Evaluating Handwriting Skills through Human-Machine Interaction: A New Digitalized System for Parameters Extraction*

Cecilia Provenzale[+], Laura Sparaci[+], Valentina Fantasia, Chiara Bonsignori, Domenico Formica, *Member, IEEE,* Fabrizio Taffoni, *Member, IEEE*

***Abstract*— Handwriting is an important component of academic curricula and grapho-motor skills (GMS) support learning, reading, memory and self-confidence. Teachers and clinicians report increase in children experiencing problems with acquiring fluid and legible handwriting. To date gold-standard tests evaluating children's GMS, mostly rely on pen and paper tests, requiring extensive coding time and subject to high inter-rater variability. This work presents preliminary data on a new digital platform for Grapho-motor Handwriting Evaluation & Exercise (GHEE), attempting to overcome limitations of available digitalized methods for GMS evalution. In fact, contrary to previous systems, GHEE design originated from comparisons among multiple standardized tests and was based on a human-machine interaction approach. GHEE hardware and software is presented as well as data on preliminary testing. Cursive handwriting data from six adult volunteers was analyzed according to six parameters of relevance, both automatically (i.e., using GHEE software) and manually (i.e., by a human coder). Comparisons among machine and human data sets allowed parsing out parameters to be extracted automatically and parameters requiring human-machine interaction. Results confirmed platform efficacy and feasibility of the proposed approach.**

I. Introduction

Spelling accuracy, memory, letter recognition, and note-taking are essential components of academic curricula which have been proven to benefit from handwriting, as compared to typing [1-5]. Cursive handwriting has also been shown to promote grapho-motor skills (GMS) and activation of complex neural networks beneficial for learning [6]. These data emerged not only from studies relying on ink and pen, but also using digital pens [7]. However, government data show consistent growth in the number of children exhibiting poor GMS resulting in dysgraphia [8], while teacher surveys highlight that many children experience problems with acquiring fluid and legible handwriting [9].

Current assessments of GMS in elementary school children are mostly based on standardized pen and paper tests. These gold-standard tools rely on analytical assessments of multiple parameters (e.g., writing velocity, letter size, etc.) to provide an evaluation of legibility and/or presence of dysgraphia [10]. In Italy the three most commonly used tests are: the Italian *Batteria per la Valutazione della Scrittura e della Competenza Ortografica* (BVSCO) [11], the Italian standardized version of the *Concise Assessment Scale for Children's Handwriting* (*Brave Handwriting Kinder*) (BHK) [12] and the Italian *Test per la Valutazione delle Difficolta' Grafo-Motorie e Posturali della Scrittura* (DGM-P) [13].

Human coding of pen and paper tests is an extremely time-consuming process and limitations are to be found also in coders' perceptual capabilities, subjective judgments and experience. These tests also require extensive training as they are subject to high inter-rater variability, leading to high-societal costs [10-14]. Therefore, use of these assessments is limited to children with evident GMS difficulties and/or contexts in which sufficient resources are available, while children experiencing minor difficulties, either remain undetected or are detected later than appropriate [14]. Furthermore, by providing an overall score, these tests hardly aid clinicians and educators in designing individually tailored intervention programs.

In recent years, to overcome limitations of pen and paper tests and/or provide more robust understanding of the hidden complexities of handwriting, several research studies have focused on: (1) translating individual pen and paper tests into automated coding systems [15-17] or (2) developing new methods for automated parameter extraction based on digitization and signal processing techniques [14, 18-21]. Both approaches incur in some limitations: (1) direct translations of individual tests, lead to replications of the original tests' limits (e.g., if a test lacks to measure one parameter which is relevant in other tests, its automated version will do the same); (2) fully automated parameter extraction by digital systems reduces coding time at the expense of limiting educators' and clinicians' access to qualitative data which is informative *per se* (e.g., specific parameters may be more quickly assessed by a digital system rather than a human having to rely on graph paper or assessment matrixes, but if the former only returns an overall error without specifying where/how this has been made the system is not informative for clinitians or educators who need to build tailored exercises).

An alternative approach is to: (A) start by comparing different tests used in GMS assessment to pinpoint parameters

*Research supported by y the H2020/ICT European project ("CONnected through roBOTS: Physically Coupling Humans to Boost Handwriting and Music Learning" (CONBOTS) (grant agree- ment no. 871803; call topic ICT-09-2019-2020).

[+]C.P. and L.S. equally contributed to the work

C. P. and F. T. are with the Advanced Technologies and Human Centred Technologies – CREO Lab, Università Campus Bio-Medico di Roma, 00128 Rome, Italy; (e-mail: c.provenzale@unicampus.it; f.taffoni@unicampus.it).

L. S. is with the Institute of Cognitive Sciences and Technologies (ISTC), National Research Council of Italy (CNR), 00185 Rome, Italy (email: laura.sparaci@istc.cnr.it).

V.F. is with the Department of Philosophy and Cognitive Science, Faculties of the Humanities and Theology, Lund University, Sweden (email valentina.fantasia@lucs.lu.se)

C. B. and D. F. are with the Research Unit of Neurophysiology and Neuroengineering of Human-Technology Interaction (NeXTlab), Università Campus Bio-Medico di Roma, 00128 Rome, Italy (email: bonsignori.ch@gmail.com; d.formica@unicampus.it).

of relevance; (B) implement these parameters in state-of-the-art hardware and software and assess which ones may be evaluated automatically and which require human-machine interaction to avoid losing relevant information.

Point A was the objective of previous studies in which we conducted structured comparisons between the BVSCO, BHK, and DGM-P tests. Outcomes allowed identifying six parameters of relevance actively contributing towards GMS assessment in elementary school children between 7 and 9 years of age who had already acquired cursive handwriting. These parameters were: Velocity, Fluctuations, Space, Dimensions, Margins and Connections [22-24]. Point B is the aim of the present work. In particular, we will present the main hardware and software of the newly designed *Grapho-motor Handwriting Evaluation & Exercise* (GHEE) platform, as well as data from preliminary testing.

Six participants were asked to write a short phrase using GHEE hardware, parameters of relevance were subsequently extracted through automatic coding (i.e., data set 1: outcome from machine processing using GHEE software) or through human coding (i.e., data set 2: outcome from human coding). Comparisons among data sets 1 and 2 were then conducted to evaluate efficacy of the software and parameters allowing for automatic assessment vs. semi-automatic human-machine interaction. Our aim was not to automatize parameter extraction processes as in [14-21] by fully *replacing* human coding, but to rely on state-of-the art technology to *support* handwriting evaluation and learning, closely following suggestions and requests made by educators and clinitians.

GHEE was specifically designed for child use, but validation was carried out on adults given that assessment of code accuracy in automatic vs. semi-automatic form does not depend on age. This also allowed for more timely data acquisition during the current pandemic.

II. Material and Methods

A . Participants

Six adult volunteers (4 females) were enrolled to this study. Participants' mean chronological age was 27.11 years ($SD = 1.38; \text{range } 25 - 29 \text{ years}$). One participant was left-handed and one wore glasses. Experimental study was conducted according to Declaration of Helsinki guidelines and approved by the Ethics Committee of the Università Campus Bio-Medico di Roma (protocol code PAR 25.21, Rome 31 March 2021). Participants signed written informed consent describing the purpose of the study and were excluded from the study if they declared to have had any GMS difficulties.

B. GHEE hardware & software procedure

GHEE is composed of both hardware and software components (Figure 1). Hardware enables handwriting acquisition while software components allow stimuli presentation and automatic extraction of parameters of relevance (see section C below). Among possible digital tools available on the market, we selected the Wacom Cintiq 16 interactive display and its stylus, the Pro Pen 2. This device was chosen in consideration of: dimension of its active area (344x194 mm) which is comparable with the standard paper

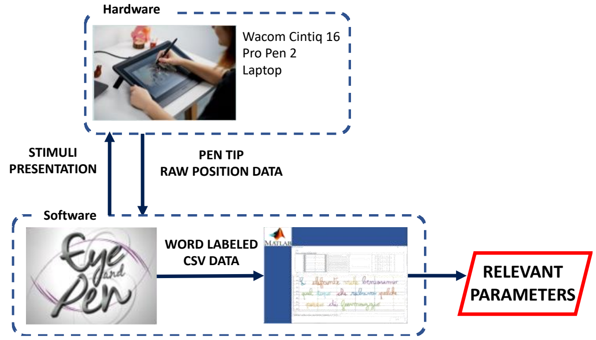

Figure 1: Assessment module of the GHEE platform: components and architecture.

size used in most school contexts in Italy (297x210 mm); for its capacity to provide visual feedback directly, where the Pro Pen 2 contacts the display, and for portability. The software part is composed of the Eye and Pen Software [25] and of a custom App developed in MATLAB. The Eye and Pen software was selected to develop stimuli presentation. Information about stylus tip position (raw position data) are split into the words composing the sentence, labelled accordingly and saved in a CSV file (csv data word labelled). This file is imported in MATLAB for feature extraction.

Participants were asked to sit at a desk with the display placed on the table in front of them and vertically oriented. The stylus is aligned with the vertical margin of the display and placed either on the right- or left-hand side in accordance with hand preference. Before each trial participants were asked to choose the paper format most familiar to them (i.e., ruled paper or grid paper). During test phase they were then presented with a standard Italian phrase on the topmost part of the screen and asked to copy it as best as they could in neat cursive handwriting on the a ruled or grid sheet that appeared immediately below the phrase. Chosen phrase was derived from [13] and contains all letters present in the Italian alphabet at least once (i.e., "*L'elefante vide benissimo quel topo che rubava qualche pezzo di formaggio*").

C. Coding

Parameters of relevance for GMS evaluation were implemented in MATLAB R2020b App Designer as part of the assessment module of the GHEE platform. **Velocity** (*aka* speed in [13]) was expressed as the ratio between the recognized letters and overall trial time. **Fluctuations** (*aka* amplitude fluctuation in [13]) was computed as the sum of the maximum vertical distance of letters above and below the line. **Space** (*aka* insufficient space between words in [12]) was measured as the horizontal distance between the last point of the last letter of the previous word and the first point of the first letter of the following one. **Dimensions** included three separate indices: maximum variation in the height of medium letters ($HV_{mL,}$ i.e., a, c, e, i, m, n, o, r, s, u, v, z,) and maximum variation in the height of ascending/descending letters ($HV_{adL,}$ b, d, f, g, h, l, p, q, t), defined as the vertical span among the corresponding letters and derived from [13], and overall heigh (OH) computed as the mean value of the larger and smaller letter of the medium class and derived from [12]. **Margins** (*aka* non-aligned left margin in [12]) was estimated as the angular coefficient of the straight line passing through the first point of the first letter of each line and compared with measures reported in [12]. **Connections**, were measured in terms of two parameters derived from [12]: broken links, and collision between two contiguous letters. The code computes the difference between the abscissa of the last point of the first letter and the first point of the second one. If such difference is lower than -0.5 mm a lost connection between letters is identified. If it is higher than 0.5 mm, the software reports an overlapping problem. The 0.5 mm threshold has been defined according to the thickness of the digital trace. For each parameter the software reported a specific value which was used for data comparisons (data set 1). Sentences produced by each participants were printed taking care not alter any aspects (e.g., dimensions) and coded by an experienced human coder (third author). Printouts were evaluated extracting the same parameters detailed above following procedures described in the corresponding standardized test from which they were derived and using graph paper and/or assessment matrixes (data set 2).

Raw data and parameters computed by the software are organized with modular structure in MATLAB. A main struct called "*Task*" is composed of two different sub-structures: "*Paper*", collecting data about the paper format (i.e., rows, columns, and margins); "*Sentence*", collecting raw position data and the performance parameters automatically extracted by the code. Raw sentence data are split in sub-structs, one for each word, according with their label. Word raw data are further split into the individual letters by asking the coder to manually select the beginning and the end of each one. Raw data for each letter are saved on separated tables for the extraction of relevant parameters reported in the previous section. Relevant parameters related to letters and sentence are saved on letters and sentence analysis tables respectively.

D. *Data analysis*

Reliability of automatic parameters extraction was tested by comparing data set 1 and data set 2. In particular, we assessed five out of the six parameters identified. Velocity was not included for analysis because it proved to be a simple ratio between number of characters and trial time and it was equivalent in both data sets. Our aim was to assess the level of agreement between the two modalities of analysis and to identify those parameters benefiting from automatic vs. human-machine interaction. We defined the error due to the software as the difference between the automatic feature extraction and the measures derived from the manual coding (considered as reference, Figure 2). We calculated the absolute value of the error for each variable and averaging it for all the participants to obtain a Mean Absolute Error (MAE). We considered 'poor agreement' all those cases in which MAE was higher than the resolution of the pen and paper test. The opportunity to exploit a collaborative approach between human and machine was critically discussed for each relevant parameter taking into account the outcome of the software and any eventual limitation in the access of qualitative data potentially informative for educators and clinicians.

III. EXPERIMENTAL RESULTS

MAE of Dimensions and Fluctuations is reported in Figure 3. The bars represent the MAE, while the whiskers the standard deviation of the absolute error. The MAE of Margins is not reported because it perfectly agrees with manual coding (i.e. $MAE = 0; SD = 0$). In manual coding, these parameters are estimated with graph paper or assessment matrixes with 1 mm spatial resolution.

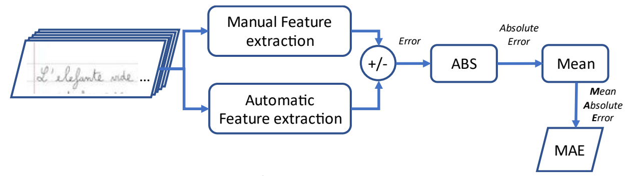

Figure 2: Data analysis of the pilot study on instrumented tools for handwriting.

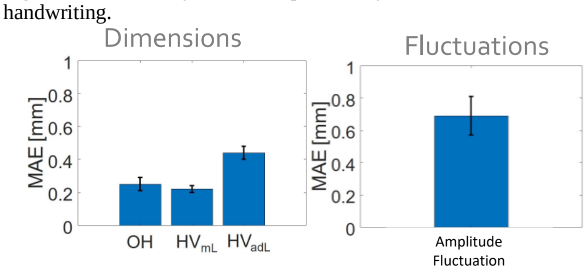

Figure 3: Mean Absolute Error (bar) and Standard Deviation (whiskers) of Dimensions and Fluctuation

As can be observed, the MAE is well below the 1 mm threshold for all parameters considered, suggesting perfect agreement between automatic feature extraction and human coder. Insufficient spaces between words extracted from the software have been compared with the ones manually assessed by the human coder. The error in the assessment represents the number of insufficient spaces missed or wrongly measured from the software, considering the ones extracted manually as reference.

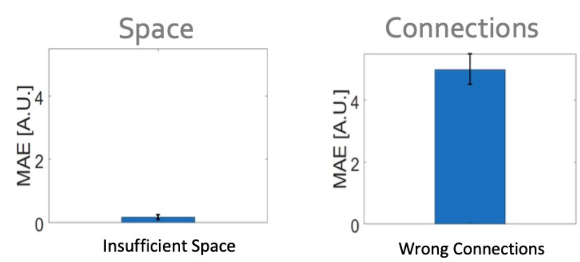

Figure 4: Insufficient space between words and Wrong connections

Figure 4 reports the MAE of this parameter on the left: it is lower than 1 occurrence, meaning that, on average, the two assessment modalities are equivalent. Finally, we compared the number of wrong connections in the same way (Figure 4 on the right). The estimation error on this parameter was higher than expected, suggesting poor agreement between the automatic and manual extraction.

IV. DISCUSSION AND CONCLUSION

In this work we present the hardware and software structure of the new GHEE platform and report data from preliminary testing. Comparisons among data sets allowed to evaluate efficacy of the software and highlighted that GHEE can reliably assess Fluctuation, Space and Margins automatically, significantly reducing GMS coding time. Moreover, GHEE software proved to correctly assess Dimensions also distinguishing specific letter types (i.e., ascending/descending vs. medium), allowing for qualitative information relevant towards tailored exercise planning. Critically, the manual coding of these parameters is an extremely time-consuming process for human coders requiring use of graph paper or assessment matrixes. Our results suggest that for all these parameters, automatic processing approach is strongly recommended and may even outperform the human coder, by cutting down time and costs, provided that other qualitative data are not lost by the automatization. As for Connections a relevant difference in MAE between data sets emerged. This suggests that this parameter requires a semi-automatic human-machine interaction approach for to qualitative aspects be appropriately captured. Therefore, in future testing we will modify the code so that it performs a pre-screening of Connections, also testing different methods to estimate distance among contiguous letters (e.g., considering not only the abscissa of letters' marginal points, but identifying the point of the second letter closest to the last point of the previous one). Subsequently, the human coder will be asked to check automatic assessment outcome: this will reduce human workload (as the coder will be working on pre-assessed data), provide quantitative information about the number of wrong connections, but avoid missing out on relevant qualitative information on where and what type of errors (e.g., missing connection vs. overlapping) are being made by writers. In conclusion, preliminary results strongly encourage use of a mixed approach, in which machine processing and human-machine interaction coexist to reduce coding time without losing relevant qualitative information needed for designing individually tailored exercise programs. Initial results are in line with the main aim of GHEE, which is to evaluate how new technologies may *support* rather than *replace* handwriting evaluation and learning. Output from the GHEE platform may support future studies on handwriting quality also relying on control models such as the Beta elliptic model [26] or the log normal one [27]. We are currently planning to evolve the GHEE platform towards a system which proposes individually tailored cursive handwriting exercises based on the evaluation of parameters of relevance and test this on a large children sample.

References

[1] A. E. Cunningham e K. E. Stanovich, "Early spelling acquisition: Writing beats the computer," J. Educ. Psychol. vol. 82(1), pp. 159–162, 1990.

[2] T. J. Smoker, C. E. Murphy, and A. K. Rockwell, "Comparing Memory for Handwriting versus Typing," Proc. Hum. Factors Ergon. Soc., vol. 53, n. 22, pp. 1744–1747, Oct. 2009.

[3] M. Longcamp, M.-T. Zerbato-Poudou, and J.-L. Velay, "The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing", Acta Psychol., vol. 119, n. 1, pp. 67–79, May 2005.

[4] M. Longcamp *et al.*, "Learning through Hand- or Typewriting Influences Visual Recognition of New Graphic Shapes: Behavioral and Functional Imaging Evidence", *J. Cogn. Neurosci.*, vol. 20, n. 5, pp. 802–815, May 2008.

[5] P. A. Mueller and D. M. Oppenheimer, "The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking", *Psychol Sci*, vol. 25, n. 6, pp. 1159–1168, June 2014.

[6] E. Ose Askvik, F. R., van der Weel, and A. L. H. van der Meer, "The Importance of Cursive Handwriting Over Typewriting for Learning in the Classroom: A High-Density EEG Study of 12-Year-Old Children and Young Adults", *Front. Psychol.*, vol. 11, pag. 1810, July 2020.

[7] K. Osugi *et al.*, "Differences in Brain Activity After Learning With the Use of a Digital Pen vs. an Ink Pen—An Electroencephalography Study", *Front. Hum. Neurosci.*, vol. 13, pag. 275, Aug. 2019.

[8] MIUR-Ufficio Gestione Patrimonio Informativo e Statistica, "I principali dati relativi agli alunni con disabilità per l'a.s. 2017/2018".

[9] C. Marquardt, M. Diaz Meyer, M. Schneider, and R. Hilgemann, "Learning handwriting at school–A teachers' survey on actual problems and future options," Trends Neurosci Educ, vol. 5(3), pp. 82–89, Sept. 2016.

[10] S. Rosenblum, P. L. Weiss, and S. Parush, "Handwriting evaluation for developmental dysgraphia: Process versus product", *Reading and Writing*, vol. 17, n. 5, pp. 433–458, July 2004.

[11] P. E. Tressoldi, C. Cornoldi, and A. M. Re, "*BVSCO-2. Batteria per la Valutazione della Scrittura e della Competenza Ortografica – 2*," Giunti Edu, 2019.

[12] C. Di Brina, "*BHK. Scala sintetica per la valutazione della scrittura in età evolutiva,"* Edizioni Erickson, 2010.

[13] M. Borean, G. Paciulli, L. Bravar, and S. Zoia, "*DGM-P. Test per la valutazione delle difficoltà grafo-motorie e posturali della scrittura,"* Organizzazioni Speciali Giunti, 2012.

[14] Z. Galaz *et al.*, "Advanced Parametrization of Graphomotor Difficulties in School-Aged Children", *IEEE Access*, vol. 8, pp. 112883–112897, 2020.

[15] G. Dimauro, V. Bevilacqua, L. Colizzi, and D. Di Pierro, "TestGraphia, a Software System for the Early Diagnosis of Dysgraphia", *IEEE Access*, vol. 8, pp. 19564–19575, 2020.

[16] L. Taverva, M. Tremolada, L. Barp, M. Blazina, and L. Lazzaro, "Esame delle prove di scrittura di bambini con DSA in termini di processo e di prodotto", procs. Giornate di Neuropsicologia dell'Età Evolutiva, Brixen, Jan. 2022.

[17] T. Gargot *et al.*, "Acquisition of handwriting in children with and without dysgraphia: A computational approach", *PLoS ONE*, vol. 15, n. 9, pag. e0237575, Sept. 2020.

[18] A. Accardo *et al.*, "A device for quantitative kinematic analysis of children's handwriting movements", in procs of MEDICON 2007, vol. 16, pp. 445–448, 2007.

[19] E. Pagliarini *et al.*, "Children's first handwriting productions show a rhythmic structure", *Sci Rep*, vol. 7, n. 1, pag. 5516, Dec. 2017.

[20] S. Rosenblum, S. Parush, and P. L. Weiss, "Computerized Temporal Handwriting Characteristics of Proficient and Non-Proficient Handwriters,"Am. J. Occup. Ther., vol. 57(2), pp. 129–138, Mar. 2003.

[21] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and Rating of Developmental Dysgraphia by Handwriting Analysis", *IEEE Trans. Human-Mach. Syst.*, vol. 47, n. 2, pp. 235–248, Apr. 2017,

[22] V. Fantasia, C. Provenzale, F. Taffoni, D. Formica, and L. Sparaci, "Many paths towards handwriting assessment: a first study comparing tools for the evaluation of grapho motor skills in Italian primary school children.", procs. Giornate di Neuropsicologia dell'Età Evolutiva, Brixen, Jan. 2021.

[23] V. Fantasia, F. Taffoni, D. Formica, and L. Sparaci, "Product versus process: Preliminary data comparing grapho motor skills assessment in Italian primary school children.", procs of Communication & Language Acquisition Studies in Typical & Atypical populations (CLASTA), May 2021.

[24] L. Sparaci, V. Fantasia, C. Provenzale, F. Lasorsa, D. Formica, and F. Taffoni, "Considerations on building new automated tools for grapho-motor skills assessment in childhood: a comparison among tests (forthcoming)", *Perceptual and motor skill*.

[25] D. Alamargot, D. Chesnet, C. Dansac, and C. Ros, "Eye and Pen: A new device for studying reading during writing", *Behavior Research Methods*, vol. 38, n. 2, pp. 287–299, May. 2006.

[26] H. Bezine, M. Kefi, e A. M. Alimi, «ON THE BETA-ELLIPTIC MODEL FOR THE CONTROL OF THE HUMAN ARM MOVEMENT», *Int. J. Pattern Recognit. Artif. Intell.*, pag. 15, 2007.

[27] R. Plamondon, C. O'Reilly, C. Rémi, e T. Duval, «The lognormal handwriter: learning, performing, and declining», *Front. Psychol.*, vol. 4, 2013, doi: 10.3389/fpsyg.2013.00945.