GRAPHOLOGIENEWS

März/April 2015

Computergestützte Validierung in der Graphologie

Eine neue Methode und die ersten Ergebnisse

Dr. Yury Chernov und Claudia Caspers

Im menschlichen Leben werden in den unterschiedlichsten Bereichen Urteile und Prognosen über Menschen getroffen, sehr häufig durch den Einsatz psychodiagnostischer Verfahren. Da die psychologische Diagnostik häufig nicht regelgerecht in Form validierter Verfahren durchgeführt wird, hat das *Deutsche Institut für Normen* im Jahr 2002 auf Betreiben von Beratungsunternehmen, der Deutschen Gesellschaft für Psychologie (DGPs) und des Berufsverbands deutscher Psychologinnen und Psychologen (BDP) die DIN 33430 veröffentlicht, in der die Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen definiert werden. Mit der DIN 33430 werden Qualitätskriterien und -standards für berufsbezogene Eignungsbeurteilungen sowie für Qualifikationsanforderungen an die an der Eignungsbeurteilung beteiligten Personen beschrieben. Die Norm, die mittlerweile auch auf internationaler Ebene in Form der ISO 10667[1] existiert, ist kein Gesetz, an das sich alle Beteiligten sklavisch halten müssen, aber sie formuliert Standards, die sogenannte „best practice". „Durch die Festlegungen und Leitsätze ergeben sich indirekt auch Hinweise für die sach- und fachgerechte Entwicklung und Evaluierung von Tests, (Bewertungs-) Verfahren, Auswahlkriterien etc."[2] In Kürze wird eine Überarbeitung der DIN 33430 erwartet, die „die veränderten Rahmenbedingungen, insbesondere die Entwicklungen bei computergestützten Eignungstests sowie einzelne Festlegungen der ISO-Normen berücksichtigen"[3].

Insofern die Graphologie auch als Testverfahren betrachtet werden kann, das Urteile und Prognosen über Menschen u.a. im Rahmen der Eignungsbeurteilung fällt, müssen wir Graphologen uns selbst fragen, ob unser Verfahren als valide angesehen werden kann. Validität als „Bestätigung durch Bereitstellung eines objektiven Nachweises, dass die Anforderungen für einen spezifischen beabsichtigten Gebrauch oder eine spezifische beabsichtigte Anwendung erfüllt worden sind"[4] stellt ein zentrales Instrument der Qualitätssicherung dar. Sie lässt den Schluss von einem Verhalten in einer Testsituation auf ein Verhalten außerhalb der Testsituation zu und ist somit eines der wichtigsten Gütekriterien neben der Objektivität und Reliabilität.

Dadurch, dass die Graphologie in der Vergangenheit keine aussagekräftigen Validitätsergebnisse aufweisen konnte, stößt man immer wieder auf eine negative Fremdwahrnehmung der Graphologie. So z.B. Dr. Kanning, „(…) dass die Graphologie schon längst auf den Müllhaufen der Geschichte skurriler Ideen gehört"[5]. Auch andere ausländische Psy-

Dr. Yury Chernov
Holderbachweg 22
CH - 8046 Zürich
y.chernov@gmx.ch

Claudia Caspers
Postfach 1132
D - 85758 Oberschleißheim
claudia.caspers@graphologie.cc

1 http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56441

2 http://de.wikipedia.org/wiki/DIN_33430

3 http://de.wikipedia.org/wiki/DIN_33430#cite_note-5

4 https://www.uke.de/zentrale-dienste/medizinische-qualitaetssicherung/index_42923.php

5 Kanning, Dr. Uwe: Von Schädeldeutern und anderen Scharlatanen: Unseriöse Methoden der Psychodiagnostik, Pabst Science Publishers, 2009

chologen, die Validitätsuntersuchungen zur Graphologie durchgeführt haben, kommen zu folgenden Ergebnissen:

- Im Bereich der kognitiven Psychologie ein Zitat von Furnham: „It was concluded that the analyst could not accurately predict personality from handwriting.“[6]
- Im Bereich der Persönlichkeitspsychologie sprechen Eysenck & Gudjonsson von einer geringen Validität bei graphologischen Analysen: „There is thus little support here for the validity of graphological analysis."[7]
- Im Bereich der Angewandten Psychologie schreibt Ben-Shakhar: „The Graphologist did not perform significantly better than a chance model.“[8]
- Im Bereich der Berufspsychologie kommen Klimovski & Rafaeli zu folgendem Schluss: „Although the literature on the topic suffers from significant methodological negligence, the general trend of findings is to suggest that graphology is not a viable assessment tool“.[9]

Wenn man von der negativen Fremdwahrnehmung zur kritischen Selbstwahrnehmung übergeht, sind einige Defizite der graphologischen Methode nicht zu leugnen:

- Kaum einer von uns folgt einem standardisierten, wiederholbaren Analyse- und Ergebnisprozess — wir sind vielmehr sehr handwerklich und intuitiv begabte Individualisten, die ihrem eigenen System folgen.
- Erschwerend hinzukommt, dass seit Jahrzehnten keine systematische Überprüfung der graphologischen Theorieansätze und Methoden existiert, wodurch die Methode aufgrund mangelnder Forschung, empirischer Bestätigung und fehlender Weiterentwicklung stagniert.
- Des Weiteren ist die Integration mit anderen Methoden nicht gegeben, wodurch das Problem der Nicht-Vergleichbarkeit entsteht.
- Zuletzt sei noch kritisch angemerkt, dass kein systematisches Lehrmaterial sowie ein geringes Angebot an Graphologie-Lehrveranstaltungen, vor allem an Universitäten existiert. Wir haben also nicht nur mit dem Problem der „Nicht-Anerkennung“, sondern auch mit der „Nachwuchsproblematik“ zu kämpfen.

Damit sich diese Punkte langfristig ändern, müssen wir an diesen Punkten mit vereinten Kräften arbeiten und das gelingt nur, wenn wir uns mit einem der wichtigsten Gütekriterien, der Validität, befassen.

Gütekriterien psychologischer Testverfahren

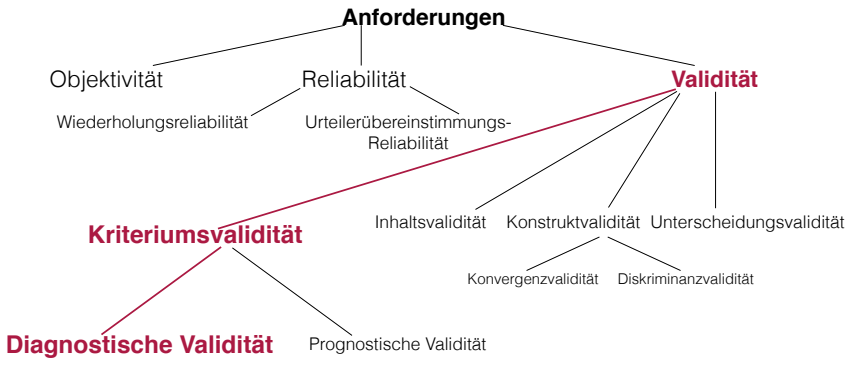
Die Güte eines Testverfahrens wird aufgrund von drei Parametern ermittelt: Objektivität, Reliabilität und Validität, wobei die beiden letztgenannten sich noch weiter untergliedern lassen wie man der nachfolgenden Grafik entnehmen kann. In unserem speziellen Fall konzentrieren wir uns auf die Kriteriumsvalidität, da ihre negativen Werte einer der wichtigsten Kritikpunkte an der Graphologie sind. Die Kriteriumsvalidität gibt an, ob eine Methode im Vergleich mit einer bereits anerkannten Methode zu ähnlichen Ergebnissen kommt.

6 Furnham, A., Chamorro-Premuzic, T., Callahan I. (2003): Does graphology predict personality and intelligence? Individual Differences Research, 1(2), 78 – 94

7 An empirical study of the validity of handwriting analysis. Personality and Individual Differences, 7, 263-264

8 Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., Ben-Abba, E., Flug, A. (1986). Can graphology predict occupational Success? Two empirical studies and some methodological ruminations. Journal of Applied Psychology 71(4), 645-653

9 Klimovski, R., Rafaeli A. (1983). Inferring personal qualities through handwriting analysis. Journal of Occupational Psychology, 56 (3), 191-202



Bisherige Ergebnisse zur Validität der Graphologie

Insofern die Graphologie bis dato noch keine überzeugenden Ergebnisse in Bezug auf die Validität geliefert hat, ist die Kritik der Graphologie-Gegner berechtigt. Allerdings mangelt es den Untersuchungen, die die Kritiker als Beweis für die mangelnde Validität der graphologischen Methode vorgenommen haben, selbst an einer wissenschaftlichen Basis[10]. Zudem sind die Autoren häufig nicht sehr objektiv. Ein Beispiel hierfür liefert uns ein Artikel von Furnham et al. aus dem Jahr 2003. Hier schreiben die Autoren: „In another review, Neter and Ben-Shaktar (1989) asked 63 graphologist and 51 non-graphologist to rate 1223 scripts“[11]. Man kann den Eindruck bekommen, dass die Autoren ein umfangreiches Experiment durchgeführt haben und uns jetzt die fundierten Ergebnisse zeigen. In Wirklichkeit ist der referenzierte Beitrag eine Meta-Analyse von 17 unterschiedlichen Arbeiten, die gar nicht einheitlich und eindeutig waren. Die Autoren haben mit keinen Graphologen zusammengearbeitet und verweisen auf eine summierte Anzahl von Teilnehmern unterschiedlicher Untersuchungen.

Da man solche Vorgehensweisen und die dadurch ermittelten Ergebnisse nicht in der Öffentlichkeit stehen lassen kann, sahen wir uns veranlasst, eine graphologische Validierungsuntersuchung durchzuführen. Da die Definition von „Validität“ generell sehr allgemein gehalten ist, lässt sie den Untersuchern viele Entscheidungsfreiräume in Bezug auf die jeweilige Durchführung des Validierungsprozesses. Im Folgenden werden wir die einzelnen Elemente des von uns umgesetzten Validierungsprozesses mittels einer für die Graphologie neuen Methode inklusive entsprechender Ergebnisse vorstellen.

Elemente einer graphologischen Validierungsuntersuchung

Einerseits existieren zahlreiche kritische Beiträge, die allerdings wenig wissenschaftlich fundiert sind in Bezug auf die Art und Weise, wie sie den graphologischen Teil modellieren. Die negativen Ergebnisse entstanden aufgrund des Vergleichs gegenüber angesehenen Persönlichkeitstests bzw. der Kriteriumsvalidität. Anderseits waren die Graphologen selbst bis jetzt nicht in der Lage, seriöse wissenschaftliche Widerlegungen in Bezug auf die Kriteriumsvalidität zu liefern. Das liegt primär an einer wenig formalisierten Weise der praktischen graphologischen Arbeit.

Um diese Lücke zu schließen, schlagen wir einen neuen Validierungsansatz vor. Das Neue besteht aus zwei Punkten: Zum einen verwenden wir in unserer Untersuchung eine Form der computergestützten Handschriftanalyse, das System HSDetect. Zum anderen modellieren wir bei der Kriteriumsvalidität die Dimensionen des Persönlichkeitstests durch einen Cluster von Persönlichkeitscharakteristiken und nicht nur durch eine einzelne Persönlichkeitseigenschaft wie es meistens gemacht wird.

10 Chernov, Y. (2014). Die Kontroverse um die Schriftpsychologie: Versuch einer kritischen Analyse. GraphologieNews, http://graphologie-news.net/cms/upload/archiv/kontroverse_schriftpsychologie_graphologie.pdf

11 Furnham, A., Chamorro-Premuzic, T., Callahan I. (2003). Does graphology predict personality and intelligence? Individual Differences Research, 1(2), S. 81

Das System HSDetect verfügt über Eigenschaften[12], die für seine Verwendung im Rahmen der Validitätsforschung besonders wichtig sind:

1. Konsolidierung: Die Integration der Informationen von mehreren Quellen (graphologische Bücher), die unterschiedliche graphologische Schulen repräsentieren, Damit sichert das Programm die Vollständigkeit des graphologischen Wissens und die Objektivität.

2. Quantitative Darstellung aller Handschriftmerkmale und Persönlichkeitseigenschaften: Quantitativ wird ihr Ausprägungsgrad in der untersuchten Handschriftprobe modelliert. Wir verwenden dafür eine Skala von 0 bis 1. Grundsätzlich spielt es keine Rolle, welche Skala man nimmt — es können auch Zehner-, Hunderter- oder Neuner-Skalen sein, wie es Wallner, Joos und Gosemärker vorschlagen. Die quantitative Form betrifft sowohl messbare als auch nicht messbare Variablen. Letztere schätzt man einfach ein.

3. Offenheit: Das System ist offen für neue zusätzliche Daten und kann aufgrund von Erfordernissen in der praktischen Anwendung flexibel angepasst werden.

4. Integrierbarkeit: Das Programm ist gut mit anderen Methoden integrierbar – man kann aufgrund der quantitativen Darstellung beliebige Dimensionen und Skalen verschiedener Tests einfach modellieren. Auch auf der technischen Ebene kann man es schnell mit verschiedenen Softwaresystemen und Datenbanken integrieren.

5. Das Programm beinhaltet eine Datenbank, in der alle eingegebenen Daten und Resultate gespeichert werden. Damit ist man in der Lage, verschiedene statistische Auswertungen durchzuführen.

6. Klare Trennung von Handschriftauswertung und Interpretation wird systemseitig sichergestellt: Ein Graphologe untersucht eine Schriftprobe, schätzt alle Handschriftmerkmale ein und trägt sie in ein Merkmalsprotokoll ein. In unserer Untersuchung war das Merkmalsprotokoll ein eigens für den Zweck erstelltes Excel-Dokument. HSDetect importiert die Daten des ausgefüllten Protokolls und berechnet die Persönlichkeitseigenschaften bzw. ihren Ausprägungsgrad. Alle diese Daten sind in der Datenbank gespeichert. Das Programm interpretiert die Resultate nicht, d.h. es generiert kein Gutachten, da es für eine Untersuchung nicht nötig ist.

Die wichtigsten Konstrukte von HSDetect sind natürlich die Handschriftmerkmale (zurzeit ca. 800 Merkmale und ca. 200 Merkmalscluster, d.h. Merkmalskombinationen), die Persönlichkeitseigenschaften (ungefähr 400 Eigenschaften und Verhaltensmuster, wobei jedes noch viele Synonyme hat, die sich auch in der Datenbank befinden) und die Verknüpfungen zwischen den Merkmalen und Eigenschaften (ca. 5000).

Um eine formalisierte Auswertung der Persönlichkeitseigenschaften durchführen zu können, muss man alle Konstrukte quantitativ darstellen. Wie bereits weiter oben ausgeführt, erhalten deshalb sowohl die erfassten Handschriftenmerkmale als auch die ausgewerteten Persönlichkeitseigenschaften in HSDetect Werte von 0 bis 1.

Die Auswertung der Persönlichkeitseigenschaften erfolgt durch folgendes Modell: Der Wert einer Eigenschaft w ist eine Kombination (Cobb-Douglass-Funktion) des Ausprägungsgrades y und des Reliabilitätswertes r:

$$w = y^{\alpha} \cdot r^{1-\alpha}$$

Unserer Erfahrung nach beträgt der optimale Wert für den Koeffizienten α 0,6.

Den Ausprägungsgrad y kalkuliert das Programm durch eine S-Funktion der nachfolgenden linearen Summe:

12 Chernov, Y. (2012). Computergestützte Graphologie: Ansätze, Modelle, Möglichkeiten und Grenzen. GraphologieNews, http://graphologie-news.net/cms/upload/archiv/computergestuetztegraphologie

$u = a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_i \cdot x_i + a_k \cdot x_k$

$y = 1/(1+ e^{-u})$

X_i ist der durch den Graphologen eingeschätzte Wert des Handschriftenmerkmals i und k die Anzahl der Merkmale, die mit der untersuchten Eigenschaft verknüpft sind. Natürlich kommt dasselbe Merkmal in mehreren Eigenschaften vor. Der Koeffizient a_i steht für das Gewicht des Merkmals i in der untersuchten Eigenschaft. Die Koeffizienten wurden aufgrund der Daten aus mehreren Quellen in der Datenbank zuvor berechnet.

Die S-Funktion oder logistische Funktion ist ein sehr häufig verwendetes Modell in der Psychologie und Ökonomie. Durch ihren Einsatz können Tendenzen besser modelliert werden, wenn die Werte in der Mitte zwischen einem minimalen und maximalen Wert weniger aussagekräftig sind. In unserem Fall steigt die Funktion stärker je mehr u gegen 1 geht, und sinkt dagegen stärker je mehr u gegen 0 läuft.

Der Wert von r hängt von der Anzahl der involvierten Merkmale ab: Je mehr Merkmale einen Wert haben, der nicht Null ist, desto höher ist r.

Das Verwenden einer Software anstelle einer manuellen Handschriftanalyse sichert die Objektivität, Reliabilität und Wiederholbarkeit des Experimentes.

Soweit zu den graphologischen Elementen, die Bestandteil unserer Validierungsuntersuchung sind. Als nächstes gehen wir über zur Beschreibung des Elements, an dem wir die graphologische Methode validiert haben, dem 16-Persönlichkeitsfaktoren-Test in der revidierten deutschen Fassung, die von Schneewind und Graf[13] erstellt wurde.

Der von R.B. Cattell in den 1940er-Jahren entwickelte Test ist ein seit weit verbreitetes Instrument zur Erfassung der Persönlichkeitsstruktur. Über mehrere Jahrzehnte wurde der 16-PF-Test entwickelt, revidiert und aktualisiert. Der Test ist differenzierter und ausgereifter als andere gängige Persönlichkeitsteste, was sich u.a. darin zeigt, dass er 16 unterschiedliche Persönlichkeitseigenschaften erfasst, die semantisch und empirisch auf 16 Skalen mit 2 Polen reduziert wurden. Wir haben uns aber auch für diesen Test entschieden, da er bereits von vielen Graphologie-Kritikern in der Vergangenheit für Validierungszwecke eingesetzt wurde. Nachfolgend wird der 16-PF-R erläutert, damit klarer wird, wie der Test abläuft und welche Persönlichkeitsmerkmale er einschätzt.

Die Testperson erhält einen Fragebogen mit 184 Fragen. In nachfolgender Abbildung sind vier Beispielfragen mit Antwortmöglichkeiten, die bei jeder Frage identisch sind, abgebildet:

9	Meine Freunde finden, dass ich etwas zerstreut und nicht immer praktisch bin.	a. stimmt b. ? c. stimmt nicht
10	Viele gehen über Leichen, um selbst voranzukommen.	a. stimmt b. ? c. stimmt nicht
11	Ich gerate in Schwierigkeiten, weil ich manchmal meine eigenen Ideen verfolge, ohne sie mit den beteiligten Menschen zu besprechen.	a. stimmt b. ? c. stimmt nicht
12	Es fällt mir leicht, über mein Leben zu reden, auch über Dinge, die andere für ziemlich persönlich halten könnten.	a. stimmt b. ? c. stimmt nicht
13	Meine Gedanken sind zu tiefgründig und kompliziert, als dass viele sie verstehen könnten.	a. so gut wie nie b. ? c. oft
14	Ich ziehe es vor ...	a. meine Probleme mit meinen Freunden zu besprechen b. ? c. meine Probleme für mich zu behalten
15	Ich neige dazu, zu empfindsam zu sein und mir zu viele Sorgen über etwas zu machen, was ich getan habe.	a. sehr selten b. ? c. oft
16	Wenn andere sich so verhalten, als würden sie mich nicht mögen …	a. macht es mir nichts aus b. ? c. bin ich gewöhnlich verletzt
17	Ich habe schon Dinge gesagt, die andere gekränkt haben.	a. stimmt b. ? c. stimmt nicht

13 Schneewind, K.A. und Graf, J. (1998): Der 16-Persönlichkeits-Faktoren-Test, Göttingen: Huber Verlag

Nach der Auswertung des vollständig ausgefüllten Tests erhält der Untersuchungsleiter bzw. die Testperson ein Persönlichkeitsprofil, auf dem 16 Persönlichkeitseigenschaften entsprechend der ausgewerteten Antworten mit einem jeweils hohen oder niedrigen Skalenwert dargestellt werden:

Skala	Niedriger Skalenwert	Hoher Skalenwert
A	Reserviertheit	Wärme
B	Log. Schlussfolgen (niedrig)	Log. Schlussfolgen (hoch)
C	Emotionale Instabilität	Emotionale Stabilität
E	Nachgiebigkeit	Durchsetzungskraft
F	Ernsthaftigkeit	Lebendigkeit
G	Unangepasstheit	Regelbewusstsein
H	Schüchternheit	Kontaktstärke
I	Sachlichkeit	Empfindsamkeit
L	Vertrauen	Misstrauen
M	Bodenständigkeit	Träumerisch
N	Offenheit	Verschlossenheit
O	Selbstsicherheit	Ängstlichkeit
Q1	Traditionalistisch	Offenheit für Veränderung
Q2	Soziale Orientierung	Introversion
Q3	Unorganisiertheit	Perfektionismus
Q4	Entspannung	Anspannung

Die Skala A wird von Schneewind und Graf beispielsweise wie folgt beschrieben:

> „Skala A erfasst das Ausmass an Wärme bzw. eine gewisse Reserviertheit im Kontakt mit anderen Menschen. Reservierte Personen (A-) sind vorsichtig beim Eingehen von zwischenmenschlichen Beziehungen. Sie arbeiten gerne alleine im maschinellen, intellektuellen oder künstlerischen Bereich. Beispielsweise geben solche Personen an, lieber in einem Maschinenpark zuständig zu sein, als Gespräche mit Mitarbeitern zu führen; lieber Architekten zu sein, als Beraterfunktion zu haben. Zuhause beschäftigen sie sich gerne für sich allein mit ihren Hobbies. In Situationen, die eine intensive Interaktion oder emotionale Nähe erfordern, fühlen sie sich unwohl. Warmherzige Personen (A +) dagegen interessieren sich für andere Menschen, mögen die Nähe zu anderen und ziehen dementsprechend Beschäftigungen vor, in denen sie mit Menschen zu tun haben. Beispielsweise arbeiten sie lieber in einem geschäftigen Büro als in einem ruhigen Raum. Sie mögen Menschen, die ihre Gefühle offen zeigen; von Freunden werden sie selbst als warmherzig und mitfühlend beschrieben. Personen mit extrem hohen Werten auf dieser Skala leiden, wenn sich so enge Beziehungen, wie sie von ihnen angestrebt werden, nicht realisieren lassen.
>
> Die Korrelationen mit den Vergleichsmaßen weisen auf den Zusammenhang zwischen Wärme und anderen Extraversionsmaßen hin. Warmherzige Personen fühlen sich wohl in der Gesellschaft anderer. Sie erregen gerne die Aufmerksamkeit anderer Menschen, um Beziehungen zu ihnen aufzubauen. Dabei geben sie sich gleichzeitig Mühe, diese auch aufrechtzuerhalten. Dass Personen mit hohen Werten auf dieser Skala besonders interessiert sind am Wohlergehen anderer, regen Anteil am Leben ihrer Freunde nehmen, einfühlsam, unterstützend, tröstend und hilfsbereit sind, zeigt sich in den positiven Korrelationen zu den Skalen Hilfsbereitschaft.“[14]

Eine ausführlichere Beschreibung der einzelnen Skalen kann auch dem Buch „Essentials of 16PF Assessment“[15] entnommen werden.

Nachdem ein Proband alle 184 Fragen beantwortet hat, erhält jede Skala einen STEN-Wert von 1 bis 10. Am Beispiel der Skala A bedeutet ein STEN-Wert von 1 bis 5, dass der Proband zur „Reserviertheit“ tendiert. Wenn der STEN-Wert zwischen 6 bis 10 liegt, ist bei jenem Proband der Gegenpol zur Reserviertheit, der hohe Skalenwert „Wärme“, stärker ausgeprägt. Je höher der Wert, desto stärker ist die Eigenschaft ausgeprägt und umgekehrt. Am Ende der 16-PF-R-Testdurchführung erhält somit jeder Proband ein Testergebnis in Form von STEN-Werten für alle 16 Skalen.

14 Schneewind, K.A. und Graf, J. (1998): Der 16-Persönlichkeits-Faktoren-Test, Göttingen: Huber Verlag

15 Cattell, H. und Schuerger, J. M.: Essentials of 16PF assessment, John Wiley & Sons, 2003

Ausgehend von den zuvor genannten Beschreibungen von Schneewind und Graf haben wir die Modellierung des 16-PF-R mit den aus HSDetect stammenden Eigenschaften vorgenommen. Am Beispiel der Persönlichkeitseigenschaft „Wärme“ mit dem Gegenpol „Reserviertheit“ haben wir folgende Eigenschaften aus HSDetect für die Modellierung der Skala A ausgewählt:

HSDetect-Eigenschaften für den hohen Skalenwert der Skala A: Rücksicht auf andere, Ansprechbarkeit, Freundlichkeit, Kontaktfähigkeit, fehlende Zurückhaltung.

HSDetect-Eigenschaften für den niedrigen Skalenwert der Skala A: Kaltherzigkeit, Distanz zur Umwelt, Reserviertheit, Ungeselligkeit.

Nun erfolgt eine Verknüpfung der selektierten HSDetect-Eigenschaften mit den 16-PF-R Skalen, wofür folgender Algorithmus gewählt wurde: HSDetect liefert für jeden Probanden Eigenschaftswerte, welche wir weiter oben durch den Wert „w“ definiert haben. Aufgrund dieser Werte erfolgt durch ein speziell entwickeltes Programm die Einschätzung der beiden Pole aller Skalen. Hierbei vergleicht das Programm die zwei Pole einer Skala miteinander, um eine richtige Modellierung zu erhalten. Ein Polwert P ist dabei einfach der Mittelwert der jeweils selektierten Eigenschaften:

$P = (w_1 + \ldots + w_m) / m$

„m“ ist hierbei die Anzahl der Eigenschaften, die diesen Pol repräsentieren. Da für einige Skalen die Anzahl der aus HSDetect selektierten Eigenschaften relativ hoch ist, wäre es nicht sinnvoll, davon auszugehen, dass alle Eigenschaften einer Skala mit einem Wert > 0 vorhanden sein müssen. Deswegen selektiert das Programm bei mehr als drei Eigenschaften pro Skala nicht alle, sondern nur zwei Drittel der Eigenschaften mit den höchsten Werten. Das ist ein übliches Vorgehen in Clusteranalyse.

Am Ende der Berechnung durch das Programm erhält man pro Skala zwei Werte: P1 für den niedrigen Skalenwert und P2 für den hohen Skalenwert. Wenn P1 größer ist als P2, befindet man sich auf dem niedrigen Skalenniveau, bei P2 größer als P1 selektiert das Programm den hohen Skalenwert. Der STEN-Wert wird aufgrund der normierten Differenz eingeschätzt:

$ABS\ (P1-P2) / Max\ (P_1, P_2)$

Je größer der Wert der Differenz, desto größer ist die Abweichung des STEN-Wertes von der Mitte. Auf diese Weise erhält man für jeden Probanden den zweiten Vektor des STEN-Wertes (der erste ist das Resultat des 16 PF-R-Testergebnisses).

Natürlich kann man den 16 PF-R-STEN-Wert nicht direkt mit dem modellierten HSDetect-STEN-Wert vergleichen. Der Vergleich wird in der Untersuchung wie folgt vorgenommen: In der STEN-Skala werden drei Zonen definiert:

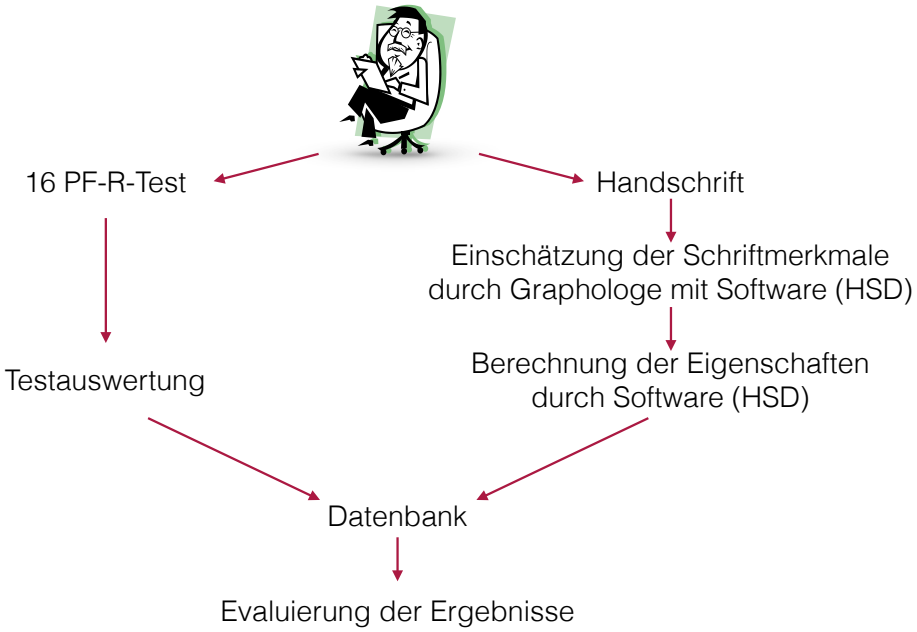
- Die Zone P1, bei der die STEN-Werte im Bereich von 1 bis 4 liegen,
- die Zone P2, bei der die STEN-Werte zwischen 7 bis 10 berücksichtigt werden und
- die P3-Zone, die für die STEN Werte 5 oder 6 steht.

Wenn die STEN-Werte des 16 PF-R- und HSDetect-Ergebnisses für einen bestimmten Probanden und eine bestimmte Skala in derselben Zone liegen, sprechen wir von einer Übereinstimmung bzw. einem positiven Resultat. Wenn sich ein Wert von einem der beiden Testergebnisse in der P1-Zone und der andere Wert dagegen in P2-Zone befindet, sprechen wir von einer eindeutigen Nicht-Übereinstimmung bzw. einem negativen Resultat. Wenn ein Wert in der P3-Zone und der Wert des zweiten Testergebnisses entweder in der P1- oder P2-Zone liegt, bezeichnen wir es als ein unbestimmtes Resultat, eine Art Grauzone. Auf diesem „Zonen-Modell“ basiert die weiter unten dargestellte Ergebnisinterpretation der Validierungsuntersuchung. Des Weiteren werden die Ergebnisse auch durch ein „Pol-Modell“ interpretiert: Man betrachtet zwei Pole und schaut, ob die STEN-Werte des 16 PF-R- und des HSDetect-Ergebnisses zu demselben Pol gehören.

Die Elemente des neuen Validierungsansatzes sind nun beschrieben, sodass wir uns im Folgenden dem Ablauf der Validierungsuntersuchung sowie der Darstellung und Besprechung der Ergebnisse widmen können.

Ablauf der Validierungsuntersuchung

Bevor wir nun zur Darstellung der Ergebnisse und den Erkenntnissen für unser Fachgebiet übergehen, möchten wir nachfolgend noch kurz den Ablauf der Untersuchung vorstellen:



54 Personen, symbolisiert durch die Figur im Sessel, haben in elektronischer oder handschriftlicher Form den 16 PF-R-Test gemacht und eine Schriftprobe abgegeben. Unabhängig voneinander wurde der 16-PF-R-Test von einem Untersuchungsleiter A ausgewertet, die Handschriftmerkmale wurden von einem Untersuchungsleiter B eingeschätzt und im Anschluss, wie zuvor im Kapitel „Elemente einer graphologischen Validierungsuntersuchung“ beschrieben, von dem Programm HSDetect ausgewertet. Mit Hilfe einer gemeinsamen Datenbank und Software wurden die errechneten Ergebnisse verglichen bzw. evaluiert.

In jeder statistischen Untersuchung muss man sich im Rahmen der Durchführung selbstverständlich die Frage nach der notwendigen Anzahl der Probanden stellen. Die von uns durchgeführte Untersuchung basiert auf der Einschätzung der Effektgrösse (Effect Size) und Teststärke (Statistical Power Analysis) gemäß Jacob Cohen[16]. Diese Einschätzung haben wir durch das an der Universität Düsseldorf entwickelte G*Power Tool[17] vorgenommen: Bei einer Binomialverteilung und unter Berücksichtigung eines in der Psychologie üblichen 5%igen Signifikanzniveaus, das besagt, dass die Nullhypothese mit einer Wahrscheinlichkeit von 5% zurückgewiesen wird, obwohl sie in Wirklichkeit wahr ist, liegt demnach die notwendige Anzahl der Probanden bei 49. Da bei vorliegender Untersuchung 54 Probanden teilgenommen haben, werden somit die Anforderungen an die statistische Relevanz erfüllt.

Ergebnisse der Validitätsuntersuchung

Die Hauptfrage dieser Validitätsuntersuchung ist, ob die Resultate der Handschriftanalyse mit dem 16 PF-R-Test übereinstimmen und somit ob die graphologischen Ergebnisse statistisch signifikant sind. Wenn die Ergebnisse des Vergleichs zufällig sind, dann müssen die Häufigkeiten der Ausgangsvarian-

16 J.Cohen (1988): Statistical Power Analysis for the Behavioral Sciences. New Jersey, Lawrence Elbraum Associates

17 http://www.gpower.hhu.de/

ten im weiter oben beschriebenen Zonen-Modell gleich sein und einer theoretischen Wahrscheinlichkeit von einem Drittel oder ungefähr 0,33 entsprechen. Damit lautet die Hypothese H_0 für jede Ausgangsvariante: Die Wahrscheinlichkeit (q) ist gleich 0,33. Die Alternativhypothese H_1 dagegen behauptet, dass die Wahrscheinlichkeit nicht 0,33 ist und damit die Verteilung der Ausgänge nicht zufällig ist, sondern einer gewissen Gesetzmäßigkeit unterliegt. Für die Überprüfung von H_0 wird die Binomialverteilung verwendet, die sich bei mehr als 10 Punkten der Normalverteilung annähert. Den Z-Wert für die standardisierte Normalverteilung berechnet man folgendermaßen:

$$Z = \text{Abs}\,(h-q) / \sqrt{(q\,(1-q) / n)}$$

h ist die experimentelle Häufigkeit (die Anzahl der Ausgänge in der Zone durch die Gesamtanzahl der Ausgänge); n – die Anzahl der Punkte (Ausgänge).

Für ein 5%iges Signifikanzniveau beträgt der Z-Wert 1,65. Von diesem Wert ausgehend kann man sehr einfach den Grenzwert für H_0 berechnen. In vorliegender Untersuchung mit 54 Probanden und 16 Skalen gibt es insgesamt 864 Punkte. Der Mittelwert beträgt 864 * 0,33 = 288. Die notwendige Anzahl der Ausgänge berechnetet man wie folgt:

$$h = Z * \sqrt{(q\,(1-q) / n)} + q$$

Unter Berücksichtigung von q = 0,33, n = 864 und Z = 1,65 erhält man für h den Wert 0,356, was einer Anzahl von 308 Ausgängen entspricht, wenn h > q ist. Der zweite Grenzwert für den Fall h < q beträgt 268. Somit liegt das Annahmeintervall für ein Signifikanzniveau von 5% zwischen 269 und 307. Wenn die Anzahl der Punkte mit einem Ausgang in diesem Intervall liegt, ist das Ergebnis zufällig und man muss die Nullhypothese (H_0) akzeptieren.

Die statistischen Ergebnisse der Untersuchung sehen vernünftig aus: Die Anzahl der Punkte pro Zone und STEN-Wert für den 16 PF-R- und HSDetect-Test sind mehr oder weniger übereinstimmend, was für eine gute Modellierung spricht.

	Zone P1				Zone P3		Zone P2			
	Sten 1	Sten 2	Sten 3	Sten 4	Sten 5	Sten 6	Sten 7	Sten 8	Sten 9	Sten 10
16PF	17	48	76	90	106	139	100	160	85	43
	231				245		388			
HSDetect	10	51	125	131	158	103	110	107	59	10
	317				261		286			

Die Ergebnisse der Überprüfung für den gesamten Test sieht man in der nachfolgenden Tabelle:

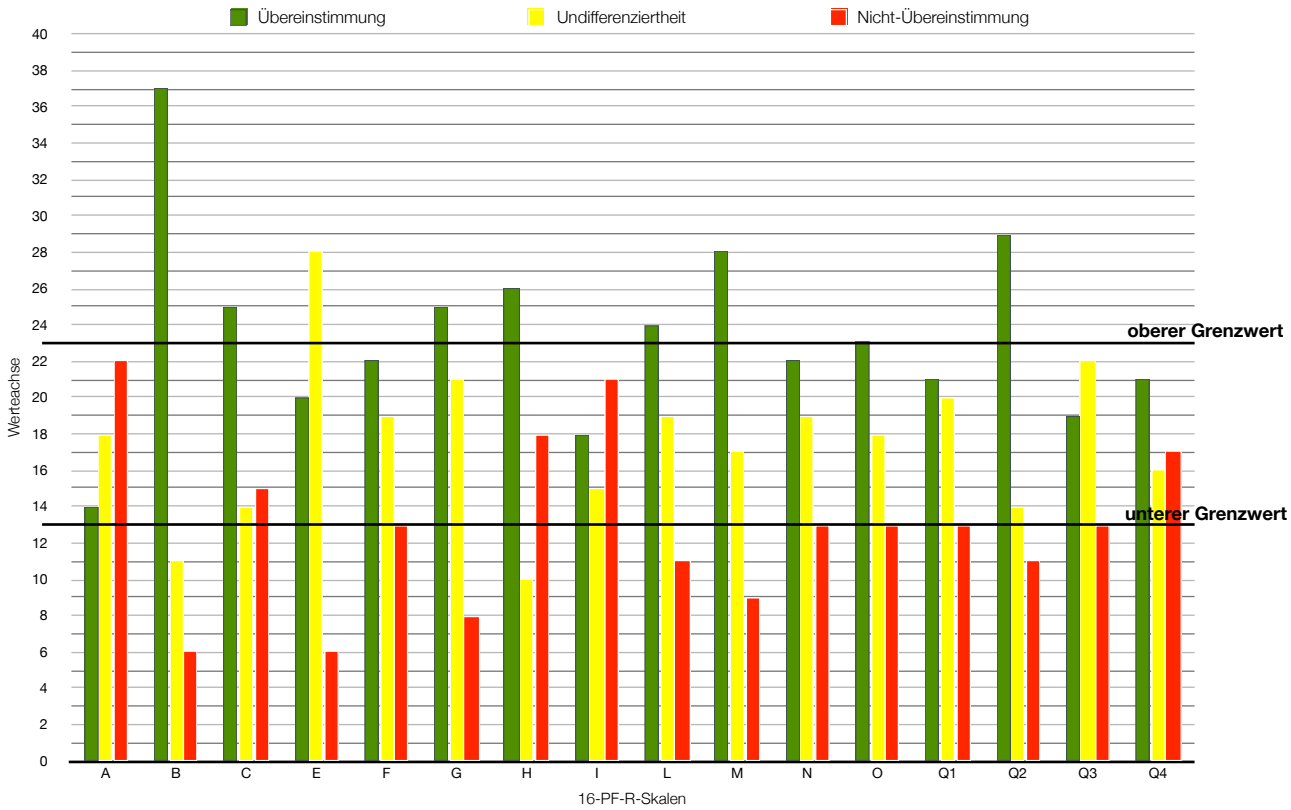
	Experimenteller Wert	Reliabilitätsintervall	H_0
Übereinstimmung	374	269 - 307	Abgelehnt
Undifferenziertheit	281	269 - 307	Angenommen
Nicht-Übereinstimmung	209	269 - 307	Abgelehnt

Man kann erkennen, dass die Ergebnisse sowohl für die Übereinstimmung als auch für die Nicht-Übereinstimmung statistisch bedeutend sind, da sie weit über bzw. unter dem jeweiligen Mittelwert liegen. Eine Übereinstimmung der Testergebnisse zwischen dem 16 PF-R- und dem HSDetect-Test ist somit nicht rein zufällig, sondern statistisch bedeutend, ebenso die niedrigen Werte der zwei alternativen Ergebnisse.

Nachfolgend findet die Ergebnisanalyse der Skalen statt. Für eine Skala mit 54 Punkten liegt das Annahmeintervall zwischen 13 und 23. Die Ergebnisstatistik sieht wie folgt aus:

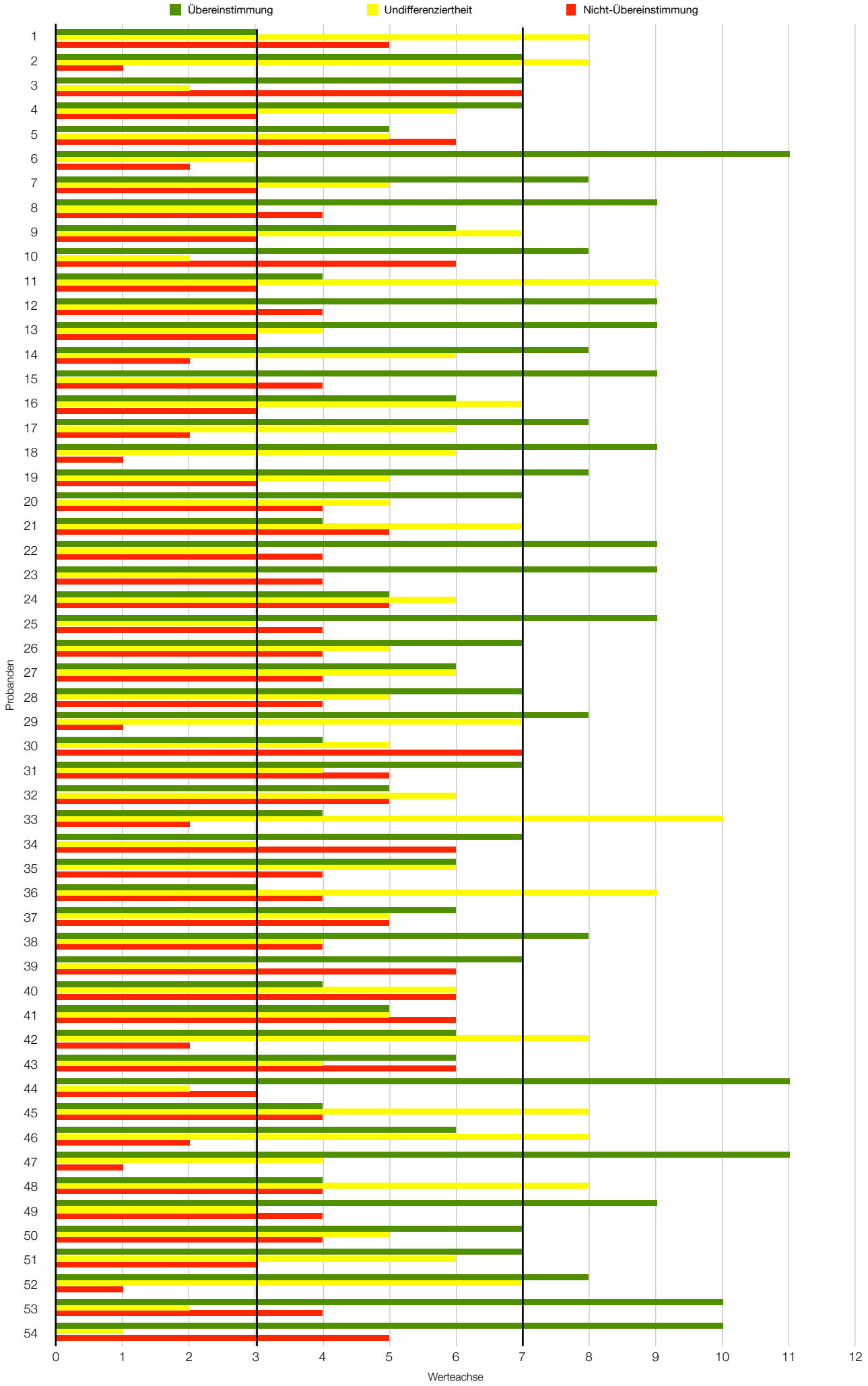
	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
Übereinstimmung	14	37	25	20	22	25	26	18	24	28	22	23	21	29	19	21
Undifferenziertheit	18	11	14	28	19	21	10	15	19	17	19	18	20	14	22	16
Nicht-Übereinstimmung	22	6	15	6	13	8	18	21	11	9	13	13	13	11	13	17

In der graphischen Darstellung sind die Ergebnisse besser nachvollziehbar:



Für keine der 16 Skalen ist die Nicht-Übereinstimmung (rot) statistisch bedeutend. Sie ist entweder statistisch unbedeutend, d.h. liegt unter dem Wert 13 (bei den Skalen B, E, G, L, M, Q2), oder liegt im Annahmeintervall der Nullhypothese. Die positiven (grünen) und unbestimmten (gelben) Ergebnisse sind dagegen für einige Skalen statistisch bedeutend, d.h. sie liegen über den Wert von 23 (bei den Skalen B, C, E, G, H, L, M, Q2). Da es sich um eine rein statistische Untersuchung handelt, ist es nicht relevant durch psychologische Interpretationen zu analysieren, warum die Ergebnisse einiger Skalen das Signifikanzniveau nicht erreichen. Man muss diese Ergebnisse im Rahmen einer statistischen Untersuchung als Black-Box-Ergebnisse akzeptieren.

Wenn man auf der folgenden Seite die Resultate des Vergleichs der Handschriftanalyse mit dem 16 PF-R-Test pro Proband betrachtet, kann man feststellen, dass es keine Personen gibt, die nur statistisch nicht-übereinstimmende Daten haben. Das für die Annahme der H_0-Hypothese relevante Intervall liegt in diesem Fall bezogen auf die 16 untersuchten Skalen des 16-PF-R zwischen 3 und 7. Dabei haben von 54 Probanden 22 einen Übereinstimmungswert von 8 und grösser, was die zuvor beschriebenen allgemeinen Ergebnisse widerspiegelt und somit folgerichtig ist.


Übereinstimmung
Undifferenziertheit
Nicht-Übereinstimmung
Probanden
Werteachse
0
1
2
3
4
5
6
7
8
9
10
11
12

Erkenntnisse für die Handschriftanalyse

Welche Schlüsse können wir aus den Resultaten dieser Untersuchung für die Graphologie als psychodiagnostische Methode ziehen?

Die Graphologie misst Persönlichkeitseigenschaften auf vergleichbare Weise wie der validierte 16-Persönlichkeitsfaktoren-Test. Mit Hilfe der eingesetzten und zuvor erläuterten Methoden konnte aufgezeigt werden, dass die Ergebnisse aus der Handschriftanalyse signifikant sind und Ergebnisse aus anderen, vorgegangenen Untersuchungen[18], die zum Ergebnis kamen, eine Korrelation zwischen Handschrift- und Persönlichkeitsmerkmalen sei rein zufälliger Natur, können widerlegt werden.

Was der Graphologie oft vorgeworfen wird, Beeinflussung durch inhaltliche Anhaltspunkte oder die Auflösung der „Grapho-Logik" in subjektive Verzerrungen, intuitives Schauen, ästhetisches Erspüren, Angemutet-Sein und das Vorhandensein eines „siebten Sinns", wird durch vorliegende Ergebnisse entkräftet. Selbstverständlich ist dieses Resultat dem eingesetzten System — der computergestützten Graphologie mittels HSDetect — geschuldet, das durch den Verzicht auf formulierte Gutachtentexte, den Einsatz einer Vielzahl von unterschiedlichen graphologisch eindeutigen Merkmals-Eigenschaft-Verknüpfungen und dem Black-Box-Prinzip in Bezug auf die Einschätzung und Ergebnisermittlung geprägt ist.

Dieser neue Ansatz zur Ermittlung der graphologischen Validität hat sich somit als wirksam und vielversprechend erwiesen und die während der Durchführung gesammelten Daten sollten zukünftig im Zusammenhang mit anderen Testen und weiteren Untersuchungen der graphologischen Gütekriterien eingesetzt und ausgebaut werden, wenngleich am eingesetzten System noch einige Punkte erweiterungs- und verbesserungsbedürftig sind.

Zusammenfassend sind also erste Schritte in die richtige Richtung erkennbar, allerdings ist noch ein weiter Weg zurückzulegen, bis die Handschriftanalyse als wissenschaftlich abgesichertes psychodiagnostisches Erkenntnismittel die Anforderungen der DIN 33430 erfüllt und sich so wieder den Eintritt als anerkannte Methode in Universitäten und Unternehmen zurück erkämpft.

18 http://www.ivha.info/publications-on-validation-of-handwriting-analysis/critical-works/