Computer-aided graphology - the basis for the scientific validation

Presented at the International Graphological Colloquium 2016

Dr. Marie Anne Nauer
Dr. Yury Chernov

For many decades, graphology has been criticised for the lack of the scientific evidence of its validity. Many studies delivered some positive results, however "the general trend of findings is to suggest that graphology is not a valid assessment method"[1]. Although the cited article, which is one of the most well known in the area of the validation research, is old, few has changed since then.

Preliminary remarks on qualitative versus quantitative research

The psychology of handwriting as the theoretic side of the handwriting analysis is a semiotic science: a science concluding from signs to facts, just like a physician who recognizes disease by symptoms, or like Freud discovered unconscious connections in the human psyche, or Sherlock Holmes concluded by certain traces on the doer of the crime. There was also, in the same historical period of the end of the 19th century, an art scientist who studied the signs of art forgers by specific details the latter did not consider enough: Giovanni Morelli (Ivan Lermolieff) could tell by the detailed forming of fingers or ears if the painting was authentic. This special method of reasoning is neither to be called inductive nor deductive, but after Peirce[2] , abductive or presumptive for its concluding quasi from the effect back to the cause. And we have to realize that, thus, we initially always deal with assumptions.

So, graphology is a science based on indication, as medicine and other social sciences, and does by its nature not belong to the quantifying sciences that are estimated higher since the times of Platon. Carlo Ginzburg[3] says: "*Actually, the use of mathematics and the experimental method infers to quantify respectively to repeat the things, while an individualizing science excludes the repeatability per definition and admits quantifying only as auxiliary method."*

For us it is important to keep an eye on that relationship while we realise the obligation to do quantifying research in order to legitimate our science. In our daily practice we are able to do this, and this will be always the starting point to any science adequate to our topic. We always start by observing individual cases, we draw upon certain experiences, we gather large amounts of data, and afterwards we formulate hypotheses to be verified or denied by an auxiliary quantifying procedure. And this means: If we do not understand anything about our

---

[1] Klimoski, J.R, Rafaeli, A.(1983). Inferring personal qualities through handwriting analysis. *Journal of Occupational Psychology, 56, pp. 191-202.*

[2] Peirce, C.S. (1976): Deduktion, Induktion und Hypothese, in: C.S. Peirce: Schriften zum Pragmatismus und Pragmatizismus, Hrsg. Von K.O. Apel, Frankfurt am Main. Zitiert nach Carlo Ginzburg (1983): Spurensicherung. Die Wissenschaft auf der Suche nach sich selbst. ISBN 3-8031-2677-1 (Anm. 22, S. 53)

[3] Ginzburg, Carlo (1983): Spurensicherung. Die Wissenschaft auf der Suche nach sich selbst. ISBN 3-8031-2677-1 (S. 23)

staple material, we will not be able to create a good research design. We may then run the risk of asking the wrong questions – as in some examples to be showed below – and of doing research in the wrong direction. Yet, the two methods are always influencing one another.

Problems of graphology as a validation object

The request for a thorough validation of a psychological method is a very basic one: "Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests."[4] That is why, if we want graphology to be treated as a scientific and well-established method, we should provide the data, studies and results on its validity.

Besides the objective reason behind the lack of proper validation results (it is extremely difficult to arrange the adequate experiments and to obtain and evaluate statistically enough data with required quality), there are several subjective problems of graphology that make the aim extremely challenging. We would highlight the following ones:

- Subjectivity - practical work based on individual training and personal experience that are not properly validated;
- Poor systematisation - no full systematic description of handwriting signs and psychological concepts behind the handwriting analysis;
- Lack of standardisation and consolidation - many graphological systems, no consensus, no integration.
- No further development - lack of research, lack of connections to academic bodies.
- Education – no new blood - the area is not attractive, not promising, poor reputation (pseudo-science).
- Non-friendly environment - negative development of the external conditions for handwriting in general.

Scientific approach is required

We can improve the poor reputation of handwriting psychology only by delivering scientific results. Moreover, we must do that in a form accepted today in the academic psychology. Psychology has been moving along a natural scientific axis. Therefore handwriting analysis must follow this track.

The only reasonable and reliable way to provide good results is, firstly, to use computer-aided handwriting analysis, and, secondly, to establish a systematic approach to the research. Instead of separate studies, we should build one methodological platform and try to reach two major goals of the validation research:

- To bridge the gap between the high potential of the handwriting analysis as a psychometric assessment method and the lack of proper validation;
- To adapt graphology to the fast changing practice of handwriting.

---

[4] Standards for Educational and Psychological Testing. (2014). *American Educational Research Association, American Psychological Association, National Council on Measurement in Education.*

Graphology as a psychometric method

Three major things must be done to achieve the formulated goals. The first one we already mentioned - the move to the natural science, i.e. to statistic methods. The second step is the establishing of consolidated, integrated graphology rather than a set of different schools, approaches and traditions. The third one is the working out of models that investigate the handwriting analysis against the standard quality requirements to the test procedures, namely objectivity, reliability and validity.

Below we are presenting our approaches and some results of the research of these important requirements for graphology. Our work is based mostly on models and corresponding computer-aided implementation (system HSDetect). The system, by its definition and by the way it was built, covers the first two specific steps, i.e. statistical evaluations and integration of different sources of the graphological knowledge-base.

Of course, during our work we thoroughly studied and took into consideration many already existing publications and studies that cover different aspect of objectivity, reliability and validity of the handwriting analysis.

Objectivity

Objectivity of a tests procedure means that its result does not depend on the tester and conditions under which the test is conducted. They define three types of objectivity.

Implementation objectivity mean that the subject (him or herself) and conditions cannot influence the test. For handwriting psychology that is practically always the case. A person that delivers their handwriting cannot purposefully influence their handwriting (only if he/she himself knows graphology - and even then it will be difficult).

Evaluation objectivity - no influence of a tester on the evaluation of handwriting signs. This is achievable when the handwriting signs are clearly and unambiguously defined, and experts are well trained.

Interpretation objectivity means that there is no influence of a tester on the evaluation of personal traits. This last requirement is the most complicated. We know that different graphological schools and individual experts can sometime disagree upon the conclusions of the graphological expertise.

Below we present the experiment where objectivity was tested. The experiment is based on the data produced in the Bachelor Thesis by Boris Peterka[5] with further evaluation done by six leading SGG graphologists[6]. The experiment involved 81 subjects, each of whom delivered three handwriting samples of the same text done on the graphical tablet Wacom Intuos3 A4, and normal samples of their handwriting. The tablet registered the movements of

[5] Peterka, B. (2009). Tempo und Sorgfalt in der Handschrift, eine computerunterstützte Studie. *Bachelorarbeit. ZHAW Angewandte Psychologie.*

[6] Nauer, M.A., Chernov, Y., Peterka, B. (2012). The Electronic Tablet: An Investigation in Graphological Validity. *Graphology, No. 90, pp. 15-22.*

a special pen, evaluated by the CSWin programon PC. The software evaluates the following handwriting variables:

- Time (total, on paper, in air)
- Path (total, on paper, in air)
- Speed (on paper, in air)
- Pressure
- Number of touches
- Number of up/down movements
- Frequency of movements

Afterwards expert-graphologists evaluated the corresponding handwritten signs from the normal writing (and not from the samples done on the graphical tablet). They covered such signs as speed, pressure, size, fullness, connectivity and variation of size and pressure. The evaluation was done on a ninth scale, i.e. from one (the lowest sign value) to nine (the strongest degree of a sign).

In the first step, we checked by means of Kendall Concordance (*Fig. 1*) how well experts agree among themselves. The red line shows the boundary for 95% reliability for 81 subjects. When the value of the Kendall Coefficient[7] (blue line) is higher than this boundary, the agreement among experts is good. We can see that for all signs the agreement is high. Therefore, we can use the average values of expert evaluation for further research.

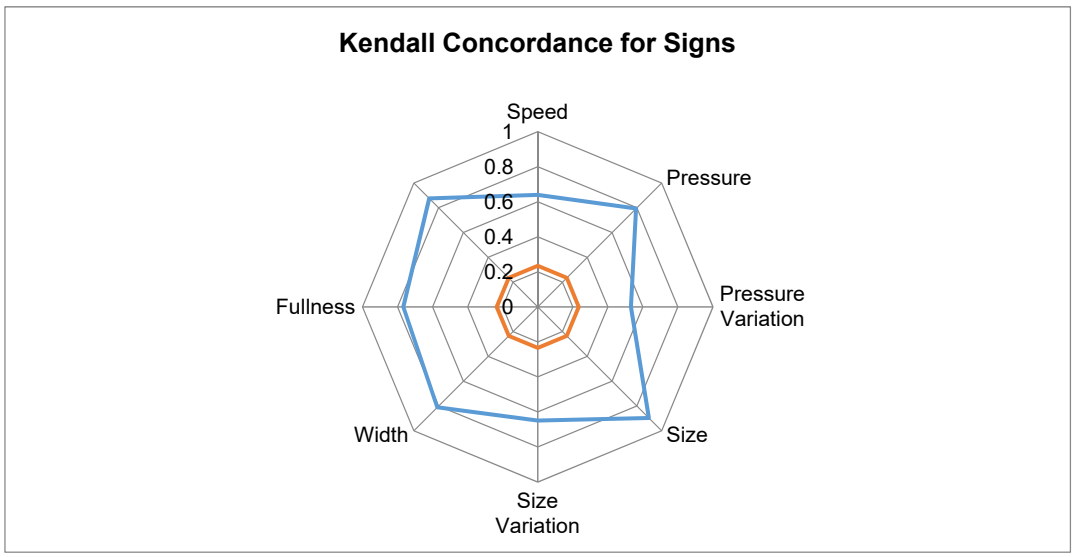

Fig 1.

In the second step, we compared the measured variables to the average values of expert evaluation. The results show that the correlation is high enough (*Fig.2*). Especially for the size, pressure and connectivity. That means two things. First, the handwriting samples done

[7] Kendall, M.G., Smith, B.B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics Vol. 10, No. 3, pp. 275-287*

under different conditions by the same subjects have close sign values. In other words, conditions do not influence the checked signs (implementation objectivity). Second, experts evaluate the signs well, because the results agree with the variables measured by a tool (evaluation objectivity).

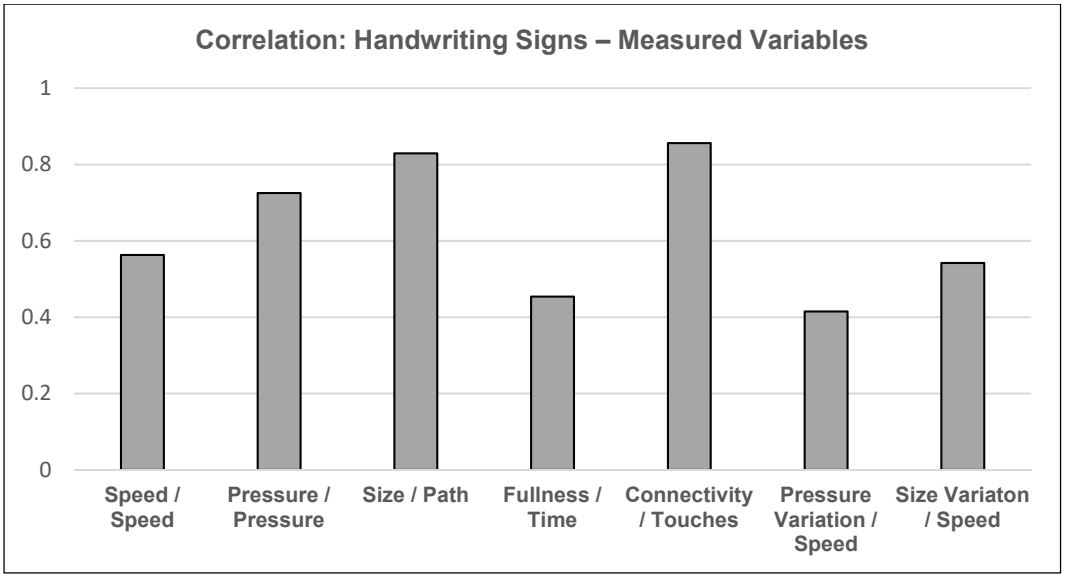

Fig. 2

Reliability

Reliability of a test procedure means that the method produces the same results under consistent conditions. Regarding the handwriting analysis, we can speak about two aspects. The first one is the ***reliability of evaluation*** – the same handwriting is similarly evaluated every time it is being analysed. In other words, when an expert takes the handwriting under investigation in a month, or in a year, or even in several years, he receives the same results. There are not many studies on the topic, however, with computer procedure and formally defined handwriting signs we can guarantee this reliability.

The second aspect is the ***reliability of handwriting*** itself. That is, the handwriting of the same person does not subsistently change from one sample to another. That is definitely a very complicated point. There is a lack of studies on this topic. To some extent, we discussed it above, saying that different writing patterns on the graphical tablet do not influence the measured variables. However, that is definitely not enough. The research of the reliability of handwriting must be provided in the future.

Validity

Validity is the most important quality criterion, it shows that a test procedure really evaluates what it is supposed to evaluate. Objectivity and reliability are just prerequisites for validity. Previously they used to define different types of validity like content validity, logical validity, criteria validity, or construct validity. In recent years, experts now define different sources of

validity evidence[8]. For such checks on objectivity researchers use either other psychological tests that are broadly accepted and well validated (***internal validity***), or non-test information (***external validity***).

There were many studies published on the validity of handwriting analysis. However, as the authors of the famous and already mentioned article wrote "…literature on this topic suffers from significant methodological negligence…"[9]. The typical problems of the validation studies lie in the difficulty of scientifically evaluating manual and subjective procedure of the handwriting analysis (this we hope to overcome with computer-aided graphology). Anyway, the major obstacle is the way the research procedures were defined. The typical deficiencies of validation studies are presented below (Tab. 1)

Table 1: Typical deficiencies of validation studies

Point	Examples
False premises	• The evaluations of graphologists do not agree with each other. • The graphologists base their reports rather on the content of the written autobiographic text.
Not valid experiments	• They take traits that cannot be evaluated by the handwriting analysis. • They take obviously improper text samples. • They take into consideration only a handful of handwriting signs and traits.
Wrong process	• Too primitive validation against personality tests. • Lack of statistical data.
Subjectivity	• Confusing experiments that deliberately lead to irrelevant results. • False interpretation of the experiment's outcomes.
Inadequate approach to handwriting analysis	• The handwriting analysis is made by the researcher himself. • Rather than evaluating handwriting analysis as a method , evaluate the particular graphologists.

Our approach to the improvement of the validation research is based on the following three principles:

1. Separating the handwriting analysis as a psychometric method from the interpretation as an assessment.
2. Usage of the formalised computer-aided handwriting analysis (ensure inner logic).
3. Appropriate modelling of psychometric constructs.

[8] Standards for Educational and Psychological Testing. (2014). *American Educational Research Association, American Psychological Association, National Council on Measurement in Education.*
[9] Klimoski, J.R, Rafaeli, (1983). A. Inferring personal qualities through handwriting analysis. *Journal of Occupational Psychology, 56, pp. 191-202.*

We describe these principles below and provide some results of the already conducted researches. Our experiments relate to the internal validity. The basic thing with internal validity is that the test against which they check a new procedure (in our case that is graphology) must be itself very well validated. Otherwise, it cannot serve as a proper reference.

Clean Separation of Method from Assessment

In one of the most influential articles, which is still referenced in all critical works against the practical use of the graphology, the authors pointed out: "We have often heard the objection that the fact that some graphologists may be incompetent does not invalidate the graphological enterprise itself. This may well be true, but until such a time as graphological analysis can be objectified, there is no way of testing the enterprise independently of its practitioners."[10] And they are right. So one of the important steps in the scientific validation of graphology is to separate the psychometric method itself from its practical implementation by graphologists. This we can achieve only by formalising the procedure and using the computer-aided handwriting analysis.

Computer-aided Handwriting Analysis: a hybrid solution with HSDetect system

For the validation researches shown below we used the computer-aided system HSDetect. The principles behind the system can be formulated as followers:

1. Consolidation, integration and harmonisation.
2. Quantitative presentation of all variables.
3. Non-ambiguity.
4. The black-box approach to the interpretation of the handwriting signs.
5. Openness und adaptability.

We will not discuss the principles in details here and will just refer to[11,12] where it has been carried out.

We call the approach implemented in HSDetect a hybrid one, since the assessment of the handwriting signs is done manually and the evaluation of the psychological traits is automated. In details HSDetect is described in some publications[13, 14] and was presented on several conferences and congresses[15, 16]. For our purposes it is enough to understand, that for

---

[10] G. Ben-Shakhar, M. Bar-Hillel, Y. Bilu, E. Ben-Abba, A. Plug. (1986). Can Graphology Predict Occupational Success? *Journal of Applied Psychology, Vol. 71, No. 4, pp. 645-653.*

[11] Chernov Y. (2014). Der Einsatz des Computers in der Graphologie. *Angewandte Graphologie und Persönlichkeitsdiagnostik, Heft 2, ss. 18-37.*

[12] Chernov Y. (2015). The Use of Computers in Graphology. *Graphology, No. 99, pp. 43-56.*

[13] Чернов, Ю. (2011). Психологический анализ почерка: системный подход и компьютерная реализация. *Москва: Генезис*. (Chernov, Y. Handwriting Psychology: System Analysis and Computer Implementation. Moscow: Genesis).

[14] Chernov Y. (2015). The Use of Computers in Graphology. *Graphology, No. 99, pp. 43-56.*

[15] Nauer, M.A, Chernov Y. (2015). Psychodynamic Diagnostic: Validation Research Based on Computer Modelling of Handwriting Psychology. *13th European Conference on Psychological Assessment, Zurich.*

every investigated handwriting sample we have a list of handwriting signs (sign protocol) quantitatively evaluated on the scale from zero to one. An expert does that manually, however based on clear guidelines. At the output we receive the list of all traits included in HSDectect (trait protocol) as well quantitatively evaluated on the scale from zero to one. These values are calculated by HSDetect. There are about four hundred traits. This data will be used further for the modelling of the test dimensions.

Modelling of psychometric constructs

The third improvement relates to the way we model the dimensions of the test, against which the research is being done. In known studies the researches would either compare the results of the test to one lexically similar trait, or ask graphologists (when they participate in the experiment) to directly evaluate the test dimension. Both methods are not good. Every dimension of a test is typically a complicated construct that should be modelled by several traits, and this is what we are doing in our experiments, which is possible due to the computer-aided graphology. For instance in EQ-i test for emotional intelligence the dimension Emotional Expression is modelled as follows:

- The positive pole is modelled with six traits – conciseness, naturalness, sincerity, emotionality, extroversion, and interpersonal skills;
- The negative pole is modelled with five traits – insincerity, artificiality, phlegmatic temperament, unsociability, and withdrawnness.

The logic behind the modelling is clear: the more traits are present in the handwriting, the higher is the value of the dimension. However the simple average of the trait values is a bad approach, since should one trait be not present in the handwriting, the average tends toward a very low value. Therefore a special model was developed, where we assign a higher weight for a trait with a higher value, and the sum of all weights equals one. Thus, a zero-value trait becomes a lower weight and its influence is much smaller. When analysing the model of the test scale both poles are always considered.

The second point in the modelling is that we check the mapping of traits on every test dimension with Cronbach-Alpha[17] that is a standard reliability criterion for the dimensions of questionnaire tests, and is appropriate in our case as well.

Algorithm of validation

The algorithm of validation against other psychometric tests is the same for all presented experiments: the subjects execute the test and provide a sample of their handwriting. We then evaluate the handwriting signs of the sample manually and import the sign protocol into HSDetect, which calculates trait values. Afterwards another program evaluates the model of

---

[16] Chernov Y. (2012). Computergestützte Graphologie: Ansätze, Modelle, Möglichkeiten und Grenzen. *26th Internationaler Kongress für Schriftpsychologie, Lindau*.

[17] Cronbach L.J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika, Vol.16, pp. 297-334.*

test scales based on the trait values, compares tests results to the model, and decides, whether they agree with each other or not.

Several words should be said about the comparison. Today the most popular method is to calculate the correlation coefficient. Many psychologists are doing that without even thinking whether this is appropriate or not. However, the correlation can be properly used only under certain conditions. First, the correlation works well for the implicitly measured variables and not for models, especially, when a model is not a regression one. Second, it checks only the linear dependence between two variables, which must be the same on the whole interval of variable values. Third, the variables must be normally distributed, which is often not the case.

That is why we use a different approach: for every scale we build three zones, namely high, middle and low and assign the results of the psychometric test and the model of the handwriting evaluation to these zones. When the result of both appear in the same zone we conclude that they agree, when they appear in opposite zones, we conclude disagreement; when they are in neighbouring zones we assume that the proper conclusion cannot be made. After this evaluation, we check the statistical confidence, based on the binomial distribution of the number of agreements and disagreements.

Below we present the results of three experiments. They are not yet finished. In the experiment with tests EQ-i and PVQ the number of subjects is still too small, so the results are preliminary. The 16PF-R experiment has enough subjects to make statistical conclusions, however we continue to add new subjects.

Experiment 1: Validation against 16PF-R

The Sixteen Personality Factor Questionnaire (16PF) was developed by R.B. Cattel and his colleagues[18], and the first version was issued in 1946. It is one of the most popular and well-evaluated tests in psychometrics. We have been using the revised version of the test 16PF-R[19]. The test consists of 184 items (questions) and evaluates 16 scales presented in Tab. 2.

Table 2: 16 PF Scales

Scale	Low Pole	High Pole
A	Reserve	Warmth
B	Reasoning bad	Reasoning good
C	Emotional Instability	Emotional Stability
E	Compliance	Dominance
F	Seriousness	Liveliness
G	Non-conformance	Rule-Consciousness
H	Timidity	Social Boldness
I	Objectivity	Sensitivity
L	Trustfulness	Vigilance

[18] Cattell, R. B., Cattell, A. K., & Cattell, H. E. P. (1993). 16PF Fifth Edition Questionnaire. *Champaign, IL: IPAT.*

[19] Schneewind, K.A., Graf J., Cattell, R.B. (1989). Der 16-Persönlichkeits-Faktoren-Test (16 PF). Bern : Hans Huber.

M	Practicality	Abstractedness
N	Openness	Privateness
O	Self-assurance	Apprehension
Q1	Resistance to Change	Openness to Change
Q2	Group-orientation	Self-Reliance
Q3	Casualty	Perfectionism
Q4	Relaxing	Tension

The experiment includes 54 subjects.[20] The results by scales are shown in Fig. 3. We can see that for nine scales the agreement between the test and the handwriting analysis is statically confident. Even more important is that there are no scales with statistically relevant disagreement.

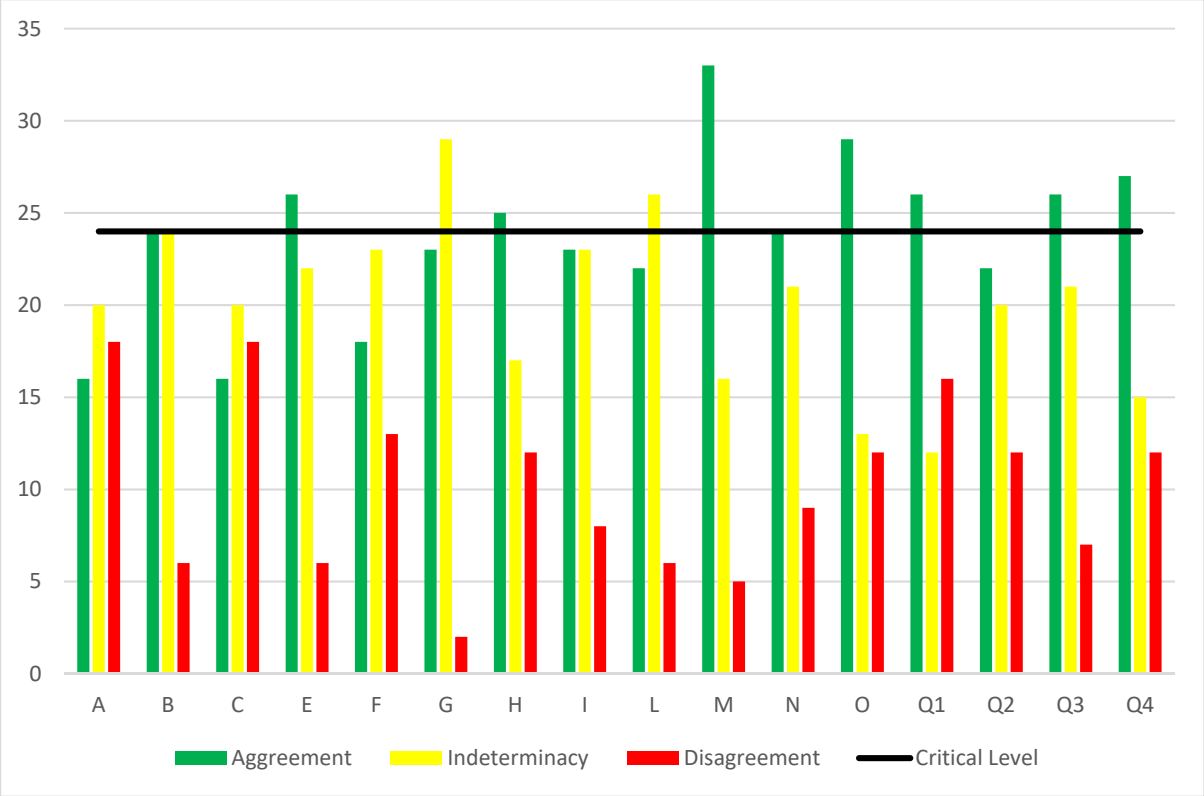

Fig. 3

[20] Chernov, Y., Caspers, C. (2015). Computergestützte Validierung in der Graphologie. *Angewandte Graphologie und Persönlichkeitsdiagnostik, Heft 2, ss. 16-19, Heft 3, ss. 41-53.*

Experiment 2: Validation against EQ-i

The Emotional Quotient Inventory (EQ-i) was developed to assess the Bar-On model[21] of emotional-social intelligence. It measures a number of constructs related to Emotional Intelligence (Tab. 3).

Table 3: EQ-i Scales

Super-Scale	Scale	Scale Name
Self-Perception	SR	Self-Regard
	SA	Self-Actualization
	SW	Emotional Self-Awareness
Self-Expression	EE	Emotional Expression
	AS	Assertiveness
	IN	Independence
Interpersonal	IR	Interpersonal Relationship
	EM	Empathy
	RS	Social Responsibility
Decision Making	PS	Problem Solving
	RT	Reality Testing
	IC	Impulse Control
Stress Management	FL	Flexibility
	ST	Stress Tolerance
	OP	Optimisms
Well-Being Indicator	HP	Happiness

The initial test data of EQ-i is for most scales normally distributed. Therefore, we can check the correlation between the test and the handwriting analysis (see Fig. 4).

[21] Bar-On, R. and Parker J.D.A (Ed.) (2000). The Handbook of Emotional Intelligence. *San Francisco: Josey-Bass a Wiley Company.*

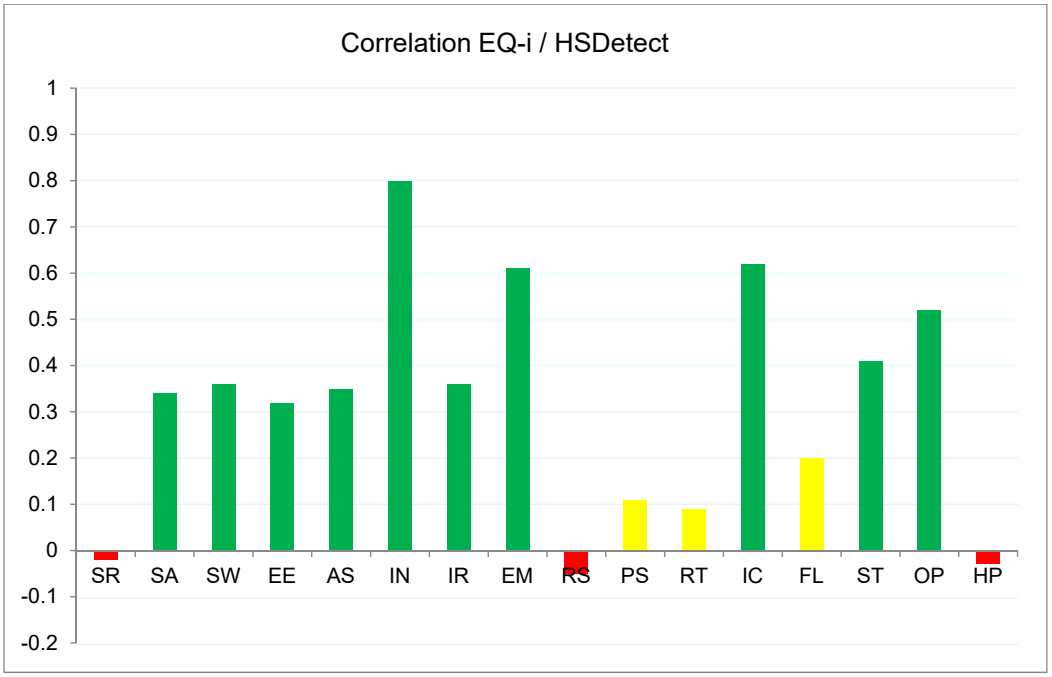

Fig. 4

Generally, the correlation for most of scales is rather good, in that the result of the handwriting analysis agree with the results of the test. Since the number of subjects, i.e. statistical points is still low, the significance level is high (0.549 for 95% reliability). So the correlation for scales IN, EM and IC is statistically significant. There are no scales with strong negative correlation, which is also a good sign. The general trend is positive, and with the increasing number of additional subjects we can expect better results.

Experiment 3: Validation against PVQ

Portrait Values Questionnaire (PVQ)[22] was designed to measure ten value orientations that S.H. Schwartz formulated in his theory of basic values. The values are listed in Tab. 4.

Table 4: PVQ Scales

Value	Value Name
AC	Achievement
BN	Benevolence
CF	Conformity
HD	Hedonism
PW	Power
SC	Security
SD	Self-Direction

[22] Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture 2 (1)*

ST	Stimulation
TR	Tradition
UN	Universalism

The questionnaire includes 40 items and the aim is not just to quantitatively evaluate the values, but rather to range them.

Three zones of ranks were defined: high, with three most high ranked values, middle, with four values and low, with three lowest ranks. The comparison for every subject is “Plus” (agreement) when the ranks from the test appear in the same zone as the results of the handwriting analysis; “Minus” – when they belong to opposite zones; “Neutral” – all other cases, when the definitive conclusion is not possible (see Fig. 5).

For seven basic values the agreement was registered more often. Five of them are statistically confident. For three values the disagreement dominates. So generally, the result is rather positive – the test agrees with the handwriting analysis.

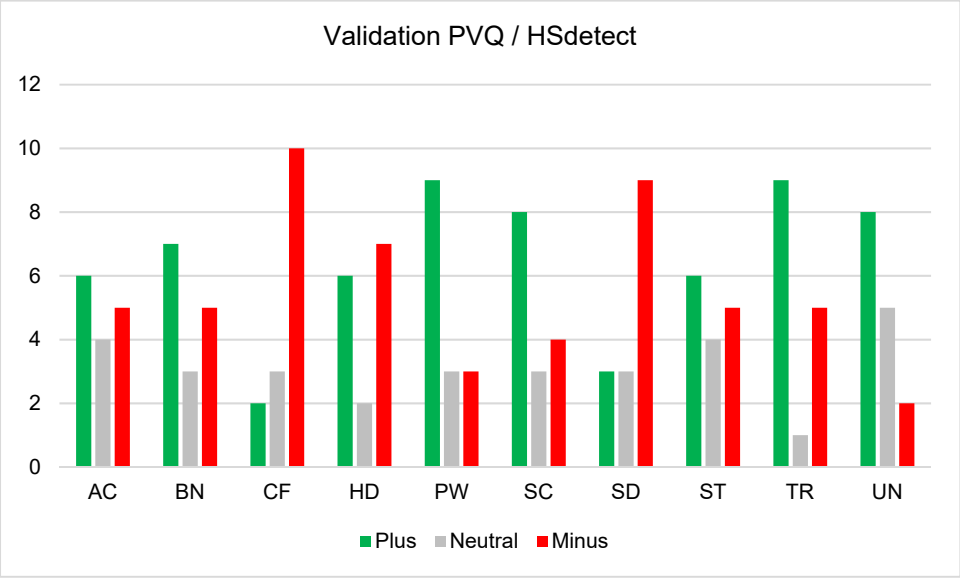

Fig. 5

Discussion

In the presented work we investigated the basic test quality criteria that are evaluated for the graphological analysis of personality. The results for objectivity and reliability are not new - their high level has been proved in many publications and are no longer disputed by the critics of graphology. The new computer-aided approach from one side just adds to this, from another side demonstrates its efficiency and perspectives in the studying of the major quality criterion, namely validation.

Promising results have been reached with the new approach to the validation research. The agreement between well-known psychometric tests and the handwriting analysis is rather high. It is much better than traditionally stated by the critics of graphology.

This was possible, due to the separation of the graphological method implemented as a computer-aided procedure, from the manual assessment done in practice. Therefore, the whole validation procedure is much more objective.

The obtained research results should be evaluated further. Based on the established procedure some additional psychological tests could be added. However, the major aim of the research is not just to prove or disprove the known graphological technique, but rather to pave the way to the better adaptation of it to the currently quickly changing writing habits.