Results of Interlab Proficiency Test No. 005 in Handwriting Analysis:

Analysing dysgraphy signs in handwriting

Table of Contents

Background and task in Interlab Proficiency Test 005

Disturbances in writing (e.g. deformed letters, illegible writing, unharmonious writing) are understood as dysgraphy. A distinction is made between acquired (e.g. after accident, stroke, tumour) and development-related writing disorders.

In France, there is a test called "Approche dynamique de l'écriture" (ADE, developed by Adeline Eloy), which allows the handwriting to be classified into a graphic stage of development corresponding to its age, thus enabling the difficulties and strengths of the handwriting to be objectively analysed to diagnose dysgraphy. From the ADE scale, 8 out of 24 handwriting signs from the categories form, space, movement and stroke were selected for the following ring trial, which should be diagnosed by the participants in six handwritings. These eight signs are by no means sufficient to diagnose dysgraphy.

Summary of results of Interlab Proficiency Test 005

17 European graphologists (Belgium, Germany, England, Switzerland) show a reasonable agreement without outliers with regard to the assessment of eight dysgraphy signs in six handwritings. There is a slightly higher agreement for handwritings No. 1, 2 and 6.

At the sign level, it can be seen that there is a substantially poorer agreement among the handwriting experts for the signs "amended letters" and "irregular upper and lower zone" than for the signs "tremors", "unstable of the direction of the line" and "disorderly layout".

In addition to the raw data, the evaluation also shows the results determined by the "Free Marginal Fleiss' Kappa" calculation method. In the section "statistical approach" of this document important hints for a better understanding of the results are explained.

Organisation, operation and evaluation

Klara Leclercq Backes, Claudia Caspers and Yury Chernov collected the data between May and November 2018. The evaluation took place in spring 2019. Special thanks to all persons who actively participated in this experiment.

Participants of the Interlab Proficiency Test

17 handwriting experts of different European countries

Data material and Methodology of the Interlab Proficiency Test 005

Data Material

- 6 different European handwritings of writers of different sexes and ages
- 136 evaluations (8 per handwriting from 17 handwriting experts)

Instructions for participants

Each of the six handwritings had to be diagnosed according to the following eight parameters:

1. Clumsy round letter distortions of ovals
2. Tremor
3. Irregular upper and lower zone
4. Incorrect letters
5. Unstable of the direction of the line
6. Disorderly layout

7. Amended letters
8. Touching letters

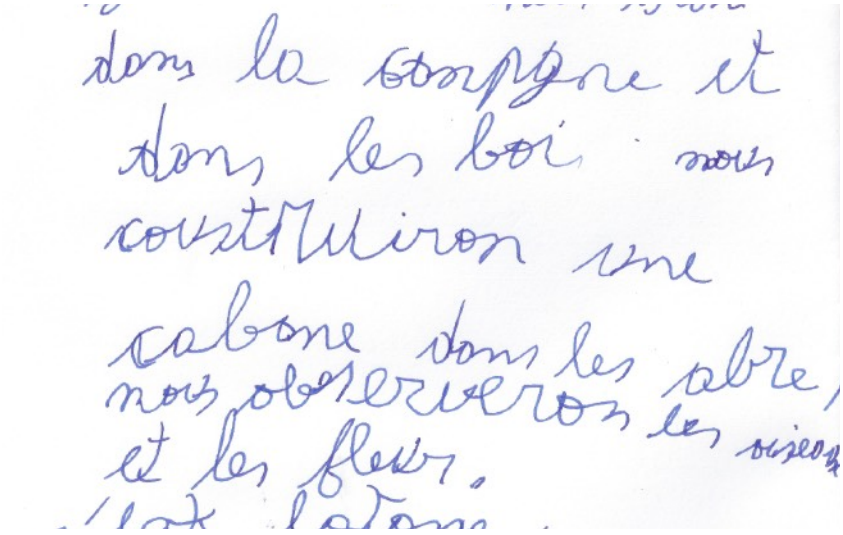
Only the existence or non-existence per parameter and handwriting was evaluated by the participants. Definitions per parameter were provided. A sample of a trilingual table per handwriting to document every results looked as follows:

	Vorhanden existent présent	Nicht-Vorhanden non existent absent
Ungeschickte runde Buchstaben | Clumsy round letter distortions of ovals | Oves maladroits		
Zitterzüge | Tremors | Tremblements		
Unregelmäßige Ober- und Unterlängen | Irregular upper and lower zone | Irrégularités de dimension des hampes/jambages		
Falsch geschriebene Buchstaben | Incorrect letters | Lettres incorrectes		
Schwankende Zeilenführung | Unstable of the direction of the line | Fluctuation de la trajectoire		
Gestörtes Raumbild | Disorderly layout | Inorganisation espace feuille		
Verbesserte Buchstaben | Amended letters | Lettres retouchées		
Aneinander stoßende Buchstaben (extreme Enge) | Touching letters | Collision de lettres		

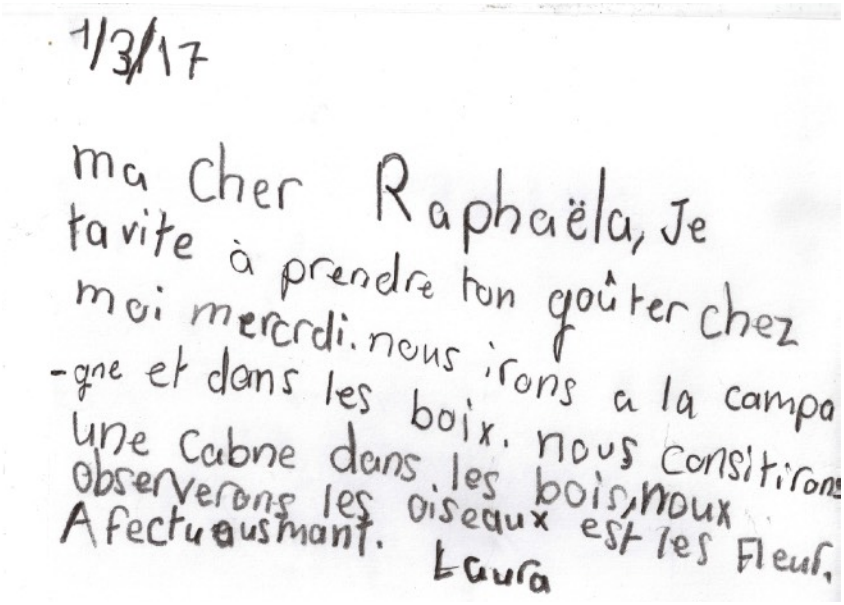
Before the statistical approach and the results of Interlab Proficiency Test 005 are presented, cuttings of the six handwritings to be judged are shown for the sake of clarity.

Cuttings of the handwritings to be judged

Cutting of handwriting 1 (Male 10.6 years old, not original size):

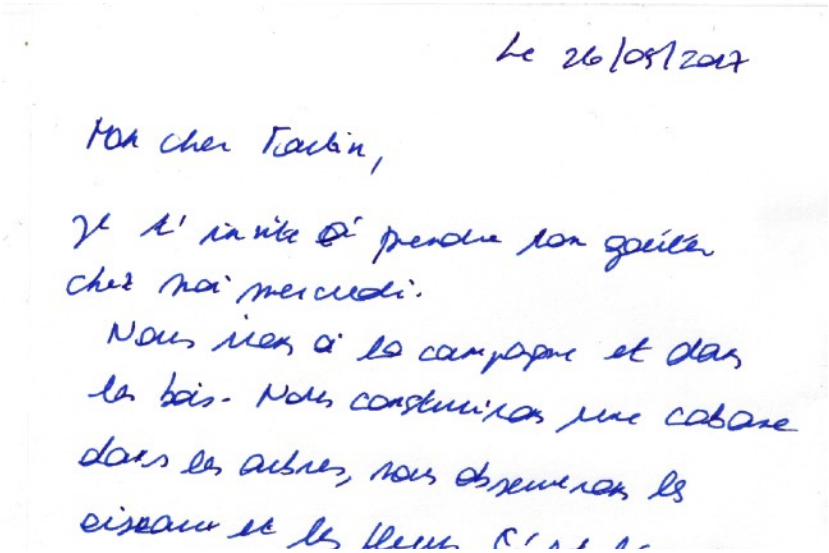


Cutting of handwriting 2 (Female, 10 years old, not original size):


1/3/17

ma Cher Raphaëla, Je
ta vite à prendre ton goûter chez
moi mercredi. nous irons a la campa
-gne et dans les boix. nous construirons
une cabne dans les bois, nous
observerons les oiseaux est les Fleur,
Afectueusmant. Laura

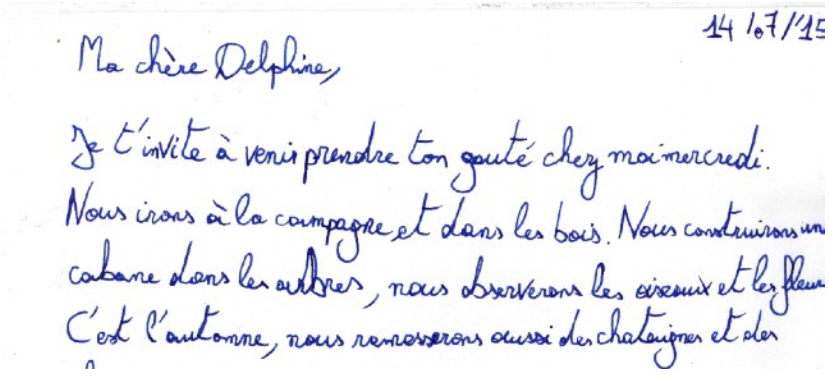
Cutting of handwriting 3 (Female, 31 years old, not original size):


Le 26/08/2017

Mon cher Fabin,

Je t'invite à prendre ton goûter
chez moi mercredi.
Nous irons à la campagne et dans
le bois. Nous construirons une cabane
dans les arbres, nous observerons les
oiseaux et les fleurs. C'est

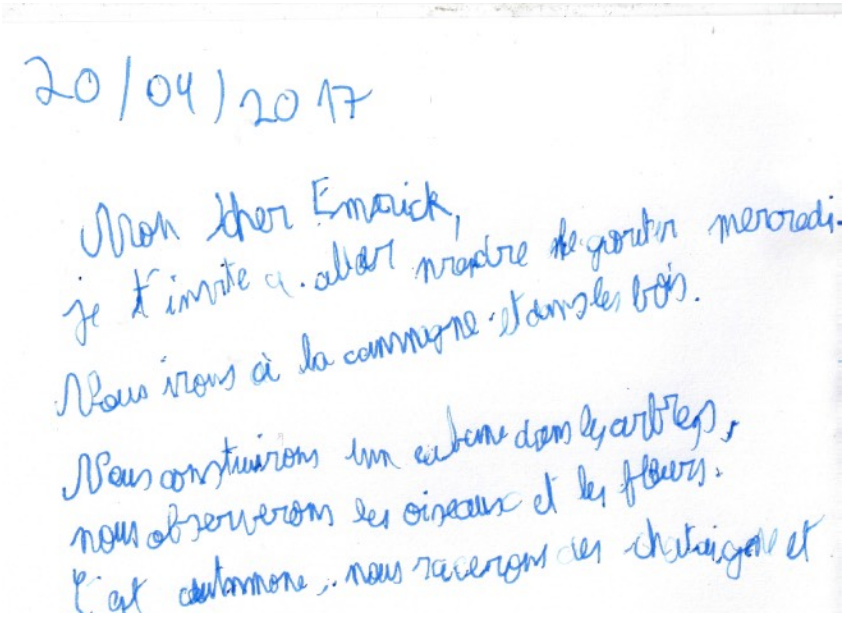
Cutting of handwriting 4 (Male, 20 years old, not original size):


14/07/15

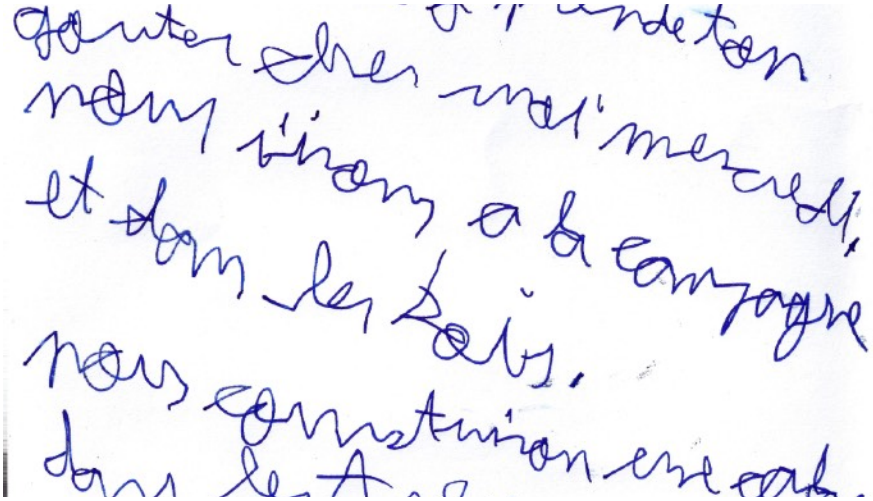
Ma chère Delphine,

Je t'invite à venir prendre ton goûté chez moi mercredi.
Nous irons à la campagne et dans les bois. Nous construirons une
cabane dans les arbres, nous observerons les oiseaux et les fleurs.
C'est l'automne, nous ramasserons aussi des chataignes et des

Cutting of handwriting 5 (Male, 12.6 years old, not original size):



Cutting of handwriting 6 (Male, 11.2 years old, not original size):



Statistical Approach

The evaluation should analyse how stable the measurements are when different evaluators are used, i.e. the agreement between several evaluators (= inter-rater reliability). In order to determine this inter-rater agreement, especially when samples are small, like in our case, the kappa measures are used in statistics. There are different kinds of kappa to measure interrater reliability, e.g. Cohen's kappa for two raters or Fleiss' kappa as an adaptation of Cohen's kappa for three and more raters. The Kappa measures the degree of agreement that can be expected above chance.

In the case of interlab proficiency test 005 there were 17 raters giving independently categorical two ratings (certain sign is existent or not in a handwriting). The most popular measure for such cases is Fleiss' kappa. The measure is simple, but has two problems:

- It is strongly influenced by prevalence and bias. This results in low values of the measure when agreements are in reality good.
- It does not work good for two-three ratings, but better when number of ratings is higher.

These problems effect the result when the structure of ratings is variable. That is, the number of specific ratings varies from one object of the experiment to another. For instance, for one handwriting (or one parameter) the specific rating strongly dominates (more raters select one rating), for another handwriting or parameter both ratings are assigned approximately equally. This situation takes place in our case.

To compensate this effect the slightly modified variant, that is, “Free Marginal Fleiss’ Kappa” according to Brennan, Prediger and Randolph was used (Randolph, 2005).

Typically, the kappa values can be interpreted as follows (Altman, 1990):

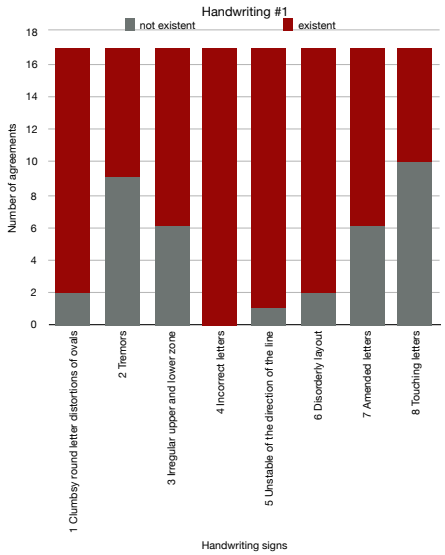
Fleiss’ Kappa values	Strength of agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

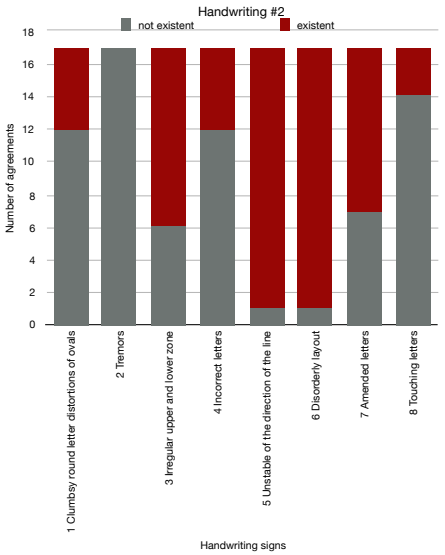
The formal results are lower then could be expected, when we just logically analyse the sample data. In such cases, the values of kappa should not be considered absolutely, but rather as a measure for comparison one handwriting to another, or one parameter to another. Therefore, we can conclude that, for instance, handwriting 6 is twice better than handwriting 4. However, in both the agreement is good. Another example, the agreement for parameter 5 “Unstable of the Direction of the line” is much better than for parameter 3 “Irregular upper and lower zone”.

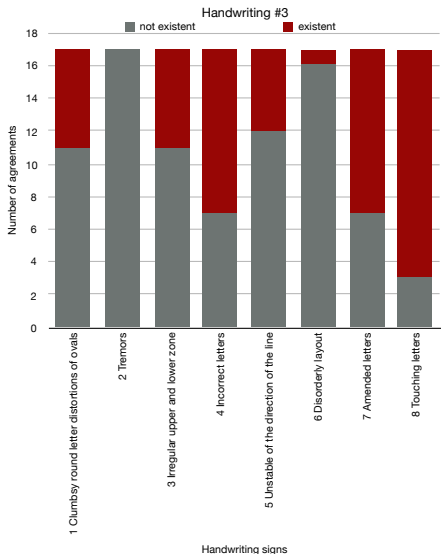
Next to the Fleiss’ Kappa results also the raw data are presented in form of diagrams.

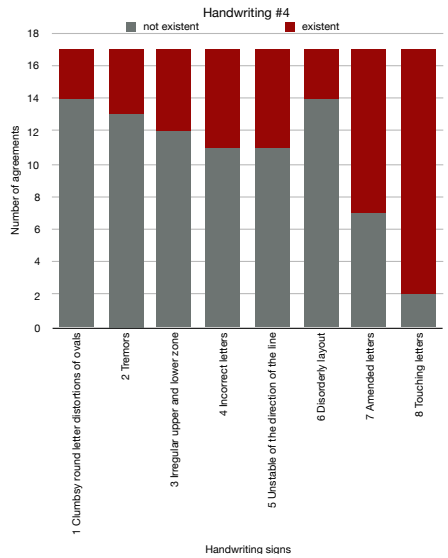
Results of agreement per handwriting (6)

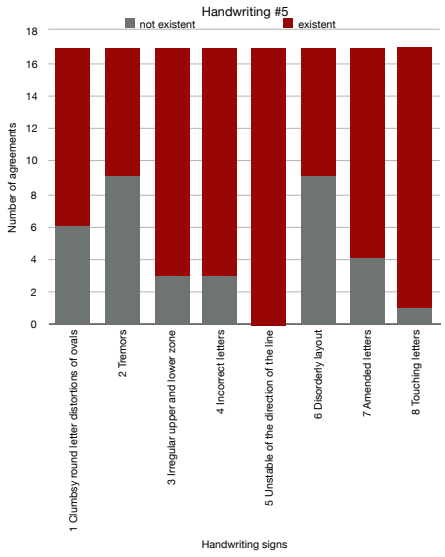
Diagram of raw values of each handwriting evaluation











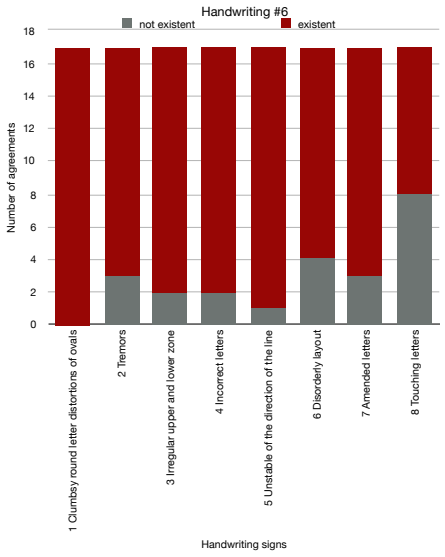
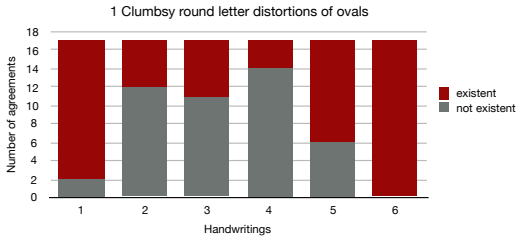


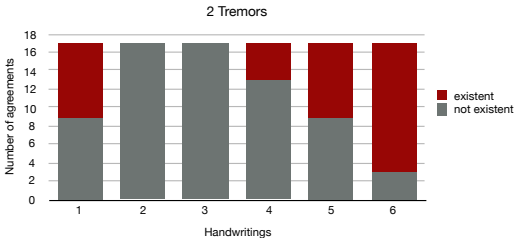
Table of Fleiss' Kappa values per handwriting

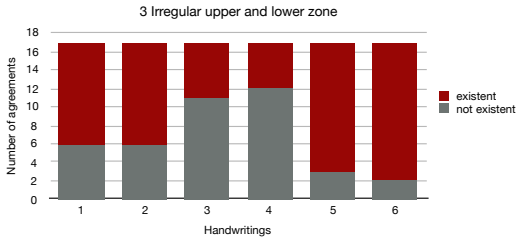
	% Overall agreement	Fleiss' Kappa values	Strength of agreement
Handwriting 1	67.83 %	0.36	Fair
Handwriting 2	69.67 %	0.39	Fair
Handwriting 3	64.15 %	0.28	Fair
Handwriting 4	60.66 %	0.21	Fair
Handwriting 5	66.73 %	0.33	Fair
Handwriting 6	73.90 %	0.48	Moderate

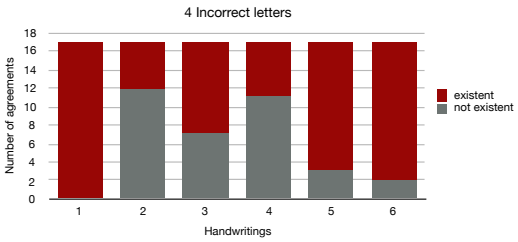
Results of agreement per handwriting sign (8)

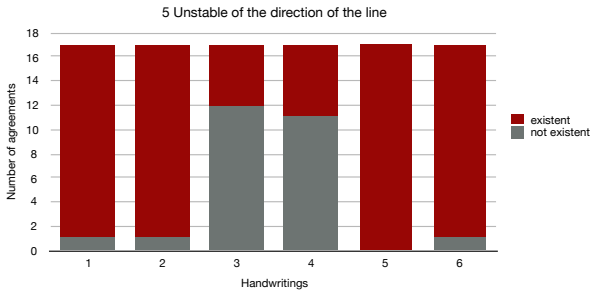
Diagram of raw values of each handwriting sign evaluation

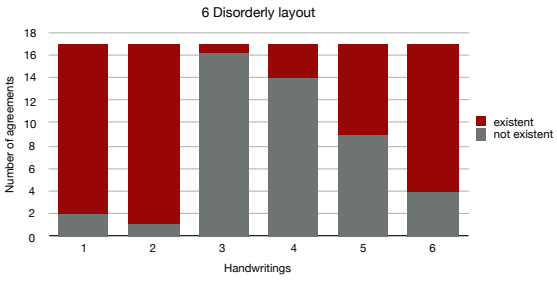


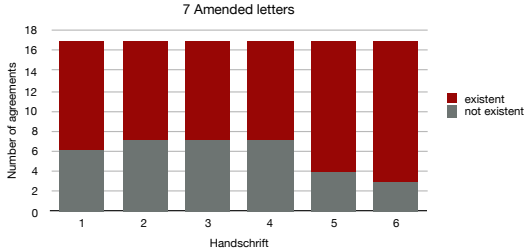












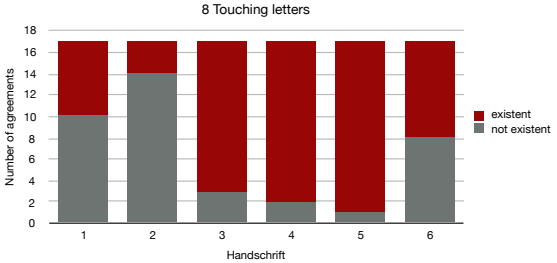


Table of Fleiss' Kappa values per handwriting sign

	% Overall agreement	Fleiss' Kappa values	Strength of agreement
1 Clumsy round letter distortions of ovals	67.65 %	0.35	Fair
2 Tremors	70.83 %	0.42	Moderate
3 Irregular upper and lower zone	59.56 %	0.19	Poor
4 Incorrect letters	67.16 %	0.34	Fair
5 Unstable of the direction of the line	78.68 %	0.57	Moderate
6 Disorderly layout	72.06 %	0.44	Moderate
7 Amended letters	54.66 %	0.09	Poor
8 Touching letters	66.67 %	0.33	Fair

References

Altman, D. G. (1990). Practical Statistics for Medical Research. London: Chapman and Hall.

MacArthur, C. A., Graham, S., & Fitzgerald, J. (Hrsg.). (2017). Handbook of writing research. (Sample Chapter: https://www.guilford.com/excerpts/macarthur.pdf?t)

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia Medica, 22(3), 276–282. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23092060

Pleau, J., & Lavoie, N. (2016). Crayon ou clavier? Effets de l'outil d'écriture sur les performances graphomotrices et rédactionnelles d'élèves de sixième année. Revue de recherches en littératie médiatique multimodale, 3. https://doi.org/10.7202/1047130ar

Publications of the department "Motology" of the University of Marburg: https://www.uni-marburg.de/fb21/motologie/mitarbeiter_seiten/mv/veroeffmv

Randolph, Justus J. (2005). 'Free-Marginal Multirater Kappa (Multirater K[Free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. Retrieved 15 April 2019, from https://www.researchgate.net/publication/224890485_Free-Marginal_Multirater_Kappa_multirater_kfree_An_Alternative_to_Fleiss_Fixed-Marginal_Multirater_Kappa

Sägesser Wyss, J. (2017). Diagnostik von Grafomotorik. Retrieved 15 April 2019, from https://www.testzentrale.de/themen/schule/artikel/Diagnostik%20von%20Grafomotorik-315

Stache, S. (2016): Viele Schüler haben Probleme mit der Handschrift. Zeit Online. Retrieved 15 April 2019, from https://www.zeit.de/gesellschaft/schule/2016-02/didacta-schueler-schreiben-motorik-schwierigkeiten