

**VALIDATION OF GRAPHOLOGICAL JUDGMENTS:
AN EXPERIMENTAL STUDY**

PSYCHOLOGICAL STUDIES

8

Editorial Committee

JOHAN T. BARENDREGT

HANS C. BRENGELMANN / GÖSTA EKMAN†

SIPKE D. FOKKEMA / NICO H. FRIJDA

JOHN P. VAN DE GEER / ADRIAAN D. DE GROOT

MAUK MULDER / RONALD WILIAM RAMSAY

SIES WIEGERSMA

Mouton - The Hague - Paris

Validation of Graphological Judgments: An Experimental Study

ABRAHAM JANSEN

University of Amsterdam

Mouton - The Hague - Paris

This book has been published with the aid of the Netherlands Organisation
for the Advancement of Pure Research (Z.W.O.)

Translated from the Dutch by J.A.A. Spiekerman

Library of Congress Catalog Card Number: 76-184753

© 1973, Mouton Publishers, The Hague, Netherlands

Printed in the Netherlands

P R E F A C E

In 1955, on the initiative of the Nederlandse Vereniging voor Bedrijfspsychologie (Netherlands Society of Industrial Psychology), a study group for graphology was set up for the purpose of studying some problems of applied graphology, particularly with regard to the practical uses of graphology in industry. Its activities almost exclusively consisted of validation research — carried out by the group itself —, which took the form of four related experiments. The study group continued to exist until the end of 1963, when it was officially dissolved and a report submitted by its Chairman to the Board of the Netherlands Society of Industrial Psychology. The study group was composed as follows:

Prof. H.W. Ouwelen (Univ. of Rotterdam)	Chairman, Psychologist
M.A. van den Hout M.A. (The Hague)	Secretary
Prof. C.J.F. Böttcher (Univ. of Leiden)	Graphologist
Prof. A.D. de Groot (Municipal Univ. of Amsterdam)	Psychologist
E.A. Hof M.A., LL.D. (The Hague)	Psychologist
Prof. D.J. van Lennep (Univ. of Utrecht)	Psychologist, Graphologist
Dr. L. Spanjaard LL.D. (Amsterdam)	Graphologist
J.J. Wittenberg (Amsterdam)	Graphologist
Prof. H.R. Wijngaarden (Free Univ. of Amsterdam)	Psychologist
The author (Municipal Univ. of Amsterdam)	Psychologist

Dr. Spanjaard and Mr. Wittenberg were at the time serving on the board of the largest graphological society in the Netherlands. Owing to pressure of work, Prof. van Lennep, Prof. Böttcher and Prof. Wijngaarden successively had to resign from the study group.

VI *Preface*

The following publications dealing with the research conducted by the study group have appeared:

Ouweleen, H.W., Een wetenschappelijk onderzoek van grafologische uitspraken (A Scientific Investigation of Graphological Judgments), The Hague: Netherlands Society of Industrial Psychology, 1963, (16 pp.).

Jansen, A., Toetsing van grafologische uitspraken (Testing Graphological Judgments), Amsterdam: van Rossen, 1963 (dissertation, 278 pp.).

The former is a concise pamphlet, comprising the Chairman's final report on the study group's activities. It describes in broad outline the experimental research and sets out the main results. For more detailed information the pamphlet refers to the present author's dissertation, a revised and translated edition of which is here offered to the reader.

The direction and scope of the research, as well as the basic design of the experiments, were decided by the study group as a whole. All important decisions were taken unanimously — sometimes after prolonged discussions. The actual experimentation, the statistical calculations and analysis of the outcomes, as well as internal reporting to the study group were the responsibility of the author, who wishes to state emphatically that he is solely answerable for the presentation and conclusions in the *present* book.

The author has pleasure in extending his thanks to the Chairman and members of the study group. In particular, he owes a large debt of gratitude to Prof. A.D. de Groot, who supervised his doctoral dissertation. Thanks are also due to all who participated in the experiments as well as to the authorities and individuals who made possible the publication of the Dutch and English editions of this book.

TABLE OF CONTENTS

PREFACE	V
1 PRELIMINARY DISCUSSIONS AND INVESTIGATIONS	1
1.1 <i>Two basic questions</i>	1
1.2 <i>Types of validation study</i>	2
1.3 <i>The criterion</i>	6
1.4 <i>An examination of the claims of graphologists</i>	8
1.5 <i>Aim and design of Experiment I</i>	14
2 EXPERIMENT I: A PILOT INVESTIGATION	15
2.1 <i>Procedure</i>	15
2.1.1 The scripts	15
2.1.2 The instructions	16
2.1.3 The rating blank	17
2.1.4 The subjects	17
2.1.5 The actual experiment	17
2.1.6 Some hypotheses	18
2.2 <i>Statistical calculations and analysis of the results</i>	19
2.2.1 Explanation of the symbols and terms. A first survey	19
2.2.2 The experimental material: The five folders	20
2.2.3 The table of hits: The results per group	23
2.2.4 The scores: Individual results	24
2.2.5 The group judgments	25
2.2.6 The certainty of the judgments	26
2.2.7 The discrepancies in the (+) judgments	26

VIII *Table of contents*

2.2.8	Analysis of variance of the hit results on 58 ink scripts	28
2.2.9	Analysis of variance of the certainty qualifications on 58 ink scripts	29
2.2.10	The relation between hit result and certainty score	31
2.2.11	Intersubjective reliability	33
2.2.12	Determination of $\varphi_{.5}$ on the S scripts	35
2.2.13	A comparison of the results of psychologists and advanced students of psychology	36
2.2.14	The time factor	36
2.3	<i>Conclusions from Experiment I</i>	38
2.3.1	The hypotheses	38
2.3.2	Final conclusions	38
2.4	<i>Criticism of Experiment I</i>	39
3	EXPERIMENT II	43
3.1	<i>Procedure</i>	43
3.1.1	Aim and design	43
3.1.2	The definitions of the terms energetic and weak	44
3.1.3	The scripts	44
3.1.4	The instructions	45
3.1.5	The rating blank	45
3.1.6	The subjects	46
3.1.7	The actual experiment	46
3.1.8	Some hypotheses	47
3.2	<i>Statistical calculations and analysis of the results</i>	48
3.2.1	A first survey	48
3.2.2	The experimental material: The two folders	48
3.2.3	The table of hits	50
3.2.4	The scores	51
3.2.5	The group judgments per script	51
3.2.6	The certainty of the judgments	52
3.2.7	The discrepancies in the (+) judgments	52

Table of contents IX

3.2.8	Analysis of variance of the hit results	53
3.2.9	Analysis of variance of the certainty qualifications	54
3.2.10	The relation between hit result and certainty score	56
3.2.11	Some reliability indices	57
3.2.12	Determination of $\varphi_{.5}$ on the S_2 scripts	58
3.2.13	A comparison of the results of graduate psychologists and advanced students of psychology	59
3.2.14	The time factor	60
3.2.15	The rank correlations of 16 subjects based on their scores in Experiments I and II	60
3.3	<i>Conclusions from Experiment II</i>	62
3.3.1	The hypotheses	62
3.3.2	Final conclusions	63
3.4	<i>Criticism of Experiment II</i>	65
4	CRITERION ANALYSIS AND SELECTION OF SCRIPT MATERIAL FOR EXPERIMENTS III AND IV	68
4.1	<i>Design of Experiments III and IV</i>	68
4.2	<i>The rating schedules of the transport company</i>	68
4.3	<i>A factor analysis</i>	71
4.4	<i>Selection of 20 scripts for Experiment III</i>	73
4.5	<i>Selection of 9 scripts for Experiment IV</i>	74
5	EXPERIMENT III	77
5.1	<i>Procedure</i>	77
5.1.1	Aim and design	77
5.1.2	The definitions of the terms energetic and weak	78
5.1.3	The scripts	78
5.1.4	The instructions	79
5.1.5	The rating blank	79
5.1.6	The subjects	79
5.1.7	The actual experiment	80

X Table of contents

5.1.8	The hypotheses	80
5.2	<i>Statistical calculations and analysis of the results</i>	81
5.2.1	The table of hits	81
5.2.2	The scores	83
5.2.3	The group judgments	84
5.2.4	The certainty of the judgments	84
5.2.5	The discrepancies in the (+) judgments	85
5.2.6	Analysis of variance of the hit results	85
5.2.7	Analysis of variance of the certainty qualifications	87
5.2.8	The relation between hit result and certainty score	88
5.2.9	A reliability index	90
5.2.10	Determination of $\varphi_{.5}$	91
5.2.11	The time factor	91
5.2.12	The coefficient of concordance W	92
5.3	<i>Conclusions from Experiment III</i>	93
5.3.1	The hypotheses	93
5.3.2	Final conclusions	94
5.4	<i>Criticism of Experiment III</i>	96
6	EXPERIMENT IV	99
6.1	<i>Procedure</i>	99
6.1.1	Aim and design	99
6.1.2	The scripts	99
6.1.3	The rating blank	100
6.1.4	The instructions	101
6.1.5	The subjects	102
6.1.6	The actual experiment	103
6.1.7	Some hypotheses	103
6.2	<i>Statistical calculations and analysis of the results</i>	104
6.2.1	The frequencies of the ratings on 18 variables	104
6.2.2	The correlations per subject (N = 162: 9 scripts and 18 variables)	105
6.2.3	The correlations per subject per variable (N = 9 : 9 scripts)	106

Table of contents XI

6.2.4	The correlations per subject, computed from the summed scores per factor group (N = 9 : 9 scripts)	108
6.2.5	Validation of profile hypotheses	110
6.2.6	The deviations from the criterion	112
6.2.7	The correlations per subject of the blanket factor judgments (N = 9 : 9 scripts)	113
6.3	<i>Conclusions from Experiment IV</i>	115
6.3.1	The hypotheses	115
6.3.2	Final conclusions	116
6.4	<i>Criticism of Experiment IV</i>	117
6.4.1	Criticism by the subjects	117
6.4.2	Critical observations with regard to the criterion	119
6.4.3	Critical observations on the contents of the scripts	120
7	SUMMARY AND FINAL OBSERVATIONS	123
7.1	<i>Summary</i>	123
7.2	<i>Final observations</i>	126
	APPENDICES	131
	APPENDIX A: TABLES AND DOCUMENTS	132
A-1	A. A survey of the 4 validation experiments	133
	B. A survey of the script material of the 4 validation experiments	134
	<i>Experiment I</i>	
A-2	The instructions (folder V)	135
A-3	Job descriptions and definitions of the concepts energetic and weak in the instructions for folders S-V-K ₁ and K ₂ -P	136
A-4	The rating blank	138
A-5	Table of hits of the graphologists	139
A-6	Hit totals of the 3 groups of subjects	140
A-7	Scores: the number of hits (+0?)	141

XII *Table of contents*

A-8	Score on 58 ink scripts in class distribution	142
A-9	Percentages of the judgments in the 3 certainty categories	142
A-10	Discrepancy table of the graphologists	143
A-11	Correlation table hit score – certainty score per script (30 subjects)	144
A-12	Hit totals of 5 graduate psychologists and 5 undergraduate students of psychology	145
<i>Experiment II</i>		
A-13	The instructions (folder S ₂)	145
A-14	The rating blank	147
A-15	Hit totals of the 4 groups of subjects	148
A-16	Scores on the 20 scripts of folder S ₂	149
A-17	Percentages of the judgments in 3 certainty categories	150
A-18	Discrepancy table of the graphologists	150
A-19	Correlation table hit score – certainty score per script (40 subjects)	151
<i>Experiments I-II</i>		
A-20	Correlation tables of judgments given in Experiment I and II on 5 identical scripts by identical subjects	152
<i>Experiment II</i>		
A-21	Hit totals of 5 graduate psychologists and 5 undergraduate students of psychology	153
<i>Experiments I-II-III</i>		
A-22	Rank orders of 18 subjects based on their scores in Exp. I (ink scripts) – Exp. I (folder S) – Exp. II (folder S ₂) – Exp. III (folder K ₃)	154
<i>Criterion analysis</i>		
A-23	Correlation matrix	155
A-24	Factor matrix	156
A-25	Business personnel ratings of the employees whose scripts were included in the material for Experiment III	157
A-26	Business personnel ratings of the employees whose scripts were included in the material for Experiment IV	158

Experiment III

A-27 The instructions	159
A-28 The rating blank	161
A-29 Hit totals of the 3 groups of subjects	162
A-30 Scores on 20 scripts	163
A-31 Percentages of the judgments in 3 certainty categories	164
A-32 Discrepancy table of the graphologists	164
A-33 Correlation hit score — certainty score per script (30 subjects)	165

Experiment IV

A-34 The instructions	166
A-35 The rating blank	168
A-36 Form for comments	167
A-37 The frequencies of the ratings on 18 variables	170
A-38 The results of the judgments given with reserve	171
A-39 Correspondence between business criterion and ratings on 18 variables: the correlation coefficients τ per subject (N = 162 : 9 scripts and 18 variables)	171
A-40 Correspondence between business criterion and ratings on 18 variables: the correlation coefficients τ per subject per variable (N = 9 : 9 scripts)	172
A-41 Correspondence between business criterion and ratings on 18 variables: the correlation coefficients τ computed from the rank orders of the summed scores (N = 9 : 9 scripts)	173
A-42 Correspondence between business criterion and ratings on 18 variables: testing of profile hypotheses	175
A-43 The deviations from the criterion of the ratings on 18 variables	176
A-44 Correspondence between business criterion and ratings on 3 blanket factors: the correlation coefficients τ per subject (N = 9 : 9 scripts)	177

XIV *Table of contents*

APPENDIX B: REFERENCES	179
B-1 <i>Works on graphology</i>	179
B-2 <i>General works</i>	186
B-3 <i>Some more recent works on graphology</i>	188

Ich verstehe unter dem Wort Graphologie fortan die *Technik* der Handschriftendeutung, soweit sie verankert ist in gesicherten, das ist *beweisbaren* und bewiesenen, Theorien der Psychologie des Schreibens.
LUDWIG KLAGES — Was die Graphologie nicht kann (1949).

Regardless of one's theory about personality and regardless of one's choice of data, whether Rorschach, MMPI, Bender, age, marital status; regardless of how these data are fused for predictive purposes — by intuition, table, equation, or rational hypotheses developed in a case conference — the honest clinician cannot avoid the question "Am I doing better than I could do by flipping pennies?"
PAUL E. MEEHL — Clinical versus statistical prediction (1954).

It would be an error to imagine that by denigrating statistics we can manage to avoid it; opposition to statistics too frequently merely means the use of bad and inefficient statistics.
H.J. EYSENCK — Uses and abuses of psychology (1954).

1 PRELIMINARY DISCUSSIONS AND INVESTIGATIONS

1.1 TWO BASIC QUESTIONS

The main question which the study group undertook to study concerned the practical value of graphology, or more concretely, as set down in the study group's terms of reference: 'What scientifically warranted use can industry make of graphology?'

The first issue that had to be settled was whether an unambiguous answer to this question could be given on the basis of the literature. In the opinion of the members of the study group this was not possible. While there exists an extensive graphological literature, the number of critical empirical studies is fairly small, and strictly designed and rigorously controlled validation experiments were almost non-existent at the time. Moreover, the results of the few studies that could more or less qualify in this respect were sometimes contradictory. See Appendix B, for the literature up to 1955.

At an early stage, therefore, the study group decided to undertake its own validation research by carrying out an experimental investigation to test the value of graphological judgments. This decision led to the formulation of the following two questions:

- With respect to what psychological variables do graphologists consider themselves qualified to make judgments?
- How can judgments of this kind be tested?

The answers to these two basic questions, it was agreed, were to determine the direction and character of the investigation. Both questions were dealt with simultaneously. We shall first address ourselves to the latter.

2 Preliminary discussions and investigations

1.2 TYPES OF VALIDATION STUDY

Fundamentally, two distinct forms of graphological validation research present themselves:

- A) The testing of hypotheses directly derived from graphological theory. There is no need to use graphologists as experimental subjects, the only purpose being verification of certain relations postulated in the theory between script* characteristics and personality variables. This type of study will enlarge our knowledge of graphological theory.
- B) The testing of graphological pronouncements or judgments. In this case graphologists will be used as subjects, while groups of non-graphologists may serve as controls. The purpose is verification of judgments made by graphologists on the basis of any or all of the script material, irrespective of the manner in which they have arrived at their judgments. This type of study will enlarge our knowledge of graphological practice.

This B type can be subdivided as follows:

- B₁. The testing of judgments about the total personality couched in a more or less complete personality sketch.
- B₂. The testing of graphological judgments of a single personality trait.

These two forms will be discussed briefly.

B₁. Graphological judgments of the total personality can be validated by testing the graphologist's report against the findings of another judge, who has independently arrived at his own conclusions by some other method. If this comparative test is to be objective, the second judge must have set down his conclusions in advance. The objectivity requirement is evidently not fulfilled if the second judge merely examines the graphologist's report to determine how far he agrees with it. The suggestive influence of a written report, for instance, and the judge's own sympathetic or critical attitude towards graphology almost inevitably constitute disturbing factors which may well affect his own, still unformulated, conclusions. While from a methodological viewpoint such a procedure is objectionable in the extreme, we more than once come across this type of 'validation' in daily life.

* 'Script' is used throughout for handwriting (specimen).

Therefore, the second judge should in advance, and independently of the report of the graphologist, commit his personality judgment to paper, and then the two reports should be compared. This procedure, however, has some serious drawbacks too. Straightforward comparison is not practically feasible since the two judges will probably deal with different aspects of the personality and/or apply different terms to the same aspect, and, in general, use a different terminology. As a result, the reports will, if not entirely, then at least in large part, frustrate comparison. On the other hand, it may be argued that the total personality picture of the subject must somehow emerge from both reports in a recognizable form and therefore permit comparison. The question then arises as to how such 'agreement on the total personality picture' can be measured.

A solution has been found in the matching procedure. The first judge is asked to sketch the personalities of a number of individuals. These sketches are given to the second judge with the instruction to match them with the correct individuals, or rather, with the personality pictures he himself has formed, whether or not in writing.* Thus, one can have the graphological reports matched by, say, colleagues of the persons judged; conversely, personality sketches by colleagues plus the scripts can be given to graphologists for matching. Experiments of this kind have been conducted by Powers (Allport & Vernon, 1933), Bobertag (1929) and de Groot (1947). Powers found positive, but very slight, deviations from chance expectation. Out of ten required matchings — of a script and a personality sketch drawn up jointly by three psychologists — 17 professional graphologists scored an average of 2.41 correct matchings, as against an average of 1.80 correct matchings by laymen and 1.00 chance expectation. In an experiment by Bobertag, in which subjects were asked to match five graphological reports with five colleagues, a percentage of 80% correct was obtained, which is a very good result. However, Bobertag's analysis of his experimental material is open to serious criticism, since he discounted part of the obtained matchings in his

* If the personality judgments of the second judge are in writing, a third judge can do the matching, which again holds certain advantages.

4 *Preliminary discussions and investigations*

statistical calculations and thus flattered the percentage of successes. De Groot's study comprised, among other things, a replication of Bobertag's experiment; each of 13 subjects was asked to match five graphological reports with five colleagues. This time, however, the result (23% correct matchings) did not deviate from chance expectation.

An advantage of the matching procedure is certainly that it is of a holistic nature, *i.e.* it starts from the total personality. The importance of this point is commonly labored by many critics of graphological experiments. However, this procedure, too, has its disadvantages. The first drawback is of a methodological nature, since the matchings cannot be made altogether independently of each other; one wrong choice necessarily implies a second incorrect matching. Especially with small numbers of matchings this may prove a serious handicap, during the experiment to the subjects, and in the evaluation of the results to the experimenter. An improved experimental design has been proposed by Fluckiger, Tripp and Weinberg (1961), in which subjects could match several reports with one person and, on the other hand, one report with several persons. A modified statistical test was developed for this procedure.

A related drawback is the circumstance that the presence of one or more persons with outstanding or particularly strong character traits, in a small group of ratees, considerably facilitates the matching. It is not unlikely that some importance should also be attached to this factor in accounting for the difference between the results of Bobertag and de Groot — both of whom used a design calling for five matchings.

A third and last problem is that the matcher frequently agrees with some judgments in the graphological report but not with others. As a result, he meets with difficulties when he has to match the report as a whole. To obviate this problem, both Bobertag and de Groot have, in addition to the overall-comparative method (matching on the basis of the report as a whole), used an analytic-quantitative method, in which subjects indicated the applicability of separate report items, of single judgments. This refinement substantially supplements the data of the overall-comparative method, especially in those cases where the two

judges — the author of the report and the matcher — radically disagree on an item. This analytic-quantitative method, however, has its own drawbacks. If the matcher believes that a certain characteristic is present in average strength, he will practically always agree with every judgment of this trait in the report that is not too extreme. Moreover, it is not uncommon for judgments in a report to be formulated in such a way that one can hardly disagree with them. These almost inevitably correct matchings will flatter the applicability percentage.

These last two objections can be overcome by means of an experimental design which does not seek to test an entire report or some other collection of judgments of the total personality, but pronouncements on an antecedently selected personality variable. This is the B_2 -form of the validation methods differentiated on p. 2. The last variant of the matching procedure, *viz.* the analytic-quantitative method, closely approaches this B_2 -form; only, the starting-point is different, which has certain consequences.

B_2 . For this form of validation, a relevant personality variable is selected in advance, after which it is established how far graphological judgments agree with (criterion-) judgments obtained in a different way, but now only on this one variable. This can, for instance, be realized in the following way: a number of persons are rank ordered according to the degree in which the quality chosen is found. This is done, for instance, by a personnel manager or psychologist on the basis of his experience or of test outcomes. Then, one or more graphologists rank order the ratees, now on the basis of a graphological analysis. Finally, these rank orders obtained by different methods are compared. Another, even more rigorous, design is based on the method of contrasting groups. Two groups are selected — on a certain criterion — which differ strongly on one variable, for instance: intelligent-stupid or energetic-weak. Manuscripts of these persons are shuffled and graphologists are asked to say which scripts have been written by the high-ranking group and which belong to low-ranking subjects.

The great advantage of this 'one variable method' is that the criterion variable is completely under control and can be varied systematically, if desired. As has been pointed out above, groups