

# Dynamic time warping: A new method in the study of poor handwriting

Carlo Di Brina, Ralph Niels, Anneloes Overvelde,  
Gabriel Levi and Wouter Hulstijn

PRE-FINAL VERSION

## Abstract

Poor handwriting is a diagnostic criterion for developmental coordination disorder. Typical of poor handwriting is its low overall quality and the high variability of the spatial characteristics of the letters, usually assessed with a subjective handwriting scale. Recently, Dynamic Time Warping (DTW), a technique originally developed for speech recognition, was introduced for pattern recognition in handwriting. The present study evaluates its application to analyze poor handwriting. Forty children attending Dutch mainstream primary schools were recruited and based on their scores on the *Concise Evaluation Scale for Children's Handwriting* (Dutch abbreviation: BHK), 20 good and 20 poor writers (of whom 13 were scheduled for handwriting intervention) were identified. The groups were matched for age (7-9 years), school grade (grades 2 and 3) and handedness. The children subsequently wrote sequences of the letter 'a' on a graphics tablet in three conditions (normal, fast, and accurate). Classical kinematics were obtained and for each individual letter DTW was used to calculate the distance from the mean shape. The DTW data revealed much higher variability in the letter forms of the poor writers that was independent of the kinematic results of larger trajectories, faster movements, and higher pen pressure. The current results suggest that DTW is a valid and objective technique for letter-form analysis in handwriting and may hence be useful to evaluate the rehabilitation treatments of children suffering from poor handwriting. In education research it may be exploited to explore how children (should) learn to write.

## 1 Introduction

At a time when even small children spend hours at their PCs, handwriting seems on its way out. Nevertheless, children still spend between 30% and 60% of their time at elementary school on handwriting tasks [10, 27]. Here, many exhibit handwriting problems, the prevalence of which has been estimated at 10 to 34% [8, 19, 20]. Handwriting disorders have been referred to as a specific learning disability [2] and developmental right-hemisphere syndrome [5], and

are increasingly seen as a warning sign for developmental coordination disorder (DCD) [12]. However, the problems can manifest themselves in quite different levels of severity and can consequently not always be properly assessed and quantified despite the time school staff dedicate to the children in question. The identification of poor handwriting in children, as a symptom of dysfunction in mechanisms underlying the production of written language [6], is clearly in need of reliable instruments that assess the deficits adequately.

One approach is to use a scale that assesses handwriting samples on a number of specified criteria. Although less subjective than the global-holistic scaling that is common practice in schools [18], the outcome is still based on subjective expert ratings. For instance, the *Concise Evaluation Scale for Children's Handwriting* (Dutch abbreviation BHK) [7] uses a product-oriented approach and assesses 13 qualitative criteria, most of which are related to spatial characteristics (e.g., the spacing of the letters and words), while others concern letter formation (e.g., letter shape and size). In the present study we used the BHK to distinguish between normal (proficient) and poor or dysgraphic handwriting. Although the latter two terms are frequently used interchangeably, in this report we will use the first term only.

Digitizer-based technology has opened new avenues for the assessment of handwriting by providing reproducible and objective measures that distinguish between the writing of children with and without writing difficulties [17, 21, 22, 24, 25, 26]. The resulting analytical approach uses the kinematics of the children's movements to provide information on the real-time strategies and the dynamic characteristics of the deviant motor performance, including pen pressure, pen velocity, and trajectory length. From kinematics and so-called 'noise spectra' it was concluded that failure to control spatial accuracy (relative to letters and text composition) is the most salient discriminating feature between poor and good writers [24]. Trajectory length or total path length, overshootings and undershootings of the target zones, and the height and width of writing segments and single letters were also used as indicators of poor size control and spatially inconsistent motor behavior in poor handwriting [16, 19, 21, 23, 29].

However, this methodology has one major drawback: objective and measurable features only consider single aspects of the allograph production in isolation; they do not take overall form criteria into account. In particular, the spatial and kinematic features of handwriting movements vary widely between, as well as within, individuals across replications of the same symbol, and even across replications of analogous letter strokes [28]. Kinematic variability can be measured adequately using the digitizer but the recording technique in itself does not allow an automated and accurate evaluation of variability in the form of a letter, which is why until now all studies on form output mainly relied on subjective ratings.

Recently, *dynamic time warping* (DTW) was introduced for the analysis of handwriting, which may compensate this deficit by facilitating an objective assessment of the spatial characteristics of poor handwriting. This so-called trajectory-matching technique, originally designed by Kruskal and Liberman [9] for speech recognition purposes, was adapted by Niels, Vuurpijl, and

Schomaker [14] for written character recognition and forensic document examination. With DTW, trajectories of pen movements are compared using both spatial and temporal information. Allograph matching is performed by point-to-point comparison of two letter trajectories, recorded on an XY-tablet.

It is important to note that the movements in handwriting, as compared to the motions in reaching and grasping, are unique in the sense that they are not aimed at a single target but are planned to follow a spatial trajectory and to leave a trace of that trajectory for visual perception. The overall form of this trajectory is its main objective, and it is a remarkable feature of writing that a more or less invariant graphic output can be realized with different effectors (e.g., the preferred or non-preferred hand, the foot, or even the mouth; for a discussion, see [11]). By objectively analyzing the spatial-temporal patterns, DTW captures the essential character of writing, i.e., the overall shape of its graphic output. The trajectory-matching technique devised by Niels et al. [14] not only offers new perspectives for fundamental research comparing various models of handwriting, but also for the analysis of the distorted shapes produced by people with handwriting disorders.

The aim of the present study was to evaluate the power of DTW to distinguish good from poor writers. The handwriting of children that experience serious difficulties with this skill shows a larger variability and inconsistency in letter size [24]. There is also less consistency in the production and form of allographs. With DTW we hoped to obtain a measure of this variability in spatial characteristics regardless of size variability. It generates more 'human-like' and visually intuitive pattern matching than previous methods and allows an individual prototype (IP) to be deduced. By having the child write the same character repeatedly, the child's IP of that character can be calculated, allowing the distance from this prototype to be determined for each written character. The means and distribution of these distances thus provide a measure of the child's ability to produce a consistent pattern. As it is only the individual spatial-temporal variability in handwriting and the individual consistency in letter formation that are being measured, the technique may reveal small deviations from the child's own average form that may otherwise remain undetected.

In sum, with the study reported here we sought to determine the potential of DTW in the evaluation of children's handwriting by comparing the DTW distance measures and a number of kinematic variables of children with poor-quality script with those of sound writing proficiency. The children were asked to write repeatedly the cursive letter 'a' (as learned at school) in three conditions. We hypothesized that the children with poor handwriting would show larger variability in the shape (global spatial characteristics) of this letter in terms of DTW distance measures. We also investigated how well these measures would correlate with the children's kinematic characteristics.

## 2 Method

### 2.1 Participants and selection procedure

The children ( $N = 40$ ) were recruited from various Dutch mainstream primary schools. Classification of their handwriting was based on the BHK. Children with a BHK total score of 29 or higher, thus falling into the scale's *dysgraphic* range, were classified as poor writers (PW), while children who scored below 22, falling into the scale's *proficient* range, as good writers (GW). Children that scored between 22 and 28 (i.e., *borderline* range) were not included in the study. All children attended grades 2 or 3, had a mean age of 8 years (range: 7-10 years), and had received an average of 19.2 months ( $SD = 7.2$ ) of instruction and practice in writing. We opted for this age group because the literature on handwriting development suggests that by the time a child reaches the end of second grade, his or her handwriting has become more or less automatic, organized, and readily available [1, 15, 31]. Thus, a deficiency in these qualities at the age of our pupils is likely to be a sign of a (developmental) problem. The selection procedure yielded 20 children for both groups; all were boys and all were right-handed. There were no significant group differences with respect to age (GW: mean age = 98.1 months,  $SD = 9.5$ ; PW: mean age = 98.8 months,  $SD = 9.6$ ), school type or grade. Thirteen of the boys in the PW group had been recruited from a group of children scheduled for handwriting intervention on an individual basis. They had been identified by their parents and teachers as having significant handwriting difficulties. Their mean age (96.0 months,  $SD = 9.9$ ) was slightly lower than that of the seven other PW children (mean age = 104.1 months,  $SD = 7.5$ ;  $t(18) = 1.93$ ,  $p = .070$ ), who were recruited from a local primary school based on their teacher's assessment of their handwriting. Thus, none of the 20 PW children had received physical or occupational therapy for their motor problems prior to the study. All children were free from organic pathologies (no major neurological impairments or physical disabilities) and from cognitive or psychopathological impairments. The study was approved by the local institutional review board and the parents of all participants had given their written informed consent.

### 2.2 Instruments

#### 2.2.1 The Concise Evaluation Scale for Children's Handwriting - BHK

The BHK is a tool designed to screen children at Dutch elementary schools (grades 1-6) for poor handwriting on the basis of a piece of writing produced during a writing test. The children are requested to copy on unlined paper a standard text - or, for very slow writers at least the first five lines - that gradually increases in complexity, within 5 min. The first five sentences that contain simple and familiar monosyllabic words are used to assess the quality of handwriting on the 13-item scale. The total score on the BHK has three score bands: *Proficient* with a total score of 0-21, *Borderline* with scores between 22 and 28,

and *Dysgraphic* with a score exceeding 28. Copying speed was determined by counting the number of letters each child had produced within the allocated time, which was then translated into a decile score scaled to the norm for the child’s grade.

### 2.2.2 Digitizing tablet and on-line data collection

All writing sessions were supervised by one of the authors. The children performed the sequence task (see below) on unlined A4 paper that was fixed to the surface of the XY-tablet (Wacom Tablet, Intuos3), using a standard-shaped, wireless electronic, inking pen. The position of the pen tip and the force exerted along the pen’s axis were recorded with a sampling frequency of 200 Hz. The data were recorded online using dedicated Delphi software (developed by Bouwhuisen at the NICI) on a Pentium-IV laptop computer; data analysis was performed off-line.

### 2.2.3 Dynamic time warping distance analysis

Dynamic time warping (DTW) was used to analyze the variations in the child’s production of the letter ‘a’. We opted to study the repeated reproduction of this well-known single cursive allograph to exclude the influence of higher-order processes like spelling or text composition. This enabled us to focus exclusively on handwriting performance. The computerized handwriting evaluation method allows the similarity between the trajectories of two allographs of the same letter to be determined through a point-to-point comparison (see Figure 1).

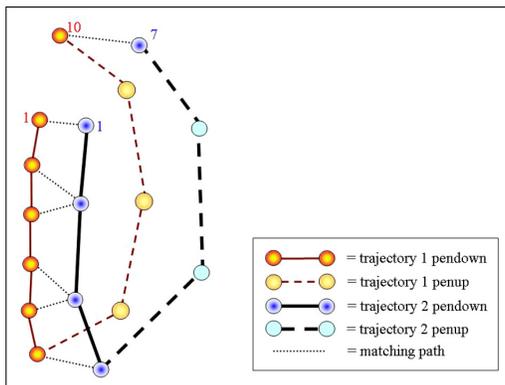


Figure 1: An example of trajectory matching using the dynamic time warping (DTW) technique illustrating the point-to-point comparison of two trajectories of the letter ‘i’, resulting in a matching path.

A specific feature of DTW is that the points are processed in the temporal order in which they were sampled during their production. The coordinates of successive points in two trajectories are compared to determine whether they

fulfill the so-called "matching criterion". Two points match if they either satisfy the "boundary" or the "pen-up/pendown" and the "continuity" criteria. The boundary criterion is satisfied while comparing the first point of the first curve with the first point of the second curve or while comparing the last points of the two curves (see the two points labeled '1' and the points labeled '10' and '7' in Figure 1). All other points need to meet the pen-up/pen-down criterion in that points from curve 1 can only match the points from curve 2 if both have the same status with respect to the position of the pen-tip vis-à-vis the writing surface (i.e., the pen either makes contact with the paper or hovers above it). They are also compared on the continuity criterion, which forces the matching between a point of the first curve and the nearest point of the second curve. If the relative temporal order of a point in curve 1 is about similar to the relative temporal order of a point of curve 2 (i.e., falls within a relative distance of plus or minus 0.13 - a value between 0 and 1 that was based on previous studies [14]), then these two points fulfill the continuity criterion (see Figure 1; for clarification and an exact definition of these criteria, see [14]). For each point of curve 1 all points of curve 2 that satisfy these requirements are evaluated and the point of curve 2 with the minimum distance to the point of curve 1 is chosen to form a pair of matching points. The path along these matching points is called the 'matching path'. The DTW distance is computed as the average Euclidean distance between all couples of matching points along the matching path.

Because the cursive letter 'a' needed to be written without pen lifts (as taught at school) we analyzed the pen-down points only; the few samples that did contain an occasional pen-up were excluded from further analysis. DTW distances are in arbitrary units because all letters were first rescaled in size (between -.5 and +.5) to derive the individual prototype (IP), the child's "personal average", for each condition. In this study the prototype was calculated by the *Merge-samples algorithm* (for details see [14]), a procedure that creates a DTW distance from the IP for every single letter. Two extreme examples of the distribution of these DTW distances are presented in Figure 2. As depicted in the top graph, the good writer (GW14) had a mean DTW distance of .0057, and only one of the letters with a DTW distance larger than .010. The poor writer (PW22; bottom graph) had a much wider distribution of DTW distances with a mean of .0390 and 85% of the letters larger than .010. In later analyses the limit of .010 proved to be an optimal criterion to separate the letters that were close to the IP from those that deviated from it. Letters with a DTW distance larger than .010 were called 'deviant letters' and their percentage is used as an additional DTW variable.

### 2.3 Procedure

All children were tested individually in an environment that was similar to the child's normal writing conditions at school. After they had copied the standard text of the BHK, they were offered a few practice trials to familiarize themselves with writing on the XY-tablet; here they learned to start and stop the recording

of a trial by placing the pen tip in the dedicated circles positioned at the left (start) and the right (stop) of the frame in which they wrote their sequences of the letter 'a' (see Figure 3). Auditory signals marked the beginning and the end of the trial. The total time needed for all the writing tasks (BHK, tablet practice and sequence test) was about 30 min.

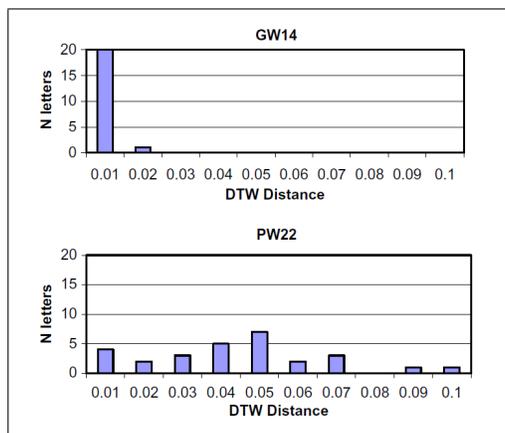


Figure 2: The distribution of the distance measures obtained through dynamic time warping (DTW) for the allographs of the letter 'a' as produced by a "good writer" (GW14, top panel) and those generated by a "poor writer" (PW22, bottom panel), with the DTW distance from the individual prototype (IP) plotted on the X-axis and the number of letters falling within the specified distance categories on the Y-axis.

The letter 'a' was produced in three conditions that were familiar to school children. In the normal condition (N), which resembled an ordinary, spontaneous writing assignment, the children were asked to write the letter 'a' several times with each letter detached from the next one while remaining within the boundaries of a wide frame. In the fast condition (F) they needed to write the letter 'a' repeatedly as fast as possible while keeping the letters legible. In the third, accurate condition (LL), the children wrote the letters 'a' between two solid horizontal lines, the width between them similar to that used during dedicated handwriting lessons. In all conditions the experimenter signaled when a sufficient number of characters (at least 20) had been written.

Only legible characters were included in the analyses. Corrected forms, characters with up-strokes or those that were not considered to represent a proper 'a' were excluded. The mean number of characters that was analyzed was slightly lower for the poor writers in the normal condition (GW: 19.5, PW: 23.2;  $t(38) = 2.008$ ,  $p = .052$  (two-sided)), but similar for the two groups in the two other conditions.

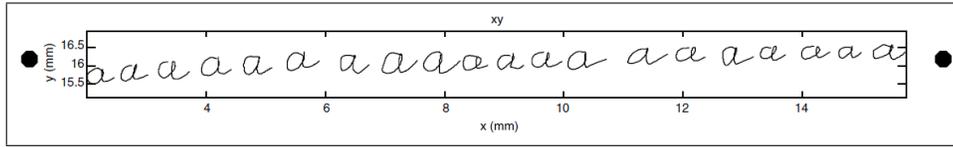


Figure 3: A sample of a sequence of allographs of the letter 'a' generated in the normal condition by a proficient writer (GW14) as recorded by means of the digitizer tablet (also showing the trial start (left) and stop circles).

## 2.4 Data analysis

The following kinematic variables were analyzed: writing time (total time in seconds taken to complete the letter 'a'), trajectory length (distance in centimeters covered by the pen tip), absolute velocity (average absolute velocity of the pen-down movements in cm/s), total "stop" time (the intervals during which the pen halted on the digitizer), and pen pressure (average axial pen pressure in Newtons).

Statistical testing was performed with SPSS. The kinematic variables, the mean DTW distance and the percentage of deviant letters were entered into a general linear model repeated-measures procedure with condition (N, F, and LL) as the repeated measure and group (PW and GW) as the between-subject variable. The percentage of deviant letters (characters with a DTW distance larger than .010) were also analyzed with a Mann-Whitney  $U$  test. In all analyses a significance criterion of  $\alpha = .05$  was used.

## 3 Results

### 3.1 BHK total score and execution times

Our selection procedure yielded two groups that differed widely as to handwriting quality. The BHK total score for the GW group ranged from 5 to 19, with a mean of 12.55 (SD 4.88), whereas the scores of the PW group ranged from 29 to 49, with a mean of 35.05 (SD 5.94), with the group difference significant ( $t(38) = 13.08$ ;  $p = .0005$ ). However, both groups had similar mean BHK-speed scores (GW: mean=3.95, range 1-8, SD 2.46; PW: mean=4.60, range 1-10, SD 3.58;  $t(38) = 0.668$ ,  $p = .51$ ).

### 3.2 Kinematic variables

Table 1 lists the means of the five kinematic variables in each of the three conditions for the two groups as well as the results of the two DTW variables. A multivariate test on all seven variables of Table 1 showed the effect for condition ( $F(14, 25) = 10.98$ ,  $p < .001$ ), the group difference ( $F(7, 32) = 4.20$ ,  $p = .002$ )

Table 1: Means of the five kinematic and the two DTW (dynamic time warping) variables in each of the three writing conditions for the good (GW) and the poor (PW) writers, and the univariate F values for Condition (Cond), Condition·Group (C·GR), and Group.

<i>Kinematics</i>	Group	Condition			<i>F</i> cond df(2,37)	<i>F</i> C*GR df(2,37)	<i>F</i> Group df(1,38)
		Normal	Fast	Accurate			
Writing time	GW	1.125	.808	1.223	38.05**	1.31	<1.00
	PW	1.226	.938	1.224			
Trajectory length	GW	1.736	2.094	1.515	21.97**	7.71**	13.69**
	PW	2.834	3.200	1.769			
Velocity	GW	1.707	2.748	1.395	43.01**	3.78**	7.03*
	PW	2.608	3.775	1.622			
Stop time	GW	.289	.154	.387	18.28**	<1.00	<1.00
	PW	.259	.144	.334			
Pen pressure	GW	119	121	108	6.01**	1.63	16.96**
	PW	265	282	240			
<i>DTW</i>		Normal	Fast	Accurate	df(2,37)	df(2,37)	df(1,38)
Mean distance	GW	.0134	.0152	.0120	1.35	3.40*	2.46
	PW	.0194	.0135	.0157			
% Deviant letters	GW	43.3	47.7	37.8	<1.00	4.09*	5.32*
	PW	62.4	48.7	55.0			

\*  $p < 0.5$ , \*\* $p < .01$

and the group by condition interaction ( $F(14, 25) = 2.90$ ,  $p = .010$ ) to be significant.

The results on *writing time* were more or less as expected, being significantly shorter in the fast condition and about equal for the two groups in each of the three conditions. *Trajectory length*, however, showed great group differences. Averaged over the three conditions, the poor writers produced larger letters, but, when forced to write between the lines (accurate condition), they wrote nearly as small as the good writers did. The significant group by condition interaction for trajectory length was therefore based on a group difference in the normal and fast conditions and had nearly disappeared in the accurate condition. The *velocity* results reflected the results on trajectory length. When writing time is equal for the two groups and trajectory length larger, then velocity must follow. *Stop times* did not discriminate between the poor and good writers. In both groups, the pen halted at approximately 25% of the entire writing time and was significantly shorter in the fast condition. Of all seven variables *pen pressure* showed the largest group difference: it was slightly higher in the fast condition but, irrespective of condition, the pen pressure of the poor writers was more than twice that of the good writers.

### 3.3 Dynamic time warping - distance measures

The mean DTW distances from the Merge-samples' individual prototype (IP) for the two groups in the three conditions are shown in Table 1. The difference between the two groups was largest in the normal condition, and completely absent or nearly reversed in the fast condition, resulting in a significant group by condition interaction. Table 1 and Figure 4 show the mean percentages of deviant letters (letters with a DTW distance  $>.010$ ) in the three conditions for the two groups. The means differed between the two groups in the normal (Mann-Whitney  $U = 110$ ,  $p = .015$ ) and in the accurate condition (Mann-Whitney  $U = 118$ ,  $p = .027$ ). Note the complete absence of a group difference in the fast condition.

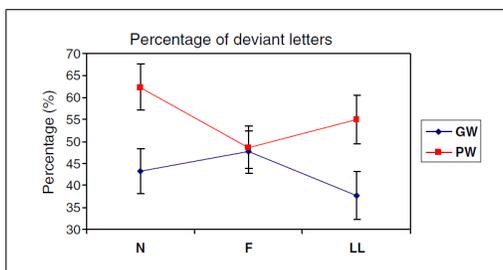


Figure 4: The percentages of deviant letters (i.e., allographs having a DTW distance  $>.010$ ) for the "good writers" (GW;  $n = 20$ ) and the "poor writers" (PW;  $n = 20$ ) in the normal (N), fast (F) and the accurate (LL) writing condition. Bars denote standard errors.

Table 2: Correlations between three kinematic variables and the two DTW (dynamic time warping) measures.

	Pen pressure	Trajectory length	Velocity	DTW mean distance	DTW % deviant
Pen pressure		.361	.281	.211	.003
Trajectory length	.361		.576**	.466*	.158
Velocity	.281	.576**		.462*	.320
DTW mean distance	.211	.466*	.462*		.754**
DTW % deviant	.003	.158	.320	.754**	

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level (1-tailed).

### 3.4 Associations between the kinematic measures and DTW-distance measures

To determine whether the DTW distance measurements had added to the traditional kinematic analyses, we inspected the DTW data obtained in the normal writing condition because this showed the largest effects. Correlations between the variables, which were calculated for the PW group only, are presented in Table 2. To simplify the presentation, only the correlations of the kinematic variables that showed the largest group effects are depicted. These are, in the order of their group  $F$  values in Table 1: pen pressure, trajectory length and velocity. As can be seen from the data in Table 2, pen pressure is rather independent from the other variables. Velocity and trajectory length were significantly correlated and the two DTW variables were highly interrelated, but only DTW mean distance correlated with the two interrelated kinematic variables of trajectory length and velocity.

To optimize these findings, we also conducted a discriminant analysis. When the seven variables of Table 1 were entered stepwise only pen pressure, trajectory length and DTW mean distance were retained (Wilks' Lambda (3,36) = .421,  $p < .001$ ), and 90.0% of the original grouped cases were correctly classified. When DTW mean distance was omitted from the discriminant analysis, the percentage of correctly classified persons dropped to 80.0%.

To test if the group differences in DTW measures would remain after correction for the kinematic variables, covariance analyses were performed. With the five kinematic variables from Table 1 as covariates, the group effect for DTW mean distance was significant ( $F(1,33) = 16.31$ ,  $p < .001$ ) and, although smaller, the group effect for DTW percentage of deviant letters remained significant ( $F(1,33) = 5.95$ ,  $p = .020$ ).

## 4 Discussion

The aim of this study was to evaluate the usefulness of dynamic time warping (DTW) as a new method for analyzing the spatial characteristics of the handwriting of young children with writing problems. The DTW distance measures we obtained in two samples of boys aged between 7 and 9 years showed that, relative to age-and grade-matched controls, the boys with poor handwriting showed larger variability in the forms of the letter 'a' and a higher percentage of deviant allographs. Interestingly, this variability in letter form constituted an extra dimension that was independent of the group differences in trajectory length, velocity, and pen pressure that were also found.

Compared to the performance of their matched peers, the performance of the poor writers was characterized by numerous deviations from their individual prototypical form (IP). In fact, in relation to the overall form characteristics, the distance of most of the letters they produced exceeded the criterion value of .010 of their IP. Even though the letter 'a' was well-practiced, the allographs of the poor writers were significantly more variable in their form than those

of the good writers. This is in agreement with the literature that identified the defining features of poor writing were letter-form errors and misproportions (letter distortions [3]), inconsistency in letter sizes (more spatial errors and inaccuracies [24]), difficulties related to the consistency of letter formation [16], and the irregularity of size and slant [13, 29].

The current results on letter size and speed confirm previous observations in children with poor handwriting. Rosenblum et al. [16] suggested their larger script may be due to the fact that it might help them achieve greater legibility and Mojet [13] posited it might simply make it easier for them to obtain an acceptable writing result because it requires less precise letter formation. However, this typical divergence in letter size cannot explain the variability in letter form we observed in our sample of poor writers because the DTW distance measures were calculated on trajectories that were normalized for size. In addition, when we controlled for the correlation between DTW distance and trajectory length (and writing velocity), as was done in the discriminant and in the covariance analyses, the DTW distance measures still showed relatively large and significant group differences. This suggests that the larger variability in the overall form characteristics we observed in the poor writers is not only a salient feature discriminating between poor and good writers, but a characteristic that is independent of a more general deviation in size.

Our manipulation of the writing conditions produced interesting and somewhat unexpected results on the DTW distance measures. Writing fast did not influence the group differences on trajectory length, writing speed and pen pressure but did greatly reduce the group difference in letter-form variability. Apparently, when pressed for time good writers start producing more variable letter forms while poor writers improve and generate less variable letter forms. Whether in the fast condition the poor writers' allographs still deviated more from the standard letter 'a' they had learned at school than the letters generated by the proficient writers, is subject for future study. This type of research is now feasible because the DTW analysis not only allows a comparison of individual trajectories with their own personal prototype but also a comparison with the teacher-set prototype.

When forced to write within a constricted space, the trajectory lengths of both groups became markedly shorter while writing time increased (viz Fitts' law) and velocity greatly decreased. Only pen pressure remained consistently high in the poor writing group. The DTW measures were affected slightly, showing somewhat smaller form variability between the two groups. Whether size constraints like the one we imposed in the accuracy condition indeed elicit (slightly) more consistent letter forms remains to be established.

The larger variability in the letter shapes in the poor writers group is consistent with earlier studies (see Section 1) that also reported poor size control and spatially inconsistent behavior in poor writers. Poor writers fail to obey spatial constraints and their handwriting lacks consistency [24]. This variability reduces the ease with which individual characters can be recognized and the readability of the handwriting, overall [4]. Until now, the assessment of script legibility was mostly based on conventional, subjective judgments as they still

discriminated between the variables of poor and proficient handwriting better than existing digitizer-based assessments [19]. Because it combines the advantages of both conventional (or expert-based) and computerized measures, the DTW technique constitutes a substantial improvement on earlier computerized methods: being digitizer-based, DTW provides objective and detailed information about letter shape and, indirectly, about the overall legibility of the handwritten product as well.

With our present investigations we only looked at the internal (in)consistencies between letter shapes by measuring their distance from the individual's average shape, the IP. It must be emphasized that also another prototype, one produced by a teacher, for instance, can be used. Comparisons with the latter prototype might provide us with more insight into how well a child is able to match this "ideal" form when learning a new letter. DTW may thus also be exploited to evaluate the efficacy of different teaching methods or special training programmes. We hope the results of the present study will prompt further research on DTW and how well it determines the degree to which the (shape of) writing movements are automated in children with poor handwriting, and, secondly, in what way their writing differs from teaching prototypes in schools and peers who are proficient writers. Such studies would also facilitate the collection of normative data on class-related IP and prototypical forms, thus providing an objective measure of the severity of a child's writing problems. In addition, the possible correlations between DTW measures and scores on traditional handwriting assessment scales like the BHK merit further scrutiny. It would be interesting to compare various aspects of letter formation such as inconsistent letter size (BHK item 8), distorted letters (BHK item 10), and ambiguous letter shapes (BHK item 11).

In their frequently cited model of handwriting, Van Galen et al. [28] posit three processing stages: the retrieval of an abstract motor program, the parameterization of this program and, thirdly, the muscular initiation of the parameterized program. Spatial variability or inconsistencies in written characters could then result from deficiencies in each of these processes (see also [4]). In his neuropsychological theory of motor-skill learning, Willingham [30] points to the importance of two additional processes in motor programming: "perceptual-motor integration" and "sequencing", which may also be deficient and hence result in poor or highly variable graphic output. Finally, the way the writer monitors his output might be impaired resulting in inadequate correction or adaptation of the movement trajectory. It is tempting to assume that children with poor handwriting function sub-optimally in all these motor control processes. Of course, matters are far more complex. In some children poor perceptual-motor integration or variable sequencing may be at the basis of their poor script. Others may not have sufficiently mastered the delicate muscle control needed in the initiation or dynamic stages of movement production. The DTW technique alone cannot answer the question on which of these processes might be responsible for the writing deficits of an individual child. Its outcome measures, however, do allow accurate comparisons of graphic output under conditions where each of these processes is taxed separately, thus enabling analysis of their separate contribu-

tions to the letter distortions in poor writing. As such, it may help clinicians determine the origin of a child's handwriting problems and may subsequently be used in the follow-up of children who have been diagnosed with handwriting disorders, allowing the rehabilitation and learning stages to be closely monitored and graded.

In conclusion, dynamic time warping seems a promising tool in the investigation of handwriting in that it provides an objective evaluation of the overall form characteristics of letters and their variability. Its measures allow an overall judgment of the quality of the output that other temporal-spatial parameters cannot provide and it supplies information that is not correlated with kinematic data. The technique may thus enhance our understanding of the processes involved in handwriting based on an unbiased assessment of the end-product. It may also be of use in clinical practice by facilitating the identification of writing problems and the monitoring of rehabilitation programs.

## Acknowledgments

We would like to thank all the children, parents and teachers for their participation and enthusiastic contributions to the study, and Rob Groot Zwaaftink for his indispensable help in testing the children and analyzing the data.

## References

- [1] V. Berninger, D. Mizokawa, and R. Bragg. Theory-based diagnosis and remediation of writing. *Journal of School Psychology*, 29:57–59, 1991.
- [2] J. K. Brown. Learning disabilities: A paediatric neurologist's point of view. *Transactions of the College of Medicine of South Africa*, pages 49–104, December 1981.
- [3] M. R. Eidlitz and M. Simner. The nature of dysgraphic handwriting in grade-1 children. In G. L. Leedham, V. Sagar, and X. Xuhong, editors, *Proceedings of the ninth biennial conference of the international graphonomics society*, pages 171–175. The Netherlands: Manigary Technological University, 1999.
- [4] S. Graham, M. Struck, J. Santoro, and V. W. Berninger. Dimensions of good and poor handwriting legibility in first and second graders: Motor programs, visual-spatial arrangement, and letter formation parameter setting. *Developmental Neuropsychology*, 29:43–60, 2006.
- [5] V. Gross-Tsur, R. S. Shalev, O. Manor, and N. Amir. Developmental right-hemisphere syndrome: Clinical spectrum of the nonverbal learning disability. *Journal of Learning Disability*, 28:80–86, 1995.
- [6] L. Hamstra-Bletz and A. Bloete. A longitudinal study on dysgraphic handwriting in primary school. *Journal of Learning Disability*, 26:689–699, 1993.

- [7] L. Hamstra-Bletz, J. De Bie, and B den Brinker. Concise evaluation scale for children's handwriting. 1987.
- [8] M. J. Jongmans, E. Linthorst-Bakker, Y. Westenberg, and B. C. M. Smits-Engelsman. Use of a taskoriented self-instruction method to support children in primary school with poor handwriting quality and speed. *Human Movement Science*, 22:549–566, 2003.
- [9] J. Kruskal and M. Liberman. The symmetric time-warping problem: from continuous to discrete. In D. Sankoff and J. Kruskal, editors, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons*. Addison-Wesley, Reading, Massachusetts, 1983.
- [10] K. McHale and S. A. Cermak. Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems. *American Journal of Occupational Therapy*, 46:898–903, 1992.
- [11] R. G. J. Meulenbroek, D. A. Rosenbaum, A. J. W. M. Thomassen, L. D. Loukopoulos, and J. Vaughan. Adaptation of a reaching model to handwriting: How different effectors can produce the same written output, and other results. *Psychological Research*, 59:64–74, 1996.
- [12] L. T. Miller, C. A. Missiuna, J. J. Macnab, T. Malloy-Miller, and H. J. Polatajko. Clinical description of children with developmental coordination disorder. *Canadian Journal of Occupational Therapy*, 68:5–15, 2001.
- [13] J. W. Mojet. Characteristics of the developing handwriting skill in elementary education. In A. M. Wing J. Wann and N. Søvik, editors, *Development of graphic skills*, pages 53–74. London: Academic Press, 1991.
- [14] Ralph Niels, Louis Vuurpijl, and Lambert Schomaker. Automatic allograph matching in forensic writer identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(1):61–81, February 2006.
- [15] A. O'Hare. Dysgraphia and dyscalculia. In K. Whitmore, H. Hart, and G. Willems, editors, *A neurodevelopmental approach to specific learning disabilities*, pages 96–119. Cambridge: MacKeith Press, 1999.
- [16] S. Rosenblum, A. Y. Dvorkin, and P. L. Weiss. Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting. *Human Movement Science*, 25:608–621, 2006.
- [17] S. Rosenblum, S. Parush, and P. Weiss. Temporal measures of poor and proficient handwriters. In R.G.J. Meulenbroek and B. Steenbergen, editors, *Proceedings of the tenth biennial conference of the international graphonomics society*, pages 119–125, Nijmegen, The Netherlands, 2001. University of Nijmegen: IGS Pub.

- [18] S. Rosenblum, P. L. Weiss, and S. Parush. Product and process evaluation of handwriting difficulties. *Educational Psychology Review*, 15:41–81, 2003.
- [19] S. Rosenblum, P. L. Weiss, and S. Parush. Handwriting evaluation for developmental dysgraphia: Process versus product. *Reading and Writing: An Interdisciplinary Journal*, 17:433–458, 2004.
- [20] N. Rubin and S.E. Henderson. Two sides of the same coin: Variation in teaching methods and failure to learn to write. *Special Education: Forward Trends*, 9:17–24, 1982.
- [21] M.M. Schoemaker, J.M.H. Schellekens, A.F. Kalverboer, and L. Kooistra. Pattern drawing by clumsy children: A problem of movement control. In M. L. Simner, W. Hulstijn, and P. Girouard, editors, *Contemporary issues in the forensic, developmental, and neurological aspects of handwriting*, volume 1, pages 45–64. Toronto, Canada: Association Forensic Document Examiners, Monograph of the Association of Forensic Document Examiners, 1994.
- [22] B. C. M. Smits-Engelsman. *Theory-based diagnosis of fine motor-coordination development and deficiencies using handwriting tasks*. PhD thesis, The Netherlands: University of Nijmegen, 1995.
- [23] B. C. M. Smits-Engelsman, A. S. Niemeijer, and G. P. Van Galen. Fine motor deficiencies in children diagnosed as dcd based on poor grapho-motor ability. *Human Movement Science*, 20:161–182, 2001.
- [24] B. C. M. Smits-Engelsman and G. P. van Galen. Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay? *Journal of Experimental Child Psychology*, 67:164–184, 1997.
- [25] N. Søvik, O. Arntzen, and R. Thygesen. Relation of spelling and writing in learning disabilities. *Perceptual and Motor Skills*, 64:219–236, 1987.
- [26] N. Søvik, O. Arntzen, and R. Thygesen. Writing characteristics of "normal", dyslexic and dysgraphic children. *Journal of Human Movement Studies*, 31:171–187, 1987.
- [27] M. H. Tseng and S. M. K. Chow. Perceptual-motor function of school-age children with slow handwriting speed. *American Journal of Occupational Therapy*, 54:83–88, 2000.
- [28] G. P. Van Galen, S. J. Portier, B. C. M. Smits-Engelsman, and L. R. B. Schomaker. Neuromotor noise and deviant movement strategies as an explanatory ground for poor handwriting in children. *Acta Psychologica*, 82:161–178, 1993.
- [29] J. P. Wann. Trends in the refinement and optimization of fine-motor trajectories: Observations from an analysis of the handwriting of primary school children. *Journal of Motor Behavior*, 19:13–37, 1987.

- [30] D. B. Willingham. A neuropsychological theory of motor skill learning. *Psychological Review*, 105:558–584, 1998.
- [31] J. Ziviani and J. Elkins. An evaluation of handwriting performance. *Educational Review*, 36:249–261, 1984.