97)
Omission, subgestalt	1.64 (1.73)	1.71 (1.73)
Reversal rotation	0.57 (0.85)	0.46 (0.72)
Fragmentation	0.71 (1.07)	0.71 (1.08)
Unspecific lower level processing	1.68 (2.01)	1.29 (1.43)
Disorganisation	0.99 (2.25)	0.33 (1.05)
Fusion of figures	1.33 (1.59)*	0.63 (1.01)*
Perseveration of single elements	0.54 (0.89)	0.92 (1.14)
Stereotypy, unspec.	0.50 (0.75)	0.54 (1.10)
Transmission of error	0.60 (1.27)	0.61 (1.03)
Other	0.62 (0.89)	0.88 (1.85)

* Significance <0.05 in an age-corrected model, LSD-adjusted for multiple comparisons.

Table 7. *Normals and diagnosed: the total number of error sheets, sensitivity, specificity for different cut-off levels*

	Sensitivity (confidence level)	Specificity (confidence level)
Cut-off ≤ 2	0.86 (0.79–0.91)	0.54 (0.46–0.62)
Cut-off ≤ 3	0.83 (0.75–0.89)	0.66 (0.58–0.73)
Cut-off ≤ 4	0.75 (0.67–0.82)	0.75 (0.68–0.81)
Cut-off ≤ 5	0.71 (0.63–0.79)	0.82 (0.76–0.88)
Cut-off ≤ 6	0.60 (0.51–0.68)	0.84 (0.78–0.89)
Cut-off ≤ 7	0.52 (0.44–0.61)	0.87 (0.81–0.91)

Note: Sensitivity: quotient of diagnosed above the cut-off level; specificity: quotient of normals below the cut-off level.

DISCUSSION

The Visual Gestalt Test seems to be a neuropsychological instrument with good psychometric capabilities. According to the results of this study, the test can discriminate between the normal and the diagnosed groups in several ways. Most importantly, the test can also discriminate significantly between the depression and the brain damage groups – a traditionally difficult distinction of the utmost clinical importance (Theilgaard & Beckmann, 1971; Goldberg, 2001).

The data on the educational levels of the participants were not sufficient to draw conclusions on the associations of education and test results. The two categories of education years: "over 10 years" and "under 10 years" are not reflecting educational differences in a modern society. It is possible and even probable that the educational level, if registered in more detail, might interact strongly with the performance on the test, but it is still not known to what extent. Data from this study suggest that the main differences would be found in the learning phase, while difference in education may be of much less importance for storing and recall. When the gestalts first have been learned, the recall of them seems less sensitive to the education level. Based on experimental analyses (Miyake, Friedman, Rettinger, Shah & Hegarty, 2001; Gruber & von Cramon, 2003) and clinical experience with the test, it can be assumed that the test results are more influenced by *the area of work* than by the length of education. If persons are highly dependent on visual organization on a daily basis, for example in occupations as craftsmen, architects, or surveyors, this might be taken into account during examination of the test results.

The relations between performance and age also need further discussion. The age groups of 15–39, 40–49 and 50–59 years were formed by analysis of the distribution of the age variable. While all error counts increased with age, there was a tendency for this association to weaken around the age of

40, which is supported by the data in Table 1. Significant differences were found between age groups under 40 and the others, differences between age groups 40–49 and 50–59 were also found, but at a non-significant level.

The major shortcoming of this study may be the lack of data for age groups older than 59, an age group often represented in clinical neuropsychological testing. An effort to collect data on normal performance in the older age group still has to be made.

When observing the error counts (Table 3), it appears that all the diagnosed groups show lower results than the normal group. Most affected of the brain damage subjects are the groups of temporal lobe resection, who do worse than the non-operated groups. It also appears that the expected differences between right- and left-sided brain damage are most evident in these groups. That is: the right-sided should do worse than the left-sided on the visuo-spatial tasks of this test.

In the observation of the qualitative traits (Table 4) the same pattern appears: generally the resected groups have a greater amount and perceptually more severe deviations than the non-operated groups. For instance, the error types of fragmentation and disorganization are more frequent in the right-resected group, but the error type of rotation is unexpectedly most frequent in the left side, non-resected group. The performances of right- vs. left-sided brain-damage groups are not the topic of this study, and will be further elaborated in a later publication. But the results from this study may reflect that the left hemisphere plays a role also in spatial processing as demonstrated by, for example, Binder (1982), Mehta, Newcombe & Damasio (1987), and Mehta & Newcombe (1991). The results may also suggest that the Visual Gestalt Test is not always as exclusively nonverbal as intended. From clinical practice, it is found that the material gives the opportunity to support the visual perception by verbalizing, for example by counting the number of radii of the first model figure (see Fig. 1) or by comparison of parts of the figures to concrete verbalized objects.

A major finding of this study was the very simple measure of the quality of pencil line showing such significant results in relation to the basic distinction between the normal and the impaired. A simple observation of the pencil lines as being shaky, unstable, uncertain, or awkward might be a good reason to suspect an impaired function and to go further in the neuropsychological investigation.

The only diagnosed group with no brain damage in this study was the depression group. A noteworthy discrimination could be made between brain damage and depression with the Visual Gestalt Test, but it may also be expected that this difference will persist with other non-brain damage aetiologies for impaired cognitive function, such as schizophrenia or personality disturbances. It could be assumed that these groups also have relatively better memory when the gestalts are first learned, but this also has to be further supported by statistical data.

Perspectives: The clinical use of the Visual Gestalt Test

On the basis of this study, the differential diagnostic possibilities of the Visual Gestalt Test can be operational in a neuropsychological investigation with the following simple rules-of-thumb:

- Impaired neuropsychological function (brain damage or other aetiology) may be suspected, if the number of learning *plus* reproduction error sheets is three or above. Suspicion of impaired function should be highly increased, if the number of sheets in the two sequences exceeds five (see Table 7 for details).
- If *impairment is demonstrated*, the next step is to *subtract* the number of learning sheets from the number of reproduction error sheets. This number will point to the probability of a brain damage diagnosis if below zero; if the subtraction results in zero or above zero, a depression diagnosis (or other non-brain damage diagnosis) will be most likely (see Results for details).
- Pencil-line quality should be noted carefully. Suspicion of impairment should be raised, if the pencil line leaves an impression of insufficient control of the fine motor coordination, where the pencil lines are clumsy, rigid, awkward, squeezed, and possibly with varying degrees of pencil pressure, maybe shaky or with endings inexact. Performance may vary from sheet to sheet during the reproduction of the figures (see Table 5 for details).

REFERENCES

Andersen, R. (1968). Learning and reproduction of Visual Gestalts. *Nordisk Psykologi*, *20*, 101–103.

Andersen, R. (1976). Verbal and visuo-spatial memory. Two clinical tests administered to a group of normal subjects. *Scandinavian Journal of Psychology*, *17*, 198–204.

Andersen, R. (1978). Cognitive changes after amygdalotomy. *Neuropsychologia*, *16*, 439–451.

Andersen, R. (1989). *The Visual Gestalt Test. Learning and memory* (Rev. edn.). Copenhagen: Dansk Psykologisk Forlag.

Bender, L. (1946). *Instructions for the use of the Visual Motor Gestalt test*. New York: American Ortopsychiatric Association.

Benton, A. L. (1963). *The Revised Visual Retention Test*. (1st edn.) New York: Psychological Corporation.

Binder, L. M. (1982). Constructional strategies on complex figure drawings after unilateral brain damage. *Journal of Clinical Neuropsychology*, *4*, 51–58.

Bruhn, P. (1987). AIDS and dementia: A quantitative neuropsychological study of unselected Danish patients. *Acta Psychiatrica Scandinavica*, *76*, 443–447.

Cronholm, B. & Molander, L. (1957). Memory disturbances after electroconvulsive therapy. I. Conditions 6 hours after electroshock treatment. *Acta Psychiatrica Scandinavica*, *32*, 306.

Cronholm, B. & Ottosen, J. O. (1961). Memory functions in endogenous depression. Before and after electroconvulsive therapy. *Archives of General Psychiatry*, *5*, 193–199.

Drejer, K., Theilgaard, A., Teasdale, T., Schulsinger, F. & Goodwin, D. W. (1985). A prospective study of young men at high risk for alcoholism. *Neuropsychological Assessment. Alcoholism: Clinical and Experimental Research*, *9*, 498–502.

Gade, A., Mortensen, E. L. & Bruhn, P. (1988). "Chronic painter syndrome". A reanalysis of psychological test data in a group of diagnosed cases, based on comparisons with matched controls. *Acta Psychiatrica Scandinavica*, *77*, 293–306.

Goldberg, E. (2001). *Frontal lobes and the civilized mind*. Oxford University Press.

Gruber, O., & von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. *Neuroimage*, *19*, 797–809.

Hamby, S. L., Wilkins, J. W. & Barry, N. S. (1993). Organizational quality on the Rey-Osterreith and Taylor Complex Figure Tests: A new scoring system. *Psychological Assessment*, *5*, 27–33.

Hansen, H., Andersen, R., Theilgaard, A. & Lunn, V. (1982). Stereotactic psychosurgery. A psychiatric and psychological investigation of the effects and side effects of the interventions. *Acta Psychiatrica Scandinavica Suppl*, *301*, 1–123.

Hesche, J., Röder, E. & Theilgaard, A. (1978). Unilateral and Bilateral ECT. *Acta Psychiatrica Scandinavica Suppl*, *275*, 180.

Jensen, I. (1977). Temporal lobe epilepsy. *Danish Medical Bulletin*, *24*, 235–240.

Mehta, Z., Newcombe, F. & Damasio, H. (1987). A left hemisphere contribution to visuospatial processing. *Cortex*, *23*, 447–461.

Mehta, Z. & Newcombe, F. (1991). A role for the left hemisphere in spatial processing. *Cortex*, *27*, 153–167.

Mikkelsen, S., Jorgensen, M., Browne, E. & Gyldensted, C. (1988). Mixed solvent exposure and organic brain damage. A study of painters. *Acta Neurologica Scandinavica*, *118*, 1–143.

Milner, B. (1968). Visual recognition and recall after temporal-lobe excision in man. *Neuropsychologia*, *6*, 191–209.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P. & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

Österreith, P. (1946). Le test de copie d'une figure complexe. *Les Archives de Psychologie*, *31*, 206–356.

Penfield, W. & Milner, B. (1958). Memory deficits produced by bilateral lesions in the hippocampal zone. *AMA Archives of Neurology and Psychiatry*, *79*, 475–497.

Piotrowski, Z. (1937). The Rorschach inkblot method in organic disturbances of the central nervous system. *Journal of Nervous and Mental Diseases*, *86*, 525–537.

Rasmussen, K., Jeppesen, H. J. & Sabroe, S. (1993). Psychometric tests for assessment of brain function after solvent exposure. *American Journal of Industrial Medicine*, *24*, 553–565.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopatie traumatiques. *Archives de Psychologie*, *28*, 286–340.

Rosenberg, N. & Andersen, R. (1990). Rorschach-profile in panic disorder. *Scandinavian Journal of Psychology*, *31*, 99–109.

Sørensen, A. S., Hansen, H., Andersen, R., Høgenhaven, H., Allerup, P. & Bolwig, T. G. (1989). Personality characteristics and epilepsy. *Acta Psychiatrica Scandinavica*, *80*, 620–631.

Teasdale, T., Hansen, H. S. & Gade, A. (1997). Neuropsychological test scores before and after brain-injury rehabilitation in relation to return to employment. *Neuropsychological Rehabilitation*, *7*, 23–42.

Theilgaard, A. (1984). A psychological study of the personalities of XYY- and XXY-men. *Acta Psychiatrica Scandinavica Suppl*, *315*, 1–133.

Theilgaard, A. & Beckmann, J. (1971). Psyciatric diagnoses and psychological test conclusions. *Acta Psychiatrica Scandinavia*, *47*, 186–205.

Wechsler, D. A. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, *19*, 87–95.

Willanger, R. (1970). *Intellectual impairment in diffuse cerebral lesions*. Munksgaard: Copenhagen.

Received 3 February 2004, accepted 5 May 2005